HISTORY of TOPOLOGY

Edited by
I.M. James

NORTH-HOLLAND

HISTORY *of* TOPOLOGY

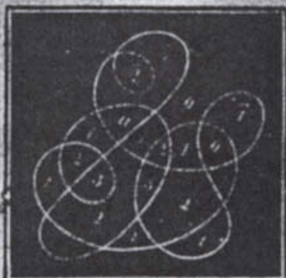
Vorstudien

zur

TOPOLOGIE.

Von

JOHANN BENEDICT LISTING.



(Mit eingedruckten Holzschnitten.)

Abgedruckt aus den Göttinger Studien. 1847.

Göttingen
bei Vandenhoeck und Ruprecht.
1848.

HISTORY
of
TOPOLOGY

Edited by

I.M. James
Oxford University, UK

1999
ELSEVIER
Amsterdam • Lausanne • New York • Oxford • Shannon • Singapore • Tokyo

ELSEVIER B.V.
Radarweg 29
P.O. Box 211, 1000 AE
Amsterdam, The Netherlands

ELSEVIER Inc.
525 B Street
Suite 1900, San Diego
CA 92101-4495, USA

ELSEVIER Ltd
The Boulevard
Langford Lane, Kidlington,
Oxford OX5 1GB, UK

ELSEVIER Ltd
84 Theobalds Road
London WC1X 8RR
UK

First edition 1999
Reprinted 2006

British Library Cataloguing in Publication Data
A catalogue record is available from the British Library.

Library of Congress Cataloging in Publication Data
History of topology / editor. I.M. James.
p. cm.
Includes bibliographical references and index.
ISBN 0-444-82375-1 (alk. Paper)
I. Topology-History. I. James, I. M. (Ioan Mackenzie), 1928–
QA611.A3H57 1999
514'.09-dc21 99-25564
CIP

ISBN: 0 444 82375 1

♾ The paper used in this publication meets the requirements of ANSI/NISO Z39.48-1992 (Permanence of Paper).
Printed in Great Britain bv MPG Books Ltd.. Bodmin. Cornwall.

Preface

Topology, for many years, has been one of the most exciting and influential fields of research in modern mathematics. Although its origins may be traced back several hundred years it was Poincaré who, to borrow an expression used of Möbius, "gave topology wings" in a classic series of articles published around the turn of the century. While the earlier history, sometimes called the prehistory, is also considered, this volume is mainly concerned with the more recent history of topology, from Poincaré onwards.

As will be seen from the list of contents the articles cover a wide range of topics. Some are more technical than others, but the reader without a great deal of technical knowledge should still find most of the articles accessible. Some are written by professional historians of mathematics, others by historically-minded mathematicians, who tend to have a different viewpoint.

Most of the material has not been published before. Topology is a large subject, with many branches, and it proved quite impossible to cover everything. The emphasis is on what might be called classical topology rather than on general (or point-set) topology: a separate history of general topology is in the process of publication under a different aegis.

However, I believe the articles will be found to cover most of the major topics. The order in which they are arranged is partly chronological and partly according to subject matter. The last part of the book is more concerned with the people who were important in the development of the subject. In particular short biographies of a number of them are given, based on material already in the literature, and rather longer biographies of some others, which contain material not previously published.

This volume is one of the fruits of a research project on the history of topology, for which I was awarded a Fellowship by the Leverhulme Trust. As well as the forty scholars, from many countries, who contributed the articles I would like to thank many others who provided valuable advice, encouragement and information. Wherever possible, sources of quotations etc. are acknowledged in the text. I must thank the mathematical societies of Brazil and The Netherlands for permission to reprint material which originally appeared in their publications, and Oxford University Press for permission to reprint an article from the recent supplement to the Dictionary of National Biography. Finally I would like to thank Arjen Sevenster, of Elsevier Science, for persuading me to take this on.

I.M. James
Mathematical Institute, Oxford
February 1998

Acknowledgement of Illustrations

Every effort has been made to discover the original sources of the photographs etc. used. It is regretted that this has not been possible in every case.

The photograph of Cantor (Chapter 1) is taken from one held by the University archive in Halle/Saale, Germany. The photograph of Hurewicz and Wallman (Chapter 1) is from the collection of Mrs Leona Wallman, that of Weyl (Chapter 7) from the collection of Mrs Ellen Weyl, and that of Freudenthal (Chapter 39) from the collection of Dr Mirjam Freudenthal. The portraits of Dehn (Chapter 36) and Seifert (Chapter 40) were taken by Frau Elizabeth Reidemeister, that of Whitehead (Chapter 32) by Professor Kosinski. The portrait of Poincaré (Chapter 6) appeared in vol. 38 of Acta Mathematica, that of Möbius (Chapter 32) in the Gesammelte Werke (Hirzel, Leipzig, 1885), that of Betti (Chapter 32) in the Opere Matematiche (Hoepli, Milan, 1903), that of Lefschetz (Chapter 32) in the Selected Papers (Chelsea, New York, 1971), that of Hurewicz (Chapter 32) in the Proceedings of the International Symposium of Algebraic Topology held in Mexico (1958), and that of Adams (Chapter 32) in Biographical Memoirs of the Royal Society, London 1990. Finally the portrait of Alexander (Chapter 32) is from the archives of the Mathematisches Forschungsinstitut, Oberwolfach, that of Listing (Chapter 33) from the University Library of Göttingen, as is the photograph of the cover of his Vorstudien used as a frontispiece to the present volume, that of Hopf (Chapter 38) from the Wissenschaftshistorische Sammlungen der ETH-Bibliotek, Zurich, as is the group portrait of the participants in the 1935 Moscow conference (Chapter 29).

Contents

This Page Intentionally Left Blank

CHAPTER 1

The Emergence of Topological Dimension Theory

Tony Crilly

Middlesex University, UK
E-mail: t.crilly@mdx.ac.uk

With the Assistance of Dale Johnson*

1. Introduction

The concept of dimension, deriving from our understanding of the dimensions of physical space, is one of the most interesting from a mathematical point of view. During the nineteenth and early twentieth century, mathematicians generalised the concept and probed its meaning. What had been a commonplace of experience became a focus for mathematical activity.

One extension of the meaning of dimension was the consideration of a mathematical space of n-dimensions. Although a revolutionary idea, the mathematical space of n-dimensions was regarded as an extrapolation from the "three dimensionality" of ordinary space. The idea of physical space being three-dimensional, an old and well accepted notion, was relatively uncontentious. While metaphysical questions concerning the meaning of four- and higher-dimensional geometry were raised, n-dimensional "hyperspace" was accepted and studied by such notable mathematicians as A.L. Cauchy (1789–1857), Arthur Cayley (1821–1895), and Hermann Grassmann (1809–1877). The principal application of "n-dimensions" in the nineteenth century was to projective and non-Euclidean geometry. Geometers studying these subjects readily accepted the notion of dimension on an intuitive basis. There was little impulse to probe the character of dimension itself.

In 1877, Georg Cantor (1845–1918) looked at dimension in a different way. He showed that the points of geometrical figures like squares, "clearly 2-dimensional", could be put into one-to-one correspondence with the points of straight line segments, "obviously 1-dimensional". The "simple" idea of dimension was immediately rendered problematic. "Dimension" came under the spotlight and more sophisticated questions were asked. For

*This article is primarily based on articles of Dale Johnson [16–19]. It also reflects information in the more recent articles of Miroslav Katetov and Petr Simon [21] and Teun Koetsier and Jan van Mill [24].

HISTORY OF TOPOLOGY
Edited by I.M. James

instance, in what sense was dimension a geometrical invariant? Could the dimension of a space and the dimension of its image under a mapping be different?

Paradoxes and contradictions have often challenged mathematicians and led them to research many problems. The long-term result of Cantor's paradoxical result was the development of an entire branch of topology: dimension theory. By this we principally mean *topological* dimension theory, that is, dimension theory free of metrical considerations. This article surveys this history, but the reader who wishes to delve further will need to consult the Selected References, in particular [16–19]. Since the appearance of these papers, other authors have considered its history and some of their works have been included in the Selected References.[1] In particular, two biographies are important: of *Georg Cantor* by Joseph Dauben [6] and of *L.E.J. Brouwer* by Walter van Stigt [36].

In essence, the three problems of *defining*, *proving* and *explaining* have been fundamental to the growth of topological dimension theory:

- The problem of defining the concept of dimension itself.
- The problem of proving that the dimension of mathematical spaces is invariant under certain types of mapping.
- The problem of explaining the number of dimensions of physical space.

The first and second problems, mathematical in nature, have been the most important direct influence on the growth of the theory of dimension. The third, a problem of physics or cosmology, has provided an indirect but significant motivation for the development of the theory from outside the mathematical domain.

2. Early history

The definitional problem seeking to answer the question "what is dimension?" is detectable in the writings of the Greek philosophers and mathematicians. To indicate that "dimension" has ancient roots we mention two of the most prominent authors.

According to Euclid, a *point* is that which has no part, a *line* is breadthless length, and a *surface* is that which has length and breadth only (*Book I*). A *solid* is that which has length, breadth and depth (*Book XI*). Euclid's definitions show a concern for a rudimentary "theory of dimension" by the recognition of a "dimension" hierarchy in the sequence of primary geometrical objects: point, line, surface, solid. A passage from Aristotle's *On the Heavens* shows a similar motivation. In it, Aristotle is more definite, even if its tone is more metaphysical:

> Of magnitude that which (extends) one way is a line, that which (extends) two ways a plane, and that which (extends) three ways a body. And there is no magnitude besides these, because the dimensions are all that there are, and thrice extended means extended all ways. For, as the Pythagoreans say, the All and all things in it are determined by three things; end, middle and beginning give the number of the All, and these give the number of the Triad [17, p. 104].

Other eminent philosophers and scientists considered questions about dimension including Galilei Galileo (1564–1642), Gottfried Wilhelm Leibnitz (1646–1716) and Immanuel Kant (1724–1804). While all of these touched on dimension in some form or other, there is

[1] I would like to thank Teun Koetsier, Jan van Mill and Petr Simon for sending me copies of their recent work on the history of dimension theory [21, 24].

no suggestion that they sought to create anything like modern dimension theory or that they were working towards modern dimension theory as it stands today. Dimension theory is primarily a modern subject of mathematics; its main historical roots lie in the early nineteenth century when the Bohemian priest Bernard Bolzano (1781–1848) examined several facets of the definitional problem and proposed some interesting solutions.

Bolzano sought precise definitions of geometrical objects. "At the present time", he wrote in 1810, "there is still lacking a precise definition of the most important concepts: line, surface, solid" [16, p. 271]. This dull essentialist problem of definitions, conceived within the limits of Euclidean geometry, led him to break from the bonds of traditional geometry. Bolzano stressed the theoretical role of mathematics and its "usefulness" in exercising and sharpening the mind. Rigour in pure mathematics was uppermost in his thoughts. For example, he regarded it a mistake to make any appeal to motion as it was foreign to pure geometry. This purge of motion from geometry is relevant to Bolzano's dimension-theoretic definitions of line, surface, and solid. For instead of taking a line as the path of a moving point, as for example was done by Abraham Kästner (1719–1800), Bolzano attempted to define the concept of line independently of any idea of motion.

A basic feature of Bolzano's outlook on research in mathematics was his view that mathematics stands in close relation to philosophy. "My special pleasure in mathematics rests only in its purely speculative part", admitted Bolzano in his autobiographical writings [16, p. 263]. His youthful *Betrachtungen* is heavily imbued with philosophy and this illustrated his deep concern for the logical and foundational issues in mathematics. Bolzano's concern over definitions required that he seek the "true" definitions for the objects of geometry. Undoubtedly, this essentialist philosophy is to blame for the main shortcomings of his geometrical investigations. The end product of his research, a seemingly endless string of definitions with hardly a theorem, must be regarded as disappointing. Yet if one asks "what is" questions – What is a line?, What is a continuum? – then one must expect essentialist answers. But, while definitions have a certain value in mathematics, no fruitful mathematical theory can consist entirely of them. Theorems and their proofs which relate definitions one to another, are much more important. Bolzano's theory is unquestionably lacking in these.

Bolzano returned to geometrical studies in the 1830s and 1840s. In writings of 1843 and 1844, though not published in his lifetime, he revised and improved his youthful findings. In these, Bolzano's topological basis, derived from his concept of "neighbour" and "isolated point", is very deep. The concept of neighbour, which in effect uses the modern notion of the boundary of a spherical neighbourhood allowed Bolzano to put forward some very clear definitions of line, surface, and solid. Later, when he discovered his notion of isolated point, he was able to arrive at an even deeper understanding of the basic figures of geometry. His geometrical insights were far more penetrating than those of his contemporaries.

3. Cantor's "paradox" of dimension

While Bolzano could be regarded as a precursor, there is little doubt that Georg Cantor is the true father of dimension theory. In 1877 Cantor discovered to his own amazement that the points of a unit line segment could be put into one-to-one correspondence with the points of a unit square or even more generally with the points of a q-dimensional

Georg Cantor (1845–1918)

cube. Cantor's probing led him to exclaim to his friend Richard Dedekind (1831–1916): "As much as you will not agree with me, I can only say: I see it but I do not believe it" [3, p. 44]. The strange result immediately called into question the very concept of dimension. Was it well-defined or even meaningful?

Cantor's work on set theory arose out of his investigations into the uniqueness of representing a function by a trigonometric series. In 1874 he published his first purely set-theoretic paper, giving proofs that the set of real algebraic numbers could be conceived in the form of an infinite sequence:

$$\omega_1, \omega_2, \ldots, \omega_\nu, \ldots .$$

This set is countable (*abzählbar*, to use Cantor's later term), while the set of all real numbers is uncountable and cannot be listed in this way. Through these results on "linear sets" Cantor saw a clear distinction between two types of infinite sets of numbers on the real number line.

As his correspondence with Dedekind shows, Cantor discovered these results in 1873. Cantor had met Dedekind by chance in Gersau during a trip to Switzerland in 1872 and their famous exchange of letters ensued [7, 8, 10, 23]. From his discoveries about linear sets of points it was perfectly natural for Cantor to wonder whether there were different types of infinite sets in the plane or in higher-dimensional spaces. In a letter to Dedekind dated 5 January 1874 he posed a tantalising new research question, a question which is basic to the growth of dimension theory:

> Can a surface (perhaps a square including its boundary) be put into one-to-one correspondence with a line (perhaps a straight line segment including its endpoints) so that to each point of the surface there corresponds a point of the line and conversely to each point of the line there corresponds a point of the surface? [17, p. 132]

From the start Cantor was convinced of the importance and difficulty of this question. He realised that most mathematicians would regard the impossibility of such a correspondence as so obvious as not to require proof. When he discussed it with a friend in Berlin during the first part of 1874, the friend explained that the matter was absurd "since it is obvious that two independent variables cannot be reduced to one" [17, p. 132]. In relating this encounter to Dedekind in a letter of 18 May, the young Cantor sought reassurance that he was not chasing a delusion! In posing his question Cantor introduced something quite new and important into thinking about dimension: he related mappings and correspondences to the dimension of figures and spaces. For Cantor this was a natural relation because he was interested in cardinality.

It is likely that Cantor only worked intermittently on this question from May 1874 until April 1877 and indeed without success. However, he persisted in regarding it as important. When he attended the *Gaussjubiläum* in Göttingen on 30 April 1877, the centenary of Gauss' birth, he told various colleagues (among them Heinrich Weber and Rudolph Lipschitz) of his problem, which he felt was fundamental. Again most thought the answer was obvious; a one-to-one correspondence between geometrical figures of differing dimensions is impossible. Still Cantor felt that *proof* was needed.

Subsequent to the Gauss Jubilee Cantor switched his line of attack. Instead of trying to prove that a one-to-one mapping could not exist he tried to construct one. This proved the key and in a letter to Dedekind of 20 June 1877 he presented a startling geometrical conclusion:

> that surfaces, solids, even continuous figures of ρ dimensions can be put into one-to-one correspondence with continuous lines, thus figures of only one dimension; therefore, that surfaces, solids, even figures of ρ dimensions have the same power as curves [17, p. 133].

Immediately Cantor saw his result on the equal power or cardinality of sets of various dimensions as a criticism of the assumptions about dimension commonly held by the geometers of the time. Dimension was not a problem for them, and many were used to basing their investigations on intuition. They casually spoke of simply infinite, twofold, threefold, $\dots$, ρ-fold infinite figures; they even regarded the infinity of points of a surface as the square of the infinity of points of a line or the infinity of points of a solid as the cube of the infinite set of points of the line. Cantor's result amounted to an attack on these "foundations" of geometry, on the very concept of dimension they were using uncritically.

In his letter to Dedekind of 20 June Cantor attempted an arithmetical demonstration of his results in which points of the ρ-cube an be put into one-to-one correspondence with values of a variable in the closed unit interval. The proof rested on his assertion that every number on the unit interval can be represented uniquely in the form of an *infinite* decimal expansion:

$$x_i = \alpha_{i,1}\frac{1}{10} + \alpha_{i,2}\frac{1}{10^2} + \cdots + \alpha_{i,\nu}\frac{1}{10^\nu} + \cdots .$$

To a system $(x_1, x_2, \ldots, x_\rho)$, which represents a point of the ρ-cube, a corresponding point y can be obtained by "interlacing" decimals in the individual x_i's, in blocks:

$$y = 0.\alpha_{1,1}\alpha_{2,1}\ldots\alpha_{\rho,1}\cdot\alpha_{1,2}\alpha_{2,2}\ldots\alpha_{\rho,2}\cdot\alpha_{1,3}\alpha_{2,3}\ldots\alpha_{\rho,3}\cdot\ \ldots .$$

Conversely, any number y can be "unlaced" to give a system $(x_1, x_2, \ldots, x_\rho)$, a point of the unit ρ-cube.

Dedekind immediately responded with an objection to Cantor's proof, since the "unlacing" of a point on the interval might produce finite decimal expansions (such as $x_2 = 0.73000\ldots$ from such a y value as $y = 0.478310507090\ldots$). Dedekind concluded: "I do not know whether my objection is of essential significance for your idea; however, I did not want to hold it back from you". Cantor's postcard reply was swift. He accepted Dedekind's objection as a criticism of the proof but not of the theorem itself. Cantor felt his result could be salvaged and his immediate reaction was to claim that he had proved more than he intended.[2] In a letter sent two days after sending his postcard, he presented Dedekind with an entirely new proof. This overcame Dedekind's objection but was not nearly as simple as the earlier one. In his new proof Cantor used the fact that every irrational number between 0 and 1 can be represented by a unique infinite *continued fraction*, a representation he used in preference to the original decimal expansion.

Having made his paradoxical discoveries Cantor was quick to draw out the mathematical and philosophical consequences. In the last paragraphs of his letter of 25 June 1877, he remarked on his interest in the efforts of Gauss, Riemann, Helmholtz, and others directed towards understanding the foundations of geometry, but his result had now made him doubt the validity of their work:

> It strikes me that all investigations taken up in this field begin for their part from an unproved assumption which does not appear to me to be obvious, but rather seems to need a proof. I mean the assumption that a q-fold extended continuous manifold requires q independent real coordinates for the determination of its element and that this number of coordinates can be neither increased nor decreased for one and the same manifold [17, p. 140].

Cantor ascribed his result of a q-fold manifold being "coordinated" by a single coordinate to the "wonderful power in the usual real and irrational numbers" [17, p. 14]. Cantor's letter to Dedekind of 25 June ended:

> Now it seems to me that all philosophical or mathematical deductions which make use of this mistaken assumption [on the number of coordinates] are inadmissible. Rather the distinction which exists between figures of different dimension numbers must be sought in entirely different aspects than in the number of independent coordinates, which is normally held to be characteristic [17, p. 141].

---

2 Cantor did not immediately repair his proof, though for us it is quite easy to fix up the argument. We merely block 0's with non-zero digits to eliminate the difficulty which Dedekind pointed out. Thus we block the value of y: 0.47831|05|07|09 .... We then treat blocks as single digits for interlacing and unlacing.

Prior to Cantor and Dedekind, Bernhard Riemann (1826–1866) and Hermann Helmholtz (1821–1894) among others, had put forward a very informal theory of continuous manifolds which included an implicit theory of dimension based on the number of coordinates. Riemann and Helmholtz merely intended their theory of manifolds as a general framework for their investigations into geometry. Objectively one can only regard their theory as vague in its mathematical details. As Cantor saw, his result was a direct and devastating blow to the "coordinate concept of dimension". Another concept of dimension was needed.

Cantor approached the concept of dimension from a position outside geometry. His point of view derived from his work on one-to-one correspondences and cardinality, and hence, from the viewpoint of the set theory which he was in the midst of creating. Consequently, he brought an entirely new set-theoretic approach to problems of geometry. Moreover, he recognised his result could easily be extended from q-dimensional manifolds to infinite-dimensional manifolds, assuming that their infinitely many dimensions have the form of a simple infinite sequence (i.e. the dimension is countably infinite). Apparently the editors of Crelle's *Journal* found Cantor's results bizarre. They wished to ignore or even reject his paper; to use Imre Lakatos' term they were monster-barrers [25]. After a frustrating wait, Cantor heard his paper would be printed.

Dedekind was not entirely persuaded by Cantor's claims of a revolution. While accepting Cantor's counterintuitive result whole-heartedly, he did not concur with Cantor's claims that the foundations of geometry were being undermined. He immediately saw a way out of the difficulty through continuity and gave credit to the older geometers for this means of escape from the consequences of Cantor's result. To a certain extent it is true that Riemann and Helmholtz included continuity (and differentiability) in their concept of dimension. However, Dedekind probably imputed a little too much to their informal theory in his desire to find "hidden lemmas" in the work of the great men of the past.

A fresh examination of the vague informal concept of dimension was now an absolute necessity. Even Dedekind realised this and he quickly saw that a proof of some kind of theorem about dimensional invariance incorporating the idea of continuity was needed. Hence, he came to state very clearly the crucial problem of dimension which the paradoxical correspondence result forced upon mathematicians. In a letter to Cantor of 2 July 1877, Dedekind arrived at the following statement, in effect a proposed invariance theorem:

> If one succeeds in setting up a one-to-one and complete correspondence between the points of a continuous manifold A of a dimensions on the one hand and the points of a continuous manifold B of b dimensions on the other, then this correspondence itself must necessarily be discontinuous throughout if a and b are unequal [17, pp. 141–142].

At once Cantor accepted that Dedekind's reading of the problem situation was superior to his own. However, he also recognised possible difficulties lurking in the background of the proposed dimensional invariance theorem. In fact, the subsequent history of the search for a proof of invariance of dimension has shown that the theorem was far easier to state than to prove. Nevertheless the sensitive spot in the problematic notion of dimension had been pinpointed.

4. The invariance of dimension

The publication of Cantor's "Contribution to the theory of manifolds" ("Ein Beitrag zur Mannigfaltigkeitslehre") towards the beginning of 1878 immediately caused a flurry of mathematical activity. The objective now was to save the concept of dimension; the paradoxical correspondence between manifolds of different "dimensions" had to be explained (one is tempted to say, explained away). During the months July to October 1878, when Cantor's paper had barely left the presses, five mathematicians attempted to demonstrate the invariance of dimension through a consideration of continuity – just as Dedekind had suggested in his letter to Cantor. These were Jakob Lüroth (1844–1910), Johannes Thomae (1840–1921), Enno Jürgens (1849–1907), Eugen Netto (1848–1919), as well as Cantor himself.

These early efforts towards showing dimensional invariance were only partially successful. The proofs, though interesting, are extremely complex. Lüroth and Jürgens (and also Cantor in his first work of 1878) only aimed at demonstrating invariance for low dimension numbers. Thomae, Netto, and Cantor also tried to prove dimensional invariance in the general case, but subsequent criticism revealed flaws in their proofs. Without doubt the greatest difficulty which faced all these mathematicians was the primitive state of "topology". The part of the subject which we now know to be most relevant, was virtually nonexistent. The offered proofs mainly used methods of real analysis (in particular, variations of the intermediate value theorem) and simple geometry.

In 1878, Jakob Lüroth was a young professor at the Technische Hochschule in Karlsruhe. He had been taught by Otto Hesse and Alfred Clebsch and, in his initial research following his doctorate, had made a contribution to invariant theory. He was the first off the mark on the subject of dimensional invariance when he presented a paper to the *Physikalisch-Medizinische Sozietät* in Erlangen on 8 July 1878.

At the same time Johannes Thomae, an *ordenlicher Professor* at Freiburg tried to prove the theorem in full generality. At a session of the 51st *Versammlung Deutscher Naturforscher und Aerzte,* held in Kassel in September 1878, Lüroth rightly criticised Thomae's "general proof" declaring that the separation property and invariance were in the same footing with respect to their importance and difficulty of proof. Though Thomae's proof was a failure, it is significant that he chose to place the problem firmly in the domain of *analysis situs*. At the same meeting, Enno Jürgens sketched an alternative proof for the 2-dimensional case. Jürgens gave a complicated rigorous proof but unfortunately the method could not be extended to give more general results.

Cantor was also at work. Several times he gently urged Dedekind to prove his conjecture on the invariance of dimension, but Dedekind did not respond. Cantor was not completely satisfied with the published proofs: "it seems to me, however", he wrote to Dedekind at the end of 1878, "that the situation is still not entirely resolved" [17, p. 157]. Eugen Netto's proof of dimensional invariance appeared last and interested Cantor most. It was published in Crelle's *Journal* at the end of 1878 and perhaps appealed because it was the only reasonably good attempt at a general proof. Netto, a student of Ernst Kummer, Leopold Kronecker, and Karl Weierstrass in Berlin, is best remembered for his contribution to group theory; his textbook *Substitutionentheorie und ihre Anwendung auf die Algebra* (1882) is a minor classic. Netto put his inductive proof of invariance squarely in the province of topology. Nevertheless, even this did not make Cantor feel that the problem was solved, as he admitted in a letter to Dedekind in early January 1879:

> No matter how commendable this penetrating attempt at a proof appears to me. I still cannot banish certain doubts about it and I fear that it is only an attempt, which nonetheless will certainly contribute to clarifying the situation [17, pp. 157–158].

In view of Cantor's lingering doubts he presented his own proof of dimensional invariance to Dedekind in a letter dated 17 January 1879. Cantor expressed the invariance theorem in the following terms:

> A continuous [i.e. connected] M_μ and a continuous M_ν, in case $\mu < \nu$, cannot be put into *continuous* correspondence with one another such that to each element of M_μ there belongs a single element of M_ν and to each element of M_ν there belongs one or more elements of M_μ [17, p. 158].

He claimed that he had a proof of this more than a year, but previously had serious doubts about its validity since it depended upon a multi-valued correspondence. Evidently he resolved these doubts, and the inductive proof he published was based on showing a contradiction to the intermediate value theorem.

By the 1880's most mathematicians thought that Cantor's paradox about dimension had been resolved by either Netto's or Cantor's proof of dimensional invariance [5, 17]. Enno Jürgens was virtually the only one to voice dissent, but his immediate and acute criticism of Netto's general proof was ignored. Cantor's invariance "theorem" is actually falsified by Peano's space filling curve. However this was not published until eleven years later and by that time the situation in topology and dimension theory had changed considerably. Among mathematicians the generally held opinion during the last two decades of the nineteenth century was that the invariance of dimension had been rigorously established.

5. The rise of point set topology

The rise of the theory of sets and particularly the theory of sets of points (*Punktmannigfaltigkeitslehre*) was due primarily to the creative work of Georg Cantor. His point set theory added an entirely new perspective to topological thinking. Cantor published his most important investigations into the theory of sets of points in a series of six papers entitled "Über unendliche, lineare Punktmannichfaltigkeiten" in the *Mathematische Annalen* during (1879–1883). These brilliant papers constitute the "quintessence of Cantor's lifework", as Ernst Zermelo (1871–1953) later declared [17, p. 163].

In Cantor's study of linear point sets and point sets in the n-dimensional arithmetic continuum (Euclidean n-space), the fundamental concept is that of limit point. The underlying theorem is the so-called Bolzano–Weierstrass theorem (that every infinite set of points in a bounded region of n-space possesses at least one limit point). From this fundamental concept and theorem flow all of Cantor's deep point set-theoretic concepts: the notions of a derived set, an everywhere dense set, an isolated point set, nowhere dense set and the Cantor "middle-third" discontinuum. By introducing these notions Cantor opened up a new field of study. Analysts were the first to see the usefulness of Cantor's imaginative ideas. Point set theory offered wonderful new instruments for a detailed study of the nature of functions, with the result that the growth of real and complex function theory was greatly accelerated in the years after the publication of Cantor's great papers. Applications of the Cantorian toolkit to the fundamental notions of geometry came a little later.

Giuseppe Peano (1858–1932) and Camille Jordan (1838–1921) were among the first to take the Cantorian ideas into the domain of geometry, so their work became a vital background to the development of point set topology. Peano and Jordan separately attacked the problems of measure and integration, but their results were remarkably close. From the point of view of the development of topology both saw the importance of applying Cantorian set theory to intuitive geometrical ideas.

Among topologists Camille Jordan is best known for his celebrated theorem on closed curves in the plane. His first enunciation of it appeared in a "Note" at the end of the first edition of his *Cours d'Analyse*. The second edition of the *Cours* was published between 1893 and 1896. Jordan incorporated the material of the earlier "Note" into the main text of the first volume and included much research material. While previous mathematicians hardly saw the need to prove the obviously true statement that a simple closed curve divides the plane into an inside region and an outside region, Jordan required a proof. In a lengthy one, Jordan considered continuous closed curves without multiple points, *i.e.*, *simple* closed curves, and so stated the theorem:

> Every continuous curve C divides the plane into two regions, the one exterior, the other interior; the latter cannot be reduced to zero, because it contains a circle of finite radius [17, p. 168].

Peano and Jordan were in many ways mathematical rivals. Peano's special contribution to the rise of point set topology, published in 1890, was the spectacular example of a space-filling curve, a curve which covers all the points of a square [29]. The ingeniously constructed curve whereby the points $(x(t), y(t))$ trace out and fill the unit square has t varies along the unit interval, completely upset the geometrical intuitions of mathematicians. Curiously Peano's construction of the continuous functions $x(t)$, $y(t)$ was published without diagrams, as a guard against error [22].

It is not known precisely how Peano came to devise such a spectacular curve, but we can see that much of his earlier work of the 1880's was directed towards a critical appraisal of commonly-held mathematical notions. His analysis text *Differential Calculus and fundamentals of integral calculus* (*Calcolo differenziale e principii di calcolo integrale*, 1884) abounds with examples demonstrating the need to revise the fundamentals of the subject then taken for granted. In his *Geometrical applications of the infinitesimal calculus* (*Applicazioni geometriche del calcolo infinitesimale*, 1887) Peano devoted a chapter to the study of geometrical magnitudes, giving definitions of the interior and exterior measure of linear sets, plane areas, and spatial volumes. He gave careful definitions of such conceptual ideas as interior points, exterior points, boundary points. Taking a set-theoretic viewpoint proved fruitful and led him to counterintuitive speculations and constructions.

Given the wealth of results from the papers of Cantor, Jordan, Peano, and many others, certain mathematicians recognised the need for an entire programme of exploration in set-theoretic topology. At the *First International Congress of Mathematicians*, held in Zurich from 9 to 11 August 1897, Adolf Hurwitz (1859–1919) sketched a plan of investigation [15]. In his lecture reviewing progress in analytic function theory Hurwitz attempted to determine the precise domain of validity of the Cauchy integral theorem. In so doing, he referred to the important work of Jordan who had scrutinized closed curves in the context of this analytic theorem. Hurwitz asked highly relevant general questions, questions reminiscent of Bolzano's essentialist forays:

> What is a simple closed line, what is a line, especially, a closed line in general, and are all or only some closed lines admissible in the enunciation of the Cauchy theorem? [17, p. 173].

In attempting to grapple with these questions he put them into the general context of the topology of closed sets. In a way his treatment is similar to Cantor's definition of cardinals and ordinals. Closed point sets can be assigned to classes, each class containing those sets which can be mapped one-to-one continuously onto one another. The sets in each class were to be called "equivalent". Then Hurwitz set the goal for the subject:

> This distribution of point sets into classes forms ... the most general foundation of analysis situs. The task of analysis situs is to search for the invariants of the single classes of point sets [17, p. 173].

Hurwitz gave a specific example: Jordan's simple closed curve could be regarded as a planar point set which belonged to the class containing the boundary of a square. In essence, Hurwitz stated Felix Klein's famous *Erlanger Programm* for *analysis situs* in a sharp form by placing the central problem of finding topological invariants into the framework of point set theory. Hurwitz himself never contributed to the programme he outlined in Zurich.

Arthur Schoenflies (1853–1928) embarked on the programme outlined by Hurwitz. Schoenflies studied at Berlin under Kummer in the period 1870–1875 and made his early mark in crystallography. Relying on the work of Cantor, Jordan, Peano, and others, Schoenflies proposed and partly developed a special programme of set-theoretic topology. At the turn of the century Schoenflies, a brilliant lecturer, was the foremost propagandist for Cantor's set theory. He wrote the article on "Mengenlehre" for the *Encyklopädie der mathematischen Wissenschaften* (1898) and composed reports on point set theory in response to a commission by the *Deutsche Mathematiker-Vereinigung*. In several publications Schoenflies intended to characterize the topology of the plane. Central to his planar topology was the Jordan curve theorem and its converse, and indeed, he was the first to state and prove a converse. This asserted that a closed curve dividing the plane into two domains, such that each of its points is *accessible* from both domains by simple paths, is homeomorphic to a circle.

Set theory became the main breeding ground for pathological geometric examples. In addition to Peano's example, two other striking entries in this list of curves which run counter to "naive geometrical intuition" may be mentioned. At the beginning of the century, William Osgood (1864–1943) and Henri Lebesgue (1875–1941) each published an example of a simple closed curve possessing a *positive* exterior measure – the bizarre result that a curve could have a measurable area. This curve is even stranger than the one which Peano suggested, because, unlike Peano's, it has no multiple points. An early catalogue of strange curves is contained in the first book on set theory in English, William H. and Grace Chisholm Young's *Theory of Sets of Points* (1906). For mathematicians well trained in Cantorian methods the topological monsters became manageable and quite familiar. They were repugnant to those who wished to pursue "normal mathematics", yet the pathological oddities were an important spur to growth in set-theoretic topology.

6. New approaches to dimension and invariance

During the first decade of the twentieth century some prominent mathematicians conceived a variety of new ideas about dimension. Among those contributing were René Baire (1874–1932), Maurice Fréchet (1878–1973) and the Hungarian Frigyes Riesz (1880–1956). These ideas were intended not just as suggestions for solving the invariance problem but more generally as novel ways of looking at the notion of dimension itself.

Among the new ideas the most interesting came from Henri Poincaré (1854–1912). As one of the giants who founded modern algebraic topology it would seem natural for Poincaré to have had an interest in the topological problem of dimension. Yet his concern with dimension did not arise from his mathematical work; rather it developed from his philosophical investigations into the origins and nature of our geometrical knowledge and into the relationship between geometry and space. He proposed a topological definition of dimension because he was grappling with cosmological or, more accurately, epistemological questions, not mathematical ones.

Poincaré's philosophy of space and geometry grew out of an examination of three main problems: first, the problem of explaining the applicability of the various geometries, Euclidean and non-Euclidean, to "our" space, second, the problem of explaining the origins of our fundamental ideas of space and geometry, and, third, the problem of explaining why we say our space has three dimensions. In the course of these investigations Poincaré constructed two theories of dimension. He developed the first theory in the 1890's based on group theory. Then in the first years of this century he proposed his second theory of dimension which is specifically topological and is the one that influenced the later development of mathematical dimension theory.

In a paper on the origins of our knowledge of space and its geometry, "L'Espace et la Géométrie" (1895) [30], Poincaré offered a brief explanation of the dimensionality of our space partly based on the theory of continuous groups due to Sophus Lie (1842–1899). This immediately drew criticism from the young Louis Couturat (1868–1914), once a student of his at the *École Normale Supérieure.* Couturat, who took a Kantian position opposed to Poincaré, bluntly accused him of arguing in a circle. Naturally Poincaré felt obliged to reply, at first briefly and then in more detail in a lengthy article published in English in 1898. In "On the Foundations of Geometry", Poincaré gave an exposition of his first theory of dimension [31].

The works of both Julius Plücker (1801–1868) and Sophus Lie strongly influenced Poincaré. In his essay of 1898 Poincaré whole-heartedly adopted the group-theoretic view of geometry. In this context he applied the Plückerian equivalence principle (that it is possible to construct an infinity of different but "equivalent" spaces by choosing different primary elements such as lines, planes, conics etc.), in order to develop a "relativistic" view of dimension in spatial geometry. If conics are taken to be elemental, for example, then the plane would be regarded as 5-dimensional.

In his essay of 1898 Poincaré was confident of the adequacy of his theory of dimension founded upon group theory for explaining the dimension of our space. From the mathematical standpoint his explanation is clever and also timely. Group theory at the turn of the twentieth century was increasingly seen as a unifying idea of various mathematical theories. However, from the philosophical standpoint his explanation is not satisfactory, since it is complicated and indirect. Explaining the 3-dimensionality of our space by means of a special representation of the Euclidean group of rigid motions acting on the conjugate space of rotation subgroups is surely a very roundabout way of explaining an apparently simple fact about physical space. Furthermore, though he explains the 3-dimensionality of space, he does not explain the concept of dimension itself. For these reasons we cannot regard Poincaré's first theory of dimension as adequate.

An event which altered Poincaré's view of the foundations of geometry was the publication of David Hilbert's *Grundlagen der Geometrie* (1899) [13]. Before this Poincaré was seemingly unaware of the developments in the pure logic of the foundations of geometry stemming from the investigations of Moritz Pasch (1843–1930), the Italian geometers, and David Hilbert (1862–1943). In his extended review of Hilbert's Grundlagen in 1902, he gave clear evidence of surprise at discovering the new approach to geometry's foundations. He seemed amazed at the extremely wide variety of geometries which are conceivable by using the logical or axiomatic-deductive method. Poincaré had grown up with the group-theoretic approach to geometry of Helmholtz, Klein, and above all Lie, but now the logical approach offered entirely new possibilities. He did not entirely give up the group-theoretic viewpoint but, nevertheless, in the last years of his career there was a discernible "retreat" from it. Yet in retreating from the group-theoretic approach, he did not thereby rush to adopt Hilbert's thoroughly logical approach, for he considered it to be too formal. Instead, he increasingly considered *analysis situs* or topology as the bedrock of geometry.

Poincaré described his theory of dimension based on group theory as a "dynamical theory". However, when investigating the notion of *place* as an empirical substitute for the mathematical concept of *point*, he put forward very tentatively a "statical theory" of dimension. In accordance with this general psycho-physiological theory is the recognition that each place in space can only be perceived inexactly – a place thus corresponds to a fuzzy area in sensible space. Our sensations of a place form something like a "wafer". We can imagine sensible space as a series or collection of overlapping wafers in which contiguous wafers are associated with one another.

Poincaré continued to struggle with the difficulties of constructing a "statical theory" of dimension in the years after the publication of his 1898 essay. The result was a new paper: "L'Espace et ses trois Dimensions" (1903), containing a novel explanation of the concept of dimension. Several years later he published a modified and extended version of the new dimension theory in an article on "Pourquoi l'Espace a trois Dimensions" (1912). These papers of 1903 and 1912 contain a substantial revision of philosophical theory of the genesis of spatial geometry [32, 33]. The centrepiece of the revised theory is his second theory of dimension. He distinguishes very clearly between the mainly mathematical problem of defining the dimension concept and the cosmological or epistemological problem of explaining why we attribute three dimensions to our space. The striking feature of the new theory for defining mathematical dimension is that it is a *topological theory*, based on a notion of a cut (*coupure*). This was outlined by Poincaré around 1904:

> If to *divide* a continuum it suffices to consider as cuts a certain number of elements all distinguishable from one another, we say that this continuum is of *one dimension*; if, on the contrary, to divide a continuum it is necessary to consider as cuts a system of elements themselves forming one or several continua, we shall say that this continuum is of *several dimensions*.
>
> If to divide a continuum C, cuts which form one or several continua of one dimension suffice, we shall say that C is a continuum of *two dimensions*; if cuts which form one or several continua of at most two dimensions suffice, we shall say that C is a continuum of *three dimensions*; and so on [4, p. 3].

Poincaré's thoughts on space occupy an unusual place in the modern history of dimension theory. His motive for attacking problems of dimension developed out of his philosophy rather than out of his mathematics and his two theories of dimension form integral parts of his philosophy of geometrical conventionalism. Poincaré accepted the idea that our physical space *is or ought to be* regarded as 3-dimensional. Unusual spaces under nonstandard interpretations are certainly possible, but they are not as convenient for our geometrical deliberations as the usual one. Ultimately the 3-dimensionality of our space rests upon a reasonable convention in line with inheritance and experience. Unlike most of his contemporaries who thought about dimension, Poincaré took no interest in the fundamental mathematical problem of dimensional invariance. In this case he was far more concerned with epistemological issues in the foundations of science and mathematics.

7. L.E.J. Brouwer

Luitzen Egbertus Jan Brouwer (1881–1966) was a dominant influence in the emergence of the new topology in the first years of the twentieth century [6]. Brouwer viewed topology in quite a different way from Poincaré. Operating within a framework of set theory and point set theory, Brouwer pushed topology to new limits with his theories of mappings, degree, and dimension. He produced his major papers between 1909 and 1913.

When Brouwer was beginning his career as a mathematician, set-theoretic topology was in a primitive state. Controversy surrounded Cantor's general set theory because of the set-theoretic paradoxes or contradictions. Point set theory was widely applied in analysis and somewhat less widely applied in geometry, but it did not have the character of a unified theory. There were some perceived benchmarks. For example, the generally held view that dimension was invariant under one-to-one continuous mappings, a view which was was echoed in Arthur Schoenflies' *Encyklopädie* article, "Mengenlehre", of 1898 [34].

Though Brouwer had already published papers on classical geometry, his doctoral thesis of 1907, *On the Foundations of Mathematics* (*Over de Grondslagen der Wiskunde*), marked the real beginning of his mathematical career. The work revealed the twin interests in mathematics that dominated his entire career: his fundamental concern with critically assessing the foundations of mathematics, which led to his creation of Intuitionism, and his deep interest in geometry, which led to his seminal work in topology [1, 2]. Brouwer quickly found that his philosophical ideas sparked controversy. D.J. Korteweg (1848–1941), his thesis supervisor, had not been pleased with the more philosophical aspects of the thesis and had even demanded that several parts of the original draft be cut from the final presentation [35]. Korteweg urged Brouwer to concentrate on more "respectable" mathematics, so that the young man might enhance his mathematical reputation and thus secure an academic career.

Brouwer was fiercely independent and did not follow in anybody's footsteps, but he apparently took his teacher's advice and set out to solve some really hard problems of mathematics. Brouwer put in a prodigious effort in these early years and rapidly produced a flood of papers on continuous group theory and topology – more than forty major papers in less than five years [36]. Brouwer's ambition can be judged by the problems he addressed, for example, a full-scale attack on Hilbert's fifth problem. This was one of twenty-three challenging problems outlined by Hilbert at the *International Congress of Mathematicians* in 1900. In the Fifth problem, Hilbert asked: "How far Lie's concept of continuous groups of transformations [of manifolds] is approachable in our investigations without the assumption of differentiability of the functions" [12, p. 12]. For Brouwer the problem was "to determine all finite continuous groups of an n-dimensional manifold", a problem that naturally followed from work in his doctoral thesis [19, p. 66]. It led him to study Arthur Schoenflies' *analysis situs.* In 1908, the year following his doctorate, Brouwer made an address to the *International Congress of Mathematicians* in Rome on the topological foundations of Lie groups.

In the summer of 1909, Brouwer discovered serious flaws in Schoenflies' second Report on point set theory, which had been published in the previous year. He wrote to Hilbert of its inadequacy: "I discovered all of a sudden that the Schoenfliesian investigations concerning the *analysis situs* of the plane, on which I had relied in the fullest way, could not be taken as correct in all parts, so that my group-theoretic results also became doubtful" [19, p. 67]. The Report was composed of all Schoenflies' work of five years standing, and Schoenflies believed he had produced a "complete manual" of planar topology – a beautiful model of mathematical rigour. By the time Brouwer had finished with it, the work of Schoenflies on *general* closed curves was left in ruins. Nearly all the theory of these curves suffered demolition. Practically the only result left intact was Schoenflies' converse of the Jordan curve theorem.

In particular, Brouwer brought his immense critical powers to bear on Schoenflies' definition of a closed curve as a bounded closed point set which divides the plane into two domains such that it is the common boundary of the two domains. In his ensuing paper "On *analysis situs*" Brouwer produced an example of a curve that divided the plane into three domains and is the boundary of each of them. Brouwer's curve made an impression and enhanced his reputation in the mathematical community. Brouwer's letter to Korteweg, written on 18 June 1909, shows some satisfaction in his triumph: "At last some fish has taken the bait! ... Schoenflies has gone into my paper in considerable detail, but I had to put the thumbscrews on rather hard" [19, p. 68]. Brouwer's paper exposed the weaknesses of Schoenflies' analysis, but from the point of view of the history of science, Schoenflies provided the raw material from which Brouwer could start. Brouwer was in a different mathematical situation from Cantor, who, thirty years before, worked from nothing.

On 12 October 1909, the twenty-eight year old Brouwer delivered his inaugural lecture as *privaat-docent* in the University of Amsterdam. In the lecture Brouwer outlined his programme for his research in topology. His lecture, "The Nature of Geometry" ("Het Wezen der Meetkunde") was similar in style to Klein's *Erlanger Programm*. Brouwer study of groups of one-to-one continuous transformations and sets of points in the plane and in higher-dimensional spaces, led him to *analysis situs* – the fledgling topology. From the need to classify the groups of one-to-one continuous transformations, Brouwer was led to hard problems in topology. One of them, which he alluded to in his lecture was to see "how far spaces of different dimension are different for our group [of transformations]. Most probably this is always the case, but it seems extremely hard to prove, and probably it will remain an unsolved problem for a long time to come" [18]. Brouwer was fascinated by the problem of dimensional invariance. Schoenflies, who had declared the problem closed in the first Report on set theory (1900), had reversed his belief and declared it open in his Second report (1908). Was it really possible for there to exist a *one-to-one and continuous* function from an m-dimensional domain to an n-dimensional domain with $m \neq n$? Brouwer sought an answer.

Brower spent the Christmas period 1909 in Paris. He stayed with his geologist brother, Aldert, at 6 rue de l'Abbé de l'Epée in the Latin Quarter. From this address he drafted a letter to Hilbert on "Neujahrsmorgen 1910" outlining the essential ideas of his pathbreaking topology papers to be published between 1911 and 1913. Remarking on the wonderful array of mathematical techniques discovered by Brouwer, Hans Freudenthal (1905–1990), saw fit to describe the place as the "cradle of modern topology" [9]. During his Christmas visit Brouwer made contact with Poincaré, Jacques Hadamard (1865–1963), and Emile Borel (1871–1956).

By 1910 the problem of dimensional invariance was a pressing one and Brouwer vigorously attacked it. By March 1910 he was able to inform Hilbert that he had arrived at a partial solution, a solution directly related to the topological distinction between spheres of odd and spheres of even dimensions and their transformations:

> I am preparing an article for submission to the *Annalen* editorial board in which I solve the invariance problem of dimension insofar as I show that at least spaces of even and odd dimension number cannot be mapped one-to-one continuously onto one another [18, p. 147].

The article submitted in June was the momentous "Proof of the invariance of the dimension number" ("Beweis der Invarianz der Dimensionenzahl"). It contained a more general proof of dimensional invariance. In the following month it was followed with a longer paper, also a masterpiece, "On the mapping of manifolds" ("Über Abbildung von Mannigfaltigkeiten"). Of these two revolutionary papers, the "Beweis" used his newly discovered "degree of a mapping" concept implicitly for the proof of dimensional invariance. Brouwer initially developed the concept of mapping degree for n-dimensional generalizations of the singularity theorem for vector fields on spheres and the fixed point theorem for mappings of spheres. The second paper spelled out the full details of the concept and included some important further consequences.

The "Beweis" – a mere five printed pages – effectively swept away all previous attempts to prove dimensional invariance. Brouwer's approach to the invariance problem depended on a lemma. This stated that the image of the unit cube, under a continuous mapping into itself which has the property of displacing every point less than half a unit, has an interior point. The proof of this lemma is firmly based on the idea of mapping degree and also on the auxiliary ideas of simplicial decomposition and the simplicial approximation of a mapping. Brouwer developed a proof full of geometrical insight. From the lemma it was relatively easy for him to prove the invariance of dimension from the two theorems [18]:

THEOREM 1. *An m-dimensional manifold cannot contain the one-to-one continuous image of a domain of higher dimension number.*

THEOREM 2. *In an m-dimensional manifold the one-to-one continuous image of a domain of lower dimension number is a nowhere dense point set.*

Thus topological mappings of manifolds can neither lower nor raise the dimension numbers. Looking back, Freudenthal concluded that the paper marked "the paradigm of an entirely new and highly promising method, now known as algebraic topology. It exhibits the ideas of simplicial mapping, barycentric extension, simplicial approximation, small modification, and, implicitly, the mapping degree and its invariance under homotopic change, and the concept of homotopy class" [2, p. 436]. It was simply "witchcraft" [2, p. xii].

Brouwer's landmark "Beweis" appeared in the same volume of the *Mathematische Annalen* as a paper by Henri Lebesgue ostensibly proving the same theorem. On a visit to Paris in the summer of 1910, Otto Blumenthal (1876–1944), the managing editor of the *Annalen* had met Lebesgue, and informed him that Brouwer had proved the invariance theorem. Lebesgue seemed unmoved and replied that he had several proofs. Blumenthal was so impressed with the slick proof he received from the great French mathematician that he decided to publish it alongside Brouwer's more difficult proof. Lebesgue's justification of dimensional invariance was based on an unfounded tiling principle:

> If an n-dimensional cube is covered by [sufficiently] small closed pieces, there is a system of $n + 1$ among them with a non-empty intersection and such that *all* systems of $n + 2$ members have empty intersection [9, p. 501].

The highly competitive Brouwer took exception to sharing the limelight with the Frenchman. To submit an unfounded proof was not Brouwer's way, and he repeatedly challenged Lebesgue to supply a rigorous proof of the tiling principle. A bitter dispute broke out between the two mathematicians of a kind which may occur when sensitive matters of priority are involved. Brouwer broke off relations with Lebesgue but kept up pressure on the Frenchman in the pages of mathematical journals "encouraging" him to provide a proof, but to no avail. To add to his irritation, Brouwer recognised the value of the tiling principle and tried to prove it himself. While waiting for Lebesgue, and after a few failed attempts of his own, Brouwer achieved a proof.

As a finale to his great period of research in topology Brouwer produced a second paper on the concept of dimension and dimensional invariance, the monumental "On the natural concept of dimension" ("Über den naturlichen Dimensionsbegriff", 1913). In this he critically examined a definition of the concept of dimension proposed by Poincaré and offered one of his own, his "Dimensiongrad". Brouwer found Poincaré's definition of dimension unsatisfactory. It assigned dimension numbers at variance with the "natural dimension", the dimension one would expect for fairly standard geometrical objects. According to Poincaré's definition the double cone was of dimension 1 since it could be cut into two pieces by the removal of a single point, whereas, intuitively, the double cone "ought" to be of dimension 2.

Brouwer's definition of dimension was more sophisticated than Poincaré's. He made early use of Fréchet's ideas on abstract spaces and, unlike Poincaré, he gave a mathematical definition of what is meant by a *continuum.* In Brouwer's scheme this was a prerequisite for the definition of dimension. Brouwer's inductive definition was intrinsic, that is, based on the internal properties of the sets under consideration and not on their situations in larger spaces. In his 1913 paper, he then proved dimensional invariance on the basis of his definition and gave his rigorous demonstration of Lebesgue's tiling principle. This paper was a highpoint in Brouwer's career as a topologist and he solved very difficult problems. He made no attempt to provide a formal theory which would serve to place his various results in relation to each other in a general setting. Formal theories were for others to create. It was an activity he came to view with contempt. Following the publication of this paper of 1913, Brouwer's interest shifted and he turned away from being a mainstream mathematician. He devoted more and more of his time to the exploration of the foundation of mathematics and intuitionism.

8. The emergence of a theory

Around the time of the publication of Brouwer's major papers and in the succeeding years point set theory and set-theoretic topology grew rapidly. Many were adding new ideas and results to the expanding field. A striking sign of these rapid developments was the publication of the classic text *Fundamentals of Set Theory* (*Grundzüge der Mengenlehre*, 1914) by Felix Hausdorff (1868–1942). In this book Hausdorff offered a synthesis of the diverse results of point set theory through his theory of topological spaces. It joined together the fundamental geometrical ideas and abstract analytical concepts of David Hilbert, Maurice Fréchet, and Hermann Weyl (1885–1955).

In contrast to Brouwer's difficult papers, Hausdorff's text was eminently readable. Of particular significance to dimension theory was the fact that mathematicians still regarded the quest for definitions started by Bolzano as still relevant. But the appearance of the topological monsters, created to put putative definitions to the test, caused Hausdorff to remark pessimistically: "We give no definition of the concept of curve; the sets which carry this name by convention are of such a heterogeneous nature that they fall under no reasonable collective concept" [18, p. 225]. While everyone has a good intuitive idea of the curve concept, no one had been able to define the general concept adequately. For example, Jordan's notion of a continuous curve was too wide, because it includes the Peano space-filling curves. The notion of a continuous simple arc, due to Nels Lennes (1874–1954) and presented in 1906, was too narrow, since it excluded many curve-like figures [26].

In the period following the Great War (1914–1918), interest in topology became vigorous. Brouwer concentrated his attention on Intuitionism but his reputation in the field of topology drew many visitors to Amsterdam. An American school of topologists produced significant papers, notably those by J.W. Alexander (1888–1971) and the Russian-born Solomon Lefschetz (1884–1972). A Polish school of set theory and topology was founded by Zygmunt Janiszewski (1888–1920), Waclaw Sierpinski (1882–1969), and Stefan Mazurkiewicz (1888–1945). These three, together with their students Bronislaw Knaster (1893–1980) and Casimir Kuratowski (1896–1980), made significant advances on topological problems which only later became relevant to dimension-theoretic questions. Through the medium of the principal publishing organ of the school, *Fundamenta Mathematicae*, started in 1920, many research papers appeared connected with Brouwer's topological work.

Hausdorff published a valuable definition of metric dimension in 1919 [11]. Starting from a generalisation of Lebesgue measure due to Constantin Carathéodory (1873–1950), Hausdorff offered a measure-theoretic characterisation of dimension for finite-dimensional Euclidean spaces. The Hausdorff theory of dimension is metrical rather than topological and its origins can be traced to Weierstrassian investigations into continuous curves without derivatives in the 1870s [4]. It assigns fractional as well as integral dimension numbers to point sets. The linear "middle-third" set, introduced by Cantor in 1883, has fractional dimension $\log 2/\log 3 = 0.6309\ldots$. Subsequent to the original work of Hausdorff many mathematicians have devoted extensive investigations to the metrical theory of dimension. Initially unaware of Hausdorff's earlier work Georges Bouligand (1889–1979) recreated the metrical theory of dimension during the 1920s. There is a provable connection between Hausdorff dimension and the topological theory of dimension of Urysohn and Menger [14].

In 1921 Henri Lebesgue finally published an adequate proof of the tiling principle along with its immediate consequence, a proof of dimensional invariance. Thus some ten years after the dispute with Brouwer he at last put a satisfactory proof of the far-reaching principle into print. Lebesgue's efforts in topology were of limited success and though they were no doubt brilliant, other mathematicians were needed to bring his insight to fruition. Later workers are often able to focus on pivotal concepts and achieve crisp formulations and polished proofs. Much better proofs of the tiling principle were produced by Emanuel Sperner (1905–1980) and Witold Hurewicz (1904–1956). The independent

From left to right: P. Alexandroff, L.E.J. Brouwer and P. Urysohn.
(Courtesy of Alexandroff Archive.)

proofs of Sperner (1928) and Hurewicz (1929) are extremely elegant and are probably the simplest conceivable, and practically reduce dimensional invariance to a topological triviality.

In the years after the publication of Brouwer's papers only a few mathematicians showed any interest in the specific problems of topological dimension. Indeed Brouwer's second paper of 1913 was hardly noticed at all. Many regarded Brouwer's proofs as difficult, particularly as they depended upon his special concepts of combinatorial topology.

In 1921 the topological theory of dimension again became a lively topic of research amongst the new generation of mathematicians. In that year two young mathematicians, unknown to one another, Pavel Urysohn (1898–1924) in Moscow and Karl Menger (1902–1985) in Vienna, began investigating these problems. A vigorous growth of the theory of dimension began, and during the 1920s and 1930s many contributed to the subject. The prime impetus for the growth of dimension theory during the 1920s and 1930s was provided by the definitions and results of Urysohn and Menger. They conceived their definitions independently, but soon others proved them to be equivalent. From their definitions sprang the full theory of dimension, which quickly attracted the attention of many topologists.

Pavel Urysohn entered Moscow University in 1915 with the intention of studying physics. However, he was soon drawn to mathematics, because he was charmed by the lectures and personalities of Nikolai Nikolaievich Luzin (1883–1950) and Dimitrii Fedorovich Egorov (1869–1931). Luzin was a dynamic mathematician and it was he who persuaded Urysohn to stay on in order to study for a doctorate during 1919–1921. Initially Urysohn's mathematical research was in analysis, but he switched to topology in 1921. In

June of that year he was awarded his doctorate and was then appointed as lecturer at the university. Egorov proposed two research problems for him. The first one was to give an intrinsic topological definition of line or curve in general which would be equivalent in the case of restriction to the plane to the known nonintrinsic definition of a Cantorian line, a continuum nowhere dense in the plane. The second one was to give an intrinsic topological definition encompassing a wide class of sets which could appropriately be regarded as surfaces. They were well-known problems, but they immediately fixed the imagination of the young man. In essence they were the definitional problems which had motivated Bolzano a century previously, but now they were posed as questions in set theory.

Without delay Urysohn started to think intensely about these definitional problems and his efforts soon led him to search for a "real" definition of the concept of dimension. During the summer of 1921 he proposed and then discarded one definition after another, giving examples and counterexamples to test his tentative proposals. Finally, on a morning near the end of August 1921, Urysohn woke up with a satisfactory definition in mind, one which he could finally accept after two months of concentrated thought. At the time he was on holiday with a group of young Moscow mathematicians in the village of Burkov (near Bolshev) on the banks of the Klyazmy. He immediately told his friend, Pavel Alexandroff (1896–1982), about his programme for transforming his conjectures into theorems.

During the subsequent academic year 1921–1922 Urysohn successfully proved the theorems of his new dimension theory. He gave a course of lectures on the topology of continua and used this platform to announce his new results, often just after he had proved them. Simultaneously he announced his results in a series of notes to the *Moscow Mathematical Society*. By the spring of 1922 his dimension theory was more or less complete and in September, Lebesgue presented Urysohn's theory to the *Académie des Sciences* in Paris. It was not until the academic year 1922–1923 that Urysohn wrote a full version of his dimension theory in his "Mémoire sur les multiplicités Cantoriennes I". At the beginning of the "Mémoire" Urysohn stated his main problem: "To indicate the most general sets that still merit being called "lines", "surfaces"" [18, p. 228].

Urysohn intended his theory to encompass more than the pair of problems which Egorov had set him originally. Not only did he seek definitions of curve and surface, but also definitions of n-dimensional Cantorian manifold and hence of dimension itself. The dimension concept was, in fact, the centre of his attention. Urysohn laid down three methodological principles. Firstly, he desired that all his definitions be intrinsic. Secondly, he attempted to provide *local* definitions wherever possible rather than integral or global ones. Indeed the local character of Urysohn's dimension definition distinguished it from Brouwer's 1913 definition of global dimension and it was this local character which made it particularly fruitful. Thirdly, Urysohn suggested that even though closed sets were his principal concern it should be possible to examine their nonclosed parts. Finally, and this gives Urysohn's work a modernity, he placed the theory in the framework of the compact metric space.

Urysohn had noticed that *compactness* was the pivotal assumption in most of his arguments. Hence, it seemed natural to him to use the compact metric space as a base for his work rather than the more concrete Euclidean spaces. Indeed, he boldly proclaimed the compact metric space to be the "natural domain of existence" ("domaine naturel d'existence") for intrinsic topology, not realising that even natural domains change in the course of history and are stretched to fit the current demands of mathematical research. In the case of dimension theory Urysohn's natural domain was soon expanded to the concept of the separable metric space by other dimension theorists of the 1920s.

Urysohn was highly successful in his enterprise. He did not work alone but often collaborated with Alexandroff, and their joint work produced a great extension and deepening of the abstract theory of metric spaces. Tragically Urysohn died on 17 August 1924 while swimming in rough seas off the coast of Brittany. Considering that he only had three years to devote to topology, he made his mark in his chosen field with brilliance and passion. He transformed the subject into a rich domain of modern mathematics. How much more might there have been, had he not died so young?

Karl Menger began thinking about the definitional problems for curves and dimension in the very same year as Urysohn: 1921. Menger, the son of the well-known Austrian economist, Carl Menger, had entered Vienna University as a student during the autumn of 1920. Like Urysohn he planned to study physics. About this time Hans Hahn (1879–1934) was called to Vienna and arrived there near the beginning of 1921. His teaching began with the announcement of a seminar on the problems associated with curves: "Neueres über den Kurvenbegriff".

At the urging of his close friend Otto Schreier (1901–1929) Menger decided to attend the first session of the seminar. At this session Hahn formulated the basic definitional problem concerning curves and introduced the first concepts of point set theory. Before the War, Hausdorff had declared the definitional problem for curves to be important but unyielding. In his seminar, Hahn in effect elaborated this challenge and declared the importance of solving the open problem of defining the curve concept generally and adequately. Menger left the seminar "in a daze", but with mounting enthusiasm he started to think about this basic problem of geometry. With a freshness of vision he boldly attacked the problem using an intuitive approach, undeterred by his lack of experience in mathematical research. After a week of intense thought he found a solution which he presented to Hahn before the next session of the seminar. Hahn was encouraging, and Menger continued to work out his ideas.

By June 1921 Menger had produced a short paper of three and one-half pages entitled "Der Begriff der Kurve". Although it contains only an informal attempt to define the concept of a curve embedded in Euclidean 3-space, and it was never published, the manuscript reveals the germ of Menger's later ideas on curve theory and implicitly on dimension theory. According to Menger the most important property of a curve is its 1-dimensionality. This view led him to call every closed connected 1-dimensional point set in 3-space a "curve in the wider sense" ("Kurve im weiteren Sinne"), where a set is 1-dimensional provided each of its points has arbitrarily small neighbourhoods with boundaries, each one of which intersects the set in finitely many points (or perhaps "countably many") – a *regular* curve in Menger's later conceptual framework. Menger continued to pursue his investigations but serious lung disease forced him to be away from Vienna for long periods and this inevitably slowed his progress. He returned to Vienna in April 1923 to prepare papers on dimension theory and complete his doctoral thesis. In June 1924 he completed his doctorate and, by the beginning of October 1924, he submitted his important paper "Über die Dimension von Punktmengen II" for publication. Brouwer and he were in correspondence by this time, and Menger's results were transmitted to the Dutch Academy in Amsterdam. Menger's paper "Über die Dimension von Punktmengen II" was eventually published in 1926 and comprised a fairly complete exposition of his dimension theory.

If we compare their principal early works on dimension, Urysohn's posthumous "Mémoire" and Menger's "Über die Dimension von Punktmengen II", Urysohn's paper is more complete and more polished. However, Menger's work has its own distinctive merits. It is a concise presentation of many significant results. Most importantly, Menger's fundamental inductive definition of dimension appeals more immediately to our geometrical intuition and leads more directly to an elegant theory than Urysohn's equivalent. Overall, the two mathematicians built a solid foundation for the subsequent development of the theory of dimension.

9. In conclusion

The twin problems of defining dimension and proving its invariance were the primary influences on the creation and growth of modern topological dimension theory. However, the cosmological problem of explaining the dimension number of physical space also was a secondary motive.

The history of modern dimension theory reveals a wide variety of definitions and theories. No single definition and theory can be regarded as uniquely correct and, contrary to the beliefs of the earliest workers, we cannot expect a single definition of dimension to reveal the "true essence" of the concept. The search for the holy grail, the single, universally acceptable definition has proved illusory. No single theory of dimension is exclusively at the centre of the mathematical stage for all time. In modern times three definitions of topological dimension are regarded as important, not counting the metrical dimension (Hausdorff/Besicovitch) which has been significant over the last twenty years in its connection with fractals. These are the "small inductive dimension" (Menger–Urysohn), the "large inductive dimension" (Brouwer–Cech) and the "covering dimension" (Cech–Lebesgue) [21].

Brouwer was perhaps the pivotal figure in the development of topology in the twentieth century. After 1913, Brouwer's contributions to topology were few but he still remained an authoritative presence. During the period 1925–1926, Alexandroff, Menger, and the Austrian Leopold Vietoris (b. 1891) visited Brouwer in Amsterdam (Vietoris is now the grand old man of Austrian mathematics). Through Menger, Witold Hurewicz became a mathematical assistant to Brouwer, and Hans Freudenthal, attracted by Brouwer's philosophical work, became another. Brouwer also influenced the American mathematician J.W. Alexander. In 1922 Alexander generalised the Jordan curve to higher dimensions, a result now known as Alexander duality. Brouwer inspired Erhard Schmidt (1876–1959), and through Schmidt's lectures, Heinz Hopf (1894–1971) was drawn into the subject. With such strong personalities as Brouwer, Menger, and Alexandroff involved, the sweet reasonableness of the earlier Cantor–Dedekind dialogue proved a rarity in the history of dimension theory. In 1926, a dispute broke out between Menger and Alexandroff, the Russian mathematician acting as guardian of the intellectual estate of his recently deceased friend Pavel Urysohn. Two years later Brouwer took exception to a passage in Menger's influential book *Dimensiontheorie* (1928) and another furious row ensued.

A history of dimension theory after the 1920s, after it became increasingly abstract and axiomatic, is outside the scope of this survey. Dimension theory as a collection of mathematical theories has grown rapidly during the twentieth century – a brief outline of technical developments in the 1920s and 1930s is given in [21]. In many ways its growth illustrates the kinds of development which mathematical theories often follow. Principally it advanced through the efforts of mathematicians to solve significant problems. New directions opened up, and the theories became more abstract. For instance, in 1932 Alexandroff extended the theory towards homology theory and algebraic topology in his paper "Di-

Hurewicz and Brouwer

mensionstheorie". Other mathematicians contributed to dimension theory, most notably L.A. Tumarkin (b. 1904), Lev Pontryagin (1908–1988), Georg Nöbeling (b. 1907) and Eduard Cech (1893–1960). In 1941 Witold Hurewicz and Henry Wallman (1915–1992), in dealing very succinctly with separable metric topological spaces set a standard for exposition in their classic *Dimension Theory* [14].

At the beginning of this article it was noted that space was generally accepted as having three dimensions and this notion was relatively unproblematic. The opening lines of W. Hurewicz and H. Wallman's *Dimension Theory* suggests an appropriate note on which to end [14, p. 3]:

> Of all the theorems of analysis situs. the most important is that which we express by saying that space has three dimensions. It is this proposition that we are about to consider, and we shall put the question in these terms: when we say that space has three dimensions, what do we mean?

Bibliography

[1] L.E.J. Brouwer, *Collected Works*, Vol. 1, A. Heyting, ed., Amsterdam (1975).
[2] L.E.J. Brouwer, *Collected Works*, Vol. 2, H. Freudenthal, ed., Amsterdam (1976).

[3] G. Cantor, *Georg Cantor: Briefe*, H. Meschkowski and W. Nilson, eds, Berlin (1991).

[4] J.-L. Chabert, *Un demi-siecle de fractales: 1870–1920*, Historia Mathematica **17** (1990), 339–365.

[5] J.W. Dauben, *The invariance of dimension: problems in the early development of set theory and topology*, Historia Mathematica **2** (1975), 273–288.

[6] J.W. Dauben, *Georg Cantor, his Mathematics and Philosophy of the Infinite*, Princeton (1979).

[7] P. Dugac, *Richard Dedekind et les Fondements des Mathématiques*, Paris (1976).

[8] A.A. Fraenkel, *Georg Cantor*, J. der D.M.-V. **39** (1930), 189–266.

[9] H. Freudenthal, *The cradle of modern topology, according to Brouwer's inedita*, Historia Mathematica **2** (1975), 495–502.

[10] I. Grattan-Guinness, *The rediscovery of the Cantor–Dedekind correspondence*, J. der D.M.-V. **76** (1974), 104–139.

[11] F. Hausdorff, *Dimension and äusseres Mass*, Math. Annalen **79** (1919), 157–179.

[12] D. Hilbert, *Mathematical Developments Arising from Hilbert's Problems*, F. Browder, ed., Amer. Math. Soc. (1976).

[13] D. Hilbert, *Grundlagen der Geometrie*, Festschrift zur Feier der Enthüllung des Gauss-Weber-Denkmals, Leipzig (1899), 3–92.

[14] W. Hurewicz and H. Wallman, *Dimension Theory*, Princeton (1948).

[15] A. Hurwitz, *Über die Entwickelung der analytischen Funktionen in neuerer Zeit*, Verhandlungen des ersten internationalen Mathematiker-Kongresses in Zürich vom 9 bis 11 August 1897, Leipzig (1897), 91–112.

[16] D.M. Johnson, *Prelude to dimension theory: the geometrical investigations of Bernard Bolzano*, Archive for History of Exact Sciences **17** (1977), 261–295.

[17] D.M. Johnson, *The problem of the invariance of dimension in the growth of modern topology, Part I*, Archive for History of Exact Sciences **20** (1979), 97–188.

[18] D.M. Johnson, *The problem of the invariance of dimension in the growth of modern topology, Part II*, Archive for History of Exact Sciences **25** (1981), 85–267.

[19] D.M. Johnson, *L.E.J. Brouwer's coming of age as a topologist*, Studies in the History of Mathematics vol. 26, E.R. Phillips, ed., The Mathematical Association of America (1987), 61–97.

[20] C. Jordan, *Cours d'Analyse de l'École Polytechnique*, 3 vols, Paris (1882–1887). 2nd ed. (1893–1896).

[21] M. Katetov and P. Simon, *Origins of dimension theory*, Handbook of the History of General Topology, C.E. Aull and R. Lowen, eds (1997).

[22] H.C. Kennedy, tr., ed., *Selected Works of Giuseppe Peano*, London, Toronto (1973).

[23] C.H. Kimberling, *Emmy Noether*, Amer. Math. Monthly **79** (1972), 136–149.

[24] T. Koetsier and J. van Mill, *General topology, in particular dimension theory, in the Netherlands: the decisive influence of Brouwer's intuitionism*, Handbook of the History of General Topology, C.E. Aull and R. Lowen, eds (1997), 135–180.

[25] I. Lakatos, *Proofs and Refutations. The Logic of Mathematical Discovery*, Cambridge (1976).

[26] N.J. Lennes, *Curves in analysis situs (abstract)*, Bull. Amer. Math. Soc. **12** (1906), 284–285.

[27] K. Menger, *Dimensiontheorie*, Leipzig, Berlin (1928).

[28] E.H. Moore, *On certain crinkly curves*, Trans. Amer. Math. Soc. **1** (1900), 72–90.

[29] G. Peano, *Sur une courbe, qui rempli une aire plane*, Math. Annalen **36** (1890), 157–160.

[30] H. Poincaré, *L'Espace et la géométrie*, Revue de Métaphysique et de Morale **3** (1895), 631–646.

[31] H. Poincaré, *On the foundations of geometry*, The Monist **9** (1) (1898), 1–43.

[32] H. Poincaré, *L'Espace et ses trois dimensions*, Revue de Métaphysique et de Morale **11** (1903), 281–301, 407–429.

[33] H. Poincaré, *Pourquoi l'espace a trois dimensions*, Revue de Métaphysique et de Morale **20** (1912), 483–504.

[34] A. Schoenflies, *Mengenlehre*, Encyclopädie der mathematischen Wissenschaften I(A5) (1898), 184–207.

[35] W.P. van Stigt, *L.E.J. Brouwer: Intuitionism and topology*, Proceedings of the Bicentennial Congress Wiskundig Genootschap, Amsterdam (1979).

[36] W.P. van Stigt, *Brouwer's Intuitionism*, Amsterdam (1990).

CHAPTER 2

The Concept of Manifold, 1850–1950

Erhard Scholz

Fachbereich Mathematik, Bergische Universität Gesamthochschule Wuppertal, Gaussstrasse 20, 42097 Wuppertal, Germany

1. Origin of the manifold concept

1.1. *n-dimensional systems geometrized*

In the early 19th century we find diverse steps towards a generalization of geometric language to higher dimensions. But they were still of a tentative and often merely metaphorical character. The analytical description of dynamical systems in classical mechanics was a field in which, from hindsight, one would expect a drive towards and a growing awareness of the usefulness of higher dimensional geometrical language.[1] But the sources do not, with some minor exceptions, imply such expectations. Although already Lagrange had used the possibility to consider time as a kind of fourth dimension in addition to the three spatial coordinates of a point in his *Mécanique analitique* (1788) and applied a contact argument to function systems in 5 variables by transfer from the 3-dimensional geometrical case in his *Theorie des fonctions analytiques* (1797, Section 3.5.25), these early indications were not immediately followed by others.

Not before the 1830-s and 1840-s do we find broader attempts to generalize geometrical language and geometrical ideas to higher dimensions: Jacobi (1834), e.g., calculated the volume of n-dimensional spheres and used orthogonal substitutions to diagonalize quadratic forms in n variables, but preferred to avoid explicit geometrical language in his investigations. Cayley's *Chapters in the analytical geometry of n dimensions* (1843) did use such explicit geometrical language – but still only in the title, *not* in the text of the article. It was the following decade about the middle of the century which brought the change. In a short time interval we find a group of authors using and exploring conceptual generalizations of geometrical thought to higher dimensions, without in general knowing about each other. Among them was Grassmann with his *Lineale Ausdehnungslehre* (1844) containing an explicit program for a new conceptual foundation for geometry on n-dimensional (lin-

[1] Such is suggested in some older historical literature, e.g., in F. Klein's *Vorlesungen über die Entwicklung der Mathematik im 19. Jahrhundert.* This is discussed with reference to more recent investigations in Section 2.3 below.

HISTORY OF TOPOLOGY
Edited by I.M. James

ear) extensional quantities,[2] Plücker with his *System der Geometrie des Raumes* (1846) and 4-dimensional line geometry in classical 3-space, and, in a certain respect most elaborated among these attempts, Schläfli with his *Theorie der vielfachen Kontinuität* (1851/1901), which was published only posthumously.[3]

Also leading mathematicians like Cauchy and Gauss started to use geometrizing language in $\mathbb{R}^n$ in publications (Cauchy, 1847) or lecture courses (Gauß, 1851/1917). Gauss, in his lecture courses, even used the vocabulary of *$(n-k)$-dimensional manifolds* (*Mannigfaltigkeiten*), but still restricted in his context to affine subspaces of the n-dimensional real space (Gauß, 1851/1917, pp. 477ff.). There is no reason to doubt that Riemann got at least some vague suggestion of how to generalize the basic conceptual frame for geometry along these lines from Gauss and developed it in a highly independent way.

1.2. *Riemann's n-dimensional manifolds*

When Riemann presented his ideas on a geometry in manifolds the first time to a scientific audience in his famous *Habilitationsvortrag* (Riemann, 1854), he was completely aware that he was working in a border region between mathematics, physics, and philosophy, not only in the sense of the pragmatic reason that his audience was mixed, but by the very nature of his exposition.[4] There was no linguistic or symbolical frame inside mathematics, which he could refer to, even only to formulate a general concept of manifold. So he openly drew on the resources of contemporary idealist, dialectical philosophy, in his case oriented at J.F. Herbart, to generalize the classical concept of *extended magnitude/quantity* for geometry and to "construct" the latter as only one specification from a more general concept.[5] Basic to such a construction was, so Riemann explained to his audience, the presupposition of any "general concept" which allows in a logical sense precise individual determinations. From the extensional point of view such a concept would form a *manifold* and the individual modes of determination were to be considered, as Riemann explicitly stated, as the *elements* or the *points* of the manifold with either "discrete" or "continuous" transition from one to the other. Thus Riemann sketched the draft for a conceptual starting point for what later was to become general set theory (discrete manifolds)[6] and topology (continuous manifolds).

Such concepts would gain mathematical value only if a sufficiently rich structure of (real or complex valued) functions on the manifold is available. Then it should be possible to describe the specification of points by the values of n properly chosen functions in a locally unique way (local coordinate system). That a change of coordinates would lead to locally invertible differentiable real functions, was not made explicit by him, but was to be understood from the context by careful listeners or readers. The distinction between *local simplicity* of manifolds, because of the presupposition of local coordinate systems, and *globally involved behaviour* was indicated by Riemann, but not particularly emphasized during the talk, although in other publications and manuscripts it was.[7]

---

[2] Hamilton's *quaternions* used 4-dimensionality for purely algebraic reasons, keeping geometry restricted to the 3-dimensional subspace of purely imaginary quaternions.

[3] See [Kolmogorov and Yushkevitch, 1996].

[4] For a detailed and very readable exposition of the width of Riemann's interests see [Laugwitz, 1996].

[5] See [Scholz, 1982a].

[6] For the line from Riemann via Dedekind and Cantor to general set theory see [Ferreirós, 1993, 1996].

[7] Compare the next two sections of this article.

Of the utmost importance was Riemann's discussion of different conceptual levels – we would say *structures* – which can be considered on a given manifold. During his talk he exemplified these by the distinction between analysis situs (combinatorial topology of differential manifolds) and differential geometry. In his works on complex function theory he moreover pursued concrete investigations of complex and birational structure in the complex one-dimensional case (Riemann, 1851, 1857).[8] And there are points in the latter publications, where Riemann indicated that it might be useful to work with even more "general concepts" of a continuous character, which would transcend the limits of the specific postulates for continuous manifolds introduced or at least presupposed in his Habilitations lecture. Thus in his dissertation Riemann (1851, p. 36) had already talked about infinite dimensional (real) function spaces and continuously varying conditions for functions in them, given by equations, which indicated nonlinear subsets in the dual of functions spaces. Moreover, Riemann had even already used the language of "continuous manifold" in this context without further specification what should be understood by that term.[9] That was a drastic generalization of Gauss's finite dimensional linear submanifolds of $\mathbb{R}^n$ and even far more general than the manifold concept as developed by Riemann in 1854.

That was three years before his Habilitations lecture. Three years after the latter, in his work on abelian functions, Riemann indicated how the complex/birational structure on a closed orientable surface of given genus p can be characterized by $3p - 3$ independent complex parameters describing a normalized branching behaviour over the complex plane. He thus started to explore the moduli space of Riemann surfaces of genus p and was cautious enough, not to talk about them as manifolds, but left it with a local description at generic points (Riemann, 1857, p. 122).

Thus Riemann presented an outline of a visionary program of a family of geometrical theories, bound together by the manifold concept, diversified by different conceptual and technical levels like topology, differential geometry, complex geometry, algebraic geometry of manifolds, and overarching the whole range from questions deep inside conceptual ("pure") mathematics to the cognition of physical space and the nature of the constitution and interaction of matter. Here is not the place to follow all these branches; we rather concentrate on the tools for a topological characterization of manifolds with some digressions into the broader context.

1.3. *Riemann on the topology of surfaces ...*

Riemann used different approaches in his studies of surfaces. Already in his dissertation he dealt with the connectivity of compact bounded surfaces. His goal was to introduce complex analytic functions on (Riemann) surfaces over a bounded region of the complex plane. For simply connected surfaces he used his famous argumentation by the Dirichlet principle to determine real and imaginary parts of a complex function by the potential equation and boundary value conditions.[10] Here he characterized simple connectedness of

[8] Some more details in [Dieudonné, 1974, pp. 42ff.; Gray, 1986; Scholz, 1980, pp. 51ff.].

[9] "These conditions, the totality of which form a continuous manifold and which can be expressed by equations between arbitrary functions ... still have to be limited or supplemented by single conditions for arbitrary constants ..." (Riemann, 1851, p. 36).

[10] See, e.g., [Bottazzini, 1986, pp. 229f.].

a surface F by the condition that F falls apart by any cross cut leading from one point of the boundary ∂F to another.[11] For not simply connected surfaces he introduced a connectivity number by a cut and count procedure.

If F can be dissected by m cuts along double-point free curves between the boundary of F or new boundary components arising from earlier cuts into n simply connected pieces, then, so Riemann argued, the difference $n - m$ is independent of the cutting procedure and a topological invariant. In fact Riemann's counting procedure can be read as a characterization of the Euler number $\chi(F)$ of the surface with each cross-cut increasing the Euler number by 1 (adding 2 zero cells and 1 one-cell) and leaving n simply connected surfaces $\chi(F) + m = n$, thus $\chi(F) = n - m$. By a specific choice of the dissection it is possible to reach exactly one simply connected piece at the end of the process, $n = 1$, giving the lowest number of cross cuts necessary, $m_0 = 1 - \chi(F)$. In this case Riemann would call the surface $(m_0 + 1)$-fold connected.

In his later work on abelian integrals and functions (1857) Riemann considered surfaces over the whole (compactified) complex plane and thus closed orientable surfaces. In order to apply his early counting method for the connectivity number he showed that "recurrent cuts (Rückkehrschnitte)" do not change the latter (adding 1 zero cell and 1 one-cell) thus allowing him to apply the old method also to this case. His interest was now directed towards a different type of question: the periods of abelian integrals of first (or higher) kind, i.e. the characterization of multivaluedness of integrals of a holomorphic (or meromorphic) differential form ω on a closed Riemann surface $F \to P_1\mathbb{C}$. Starting from a general 2-dimensional version of the Gauss–Stokes theorem and the Cauchy–Riemann equations for the coefficients of the holomorphic form ω, he realized that (in modernized notation) $\mathrm{d}\omega = 0$ and therefore for any set of closed (oriented) curves c_i, $1 \leqslant i \leqslant k$, forming a complete boundary of a part F' of the surface,

$$\bigcup_{i=1,\dots,k} c_i = \partial F',$$

the evaluation of the integral will give zero:[12]

$$\int_{c_1,\dots,c_k} \omega = \int_{F'} \mathrm{d}\omega = 0.$$

Therefore, so Riemann concluded, the multivaluedness of integrals of holomorphic 1-forms (abelian integrals of the first kind) depends only (and still to a high degree in the case of meromorphic 1-forms, the abelian integrals of second and third kind)[13] on the topology of the surface. So it was reasonable to characterize the topology of closed (orientable) surfaces in this context by a method of boundary relations between systems of curves, which from the later point of view reads as a first step towards a homology theory of 2-dimensional manifolds.

---

[11] Riemann thus used a purely homological characterization of simple connectedness, in contrast to the modern post-Poincaréan view. Compare the contribution by R. Vanden Eynde, this volume.

[12] (Riemann, 1857, pp. 91ff.), compare also the contribution of R. Vanden Eynde, this volume.

[13] $\int_c \omega$ is an *abelian integral of second kind* if ω is a meromorphic differential form only with poles of order $m > 2$ and *abelian integral of third kind* if ω is a meromorphic form with poles of order 1 but with sum of residues 0.

For the purely topological part of his investigation Riemann did not take into account the orientation of curves or surface parts, thus simplifying the calculations. He introduced an equivalence between systems of curves C and C' if both together form the complete boundary of part of the surface $C + C' = \partial F'$, as in this case C and C' "achieve the same with respect to forming complete boundaries" with other curves (Riemann, 1857, p. 124). In slightly modernized reading Riemann thus worked with a geometrical description of bordance homology of submanifolds in F modulo 2, or, in another translation, with simplicial homology, if F is simplicially decomposed by cuts along the curves c_i such that the latter represent 2-cycles of the decomposition. Indeed Riemann showed that there is a well-determined number n of homologically independent curves, independent of the choice of the specific realization of the curve system, and that in the case of his surfaces this number is even, $n = 2p$ (Riemann's notation (Riemann, 1857, p. 136)).

Of course, Riemann did not keep to the modulo 2 reduction of homology when working with integrals of differential forms. Once a complete set of generators of the homology $c_1, \ldots, c_{2p}$ and corresponding periods $w_i = \int_{c_i} \omega$ $(1 \leqslant i \leqslant 2p)$ of a differential form were determined, he worked with integral linear combinations of the periods and thus (at least implicitly) with unreduced integral combinations of cycles (Riemann, 1857, pp. 137ff.). So the modulo 2 reduction was for him nothing more than a method to simplify the calculation of the topological invariants and in fact a result of a context dependent abstraction from orientation.

1.4. *... and on the connectivity of higher dimensional manifolds*

In the edition from Riemann's Nachlass Weber edited three fragments about analysis situs (Riemann, 1876a) in which Riemann explored first thoughts on the topological characterization of higher dimensional manifolds. These fragments can be dated with great probability to the time of Riemann's work on his Habilitationsschrift, thus about the years 1852/1853.[14] Here Riemann described the introduction of higher connectivity numbers using a bordance homological approach similar to the one later published in his theory of abelian functions and discussed in the last section. He considered closed connected submanifolds U_i, $1 \leqslant i \leqslant m$, of dimension n in a manifold M of dimension k,[15] which "taken each once, neither individually nor jointly" form the complete boundary of an $(n+1)$-dimensional submanifold, which means, expressed in more recent terminology, they form a set of homologically independent n-cycles.[16]

Riemann explicitly defined homological equivalence of n-cycles A and B, using the terminology of "transmutability" of A into B.[17] Riemann then argued with an exchange argument which algebraically expressed would be the Steinitz lemma and the change of generators in the homology vector-space (take into consideration that Riemann worked

[14] For more details see [Scholz, 1982b].

[15] Riemann used the terminology "innere zusammenhängende unbegrenzte n-Strecke" for the U_i, without further specification of the objects considered. From a recent mathematical perspective "n-Streck" should perhaps not be understood as submanifold, but as "subvariety" admitting certain controlled singularities like the topological varieties in (Kreck, 1998).

[16] As in the last section suppose there is a simplicial decomposition of M, in which the U_i represent n-cycles.

[17] "Ein n-Streck A heisst in ein anderes B veränderlich, wenn durch A und durch Stücke von B ein inneres $(n+1)$-Streck vollständig begrenzt werden kann." (Riemann, 1876a, p. 479).

mod 2): If V_i $(1 \leqslant i \leqslant m)$ is another set of n-submanifolds which fulfill the same boundary conditions as the U_i, each of which forms jointly with some of the U_i the complete boundary of an $(n+1)$-submanifold, then with respect to the formation of bounding relations the V_i can be substituted step by step for the U_i and in the end the V_i and the U_i $(1 \leqslant i \leqslant m)$ can be considered equivalent in the context of forming boundary relations inside the manifold M.

Riemann thus introduced the maximal number m of (mod 2) homologically independent n-cycles, i.e. the nth Betti number mod 2, and called the manifold M $(m+1)$*-fold connected in dimension n* (ibid.). In particular, he called M simply connected if all connectivity numbers (Betti numbers) mod 2 of M are zero, thus deviating from the modern, post Poincaréan, terminology (or better the other way round).[18] He started to investigate the decomposition of a k-dimensional manifold by dissection along lower dimensional submanifolds, and tried to generalize his decomposition method from 1851 for surfaces to higher dimension, although he did not fully elaborate a symbolism to characterize types of such decompositions or topological invariants. The fragments leave no doubt, however, that already at the time of his Habilitationsvortrag he had a rather clear conceptual construction of Betti numbers modulo 2 in mind, taking into account the level of elaboration of symbolical characterization of manifolds. Enrico Betti seems to have been the only mathematician to whom he talked about these concepts in sufficient detail to transmit the essentials of his ideas. At least Betti was the only one in Riemann's lifetime, who understood what the latter was heading for.[19]

2. Dissemination of manifold ideas

2.1. *The problem of how to characterize manifolds*

The reception and assimilation of Riemann's concept of manifold to the mathematics of the 19th century was slow and inhibited by severe conceptual problems. Of course it was difficult to understand what a manifold in general should be. The easiest way was to translate it as a "number manifold" in the 1870-s and later. At that time the former real quantities had been arithmetically reconstructed by Meray, Cantor, Dedekind, and Weierstrass, and it appeared as perfectly clear to talk about concretely given submanifolds of $\mathbb{R}^m$ or of projective spaces $P_m\mathbb{R}$ or $P_m\mathbb{C}$. Such submanifolds were in the easiest approach defined by inequalities as m-dimensional (usually connected) subsets in the works of Beltrami (1868a, 1868b) Helmholtz (1868), and even of the young Klein during his investigations on non-Euclidean geometry and the Erlangen program (1871).

That was of course a reduction of Riemann's intention and suppressed the distinction between local simplicity and global complexity of manifolds. That global behaviour was an essential ingredient for Riemann's concept, was most clearly understood in the 1860-s and 1870-s in the special context of geometric function theory and the dissemination of knowledge about the topology of Riemann surfaces (Lüroth, Clebsch, Neumann, Clifford et al.) An additional aspect was the problem of compactification of geometrical objects "in the infinite", which in a discussion between Schläfli and Klein was realized, when they

[18] Compare the contribution of R. Vanden Eynde and footnote 11.

[19] For the relationship between Riemann and Betti see [Bottazzini, 1977].

debated the difference between one-point compactification of the plane like $\mathbb{C}$ to $P_1\mathbb{C}$ and line compactification of $\mathbb{R}^2$ to $P_2\mathbb{R}$ and its topological consequences (Schläfli, 1872; Klein, 1873, 1874–1876).

Only after discussions with Clifford on space forms during the 1873 meeting of the British Association for the Advancement of Science, Klein modified his earlier restricted concept of "manifold" and introduced the distinction between *relative* properties of a "number manifold", which depended on the embedding, and *absolute* ones which did not, orientability given as an example for the latter, without going into technical details of how to identify the "absolute" properties (Klein, 1874–1876).

A more general characterization of "number manifolds" was the consideration of zero sets by equations (and inequalities), which usually were supposed to be nonsingular without further specification. This approach was taken by Betti (1871) in his paper on the topology of higher dimensional manifolds and by Lipschitz in his investigations of higher dimensional state spaces of mechanical systems (Lipschitz, 1872).[20] In Betti's case global complexity was, of course, part of his object of study. The local simplicity, however, remained unanalyzed before the proof of the implicit function theorem, including an explicit statement of the condition under which it holds, became generally known. The theorem and its proof was developed by U. Dini during his lecture courses in the late 1870-s and spread in analysis courses and monographs during the late 1880-s and early 1890-s.[21]

Finally, a first, still clumsy and vaguely described, combinatorial approach to a characterization of n-dimensional manifolds was used by Klein's student W. Dyck in addition to the characterization as a "number manifold". Although starting as Klein had done from a submanifold M of $\mathbb{R}^n$, Dyck gave a vague description of how to build M from an n-ball E_n by cutting and pasting along submanifolds of type E_k isomorphic to k-balls (von Dyck, 1888, 1890). This process was not uniquely described in Dyck's symbolism and presupposed sufficient intuition to be applied to a manifold defined by other means. It still sufficed for Dyck's purpose, as his procedure served only as an aid for the topological characterization of manifolds, not for their definition or construction.

2.2. *The changing concept of geometry*

During the 19th century the perception, structure and role of geometry was fundamentally transformed. Classically there existed but one, Euclidean geometry, and its unique role in the framework of knowledge at the turn from early modernity to "high" modernity was paradigmatically exemplified in Kant's philosophy of space. The breakthrough in the studies of the foundations of geometry has been described by I. Tòth as the shift from the "anti-Euclidean" hypothesis to the non-Euclidean point of view;[22] it was realized independently, as is well known, by Gauss, Lobachevsky, and J. Bolyai. Until the 1860-s this change of view was shared only by a small minority of mathematicians, and was moreover conceptually still rather fragile, as long as *only* the theoretical structure of non-Euclidean

[20] For Lipschitz compare [Lützen, 1995] and the Section 2.3 below.

[21] Two important publications for the dissemination of the implicit function theorem were (Peano and Genocchi, 1884) and Jordan 2nd edition of the *Cours d'analyse* (Jordan, 1893, pp. 80ff.). For Dini's broader contribution to the foundation of real analysis see [Bottazzini, 1985].

[22] [Tòth, 1972, 1980].

geometries had been outlined, with no mathematical (or physical) interpretation in terms of accepted objects and relations being given.[23]

Gauss was apparently well aware that his differential geometry of surfaces might carry the potential to open a route towards such a missing interpretation, but he could not (or at least did not) solve the dilemma from a foundational point of view, that his surfaces were constructed inside the framework of Euclidean geometry. His reaction to Riemann's Habilitations lecture shows how well Gauss understood that Riemann had given a beautiful outline and far reaching program for another and much deeper conceptual step towards a trans-Euclidean geometry, which would reduce non-Euclidean theory in the sense of Bolyai and Lobachevsky to nothing but a special case. Riemann even sketched such a reduction in the last section of his talk, although he apparently had no knowledge of Bolyai's or Lobachevsky's studies in the foundations of geometry.[24]

But the concept of manifold became essential for the understanding of non-Euclidean geometries in the late 1860-s and early 1870-s when the latter became finally absorbed into the general knowledge of mathematics. All three main contributors to non-Euclidean geometry in this phase, Beltrami, Helmholtz, and Klein, did refer to Riemann, whose Habilitations lecture became accessible to a wider scientific public outside Göttingen in 1867 after the publication in the *Göttinger Abhandlungen* (vol. 13). Here is not the place to discuss the role of Riemannian ideas in the development of knowledge and the discourse on non-Euclidean geometry in detail. It has to be said, however, that among the just mentioned authors, involved in the development of non-Euclidean geometry in the 1860-s, only Klein had been in contact with Riemannian ideas before he started to work on non-Euclidean geometry, through his close cooperation with A. Clebsch from 1866 onward. Beltrami and Helmholtz, in contrast, started to develop their ideas independently and progressed considerably before they learned to know of Riemann's lecture and adapted their presentation according to the latter's outlook. The shift in Beltrami's argument due to the influence of Riemann's view was particularly clear and seems to be characteristic for the broader turn geometry went through in the 1860-s and 1870-s and in particular to the role of the manifold concept in it.

E. Beltrami had started on his own in 1866 and 1867 to explore the possibilities inherent in the Gaussian theory of surfaces for an interpretation and understanding of non-Euclidean geometry. In early 1867 he realized that the geometry of the non-Euclidean plane can be gained in terms of a *generalized* Gaussian surface, i.e. the region

$$A = \{x \mid |x|^2 < a^2\} \subset \mathbb{R}^2$$

with metric not induced by an embedding in Euclidean 3-space, but "formally" given by

$$\mathrm{d}s^2 = \frac{r^2}{(a^2 - x_1^2 - x_2^2)^2}\left(\left(a^2 - x_2^2\right)\mathrm{d}x_1^2 + 2x_1x_2\,\mathrm{d}x_1\,\mathrm{d}x_2 + \left(a^2 - x_1^2\right)\mathrm{d}x_2^2\right).$$

[23] The problematics of this type was addressed in Riemann's 1854 lecture by his opening remark, that earlier investigations on the foundations of geometry worked with purely "nominal" definitions. Although this remark was addressed at classical Euclidean definitions, Riemann hit a point which was even of higher importance for the contemporary status of non-Euclidean geometry, the discourse of which was apparently not known to him.

[24] Compare [Scholz, 1982a, pp. 220f.] and [Laugwitz, 1996].

He derived all properties essential for what would later be called the "Beltrami model" of the non-Euclidean plane, but insisted on the necessity to find a *real substrate* (*substrato reale*) of this "purely formally given" system in order to understand its geometric meaning. He was glad to find such a "real substrate", by local isometric embeddings in classical Gaussian surfaces of constant negative curvature $\kappa = -r^{-2}$, embedded in Euclidean 3-space. So he sent a manuscript under the title *Saggio di interpretazione della Geometria non Euclidea* (published as (Beltrami, 1868a)) to Cremona as editor of the *Giornale di Mathematiche*. Cremona disagreed with Beltrami's narrow conception of "real substrate" of geometry, but nevertheless voted for publication after some period of hesitation and exchange with Beltrami about his views. Probably he doubted among others the mathematical value (not the correctness) of Beltrami's observation that, although a "real substrate" could be given for the non-Euclidean *plane* by local isometric embeddings in classical Euclidean space, nothing similar could be hoped for in the case of three-dimensional non-Euclidean geometry (Beltrami, 1868a, p. 284).

After delivering the manuscript of his *Saggio* Beltrami got to know Riemann's Habilitations lecture (maybe through a hint by Cremona) and changed his mind with respect to the epistemological (or even "ontological") role of a classical interpretation for non-Euclidean concepts. He immediately prepared a second publication in which the two-dimensional case was generalized and an n-dimensional differential geometrical model for non-Euclidean geometry, using a simple Riemannian manifold representation, was given: $M \subset \mathbb{R}^{n+1}$ defined as a hemisphere, $|x|^2 = a^2$, $x_{n+1} > 0$, with metric induced by

$$ds^2 = \left(r^2/x_{n+1}^2\right) \sum_{i=1}^{n+1} dx_i^2$$

on $\mathbb{R}^{n+1}$. Parametrization of M by the open ball $|\tilde{x}|^2 < a^2$ with $\tilde{x} \in \mathbb{R}^n$ leads back to the case presented in the *Saggio* for the two-dimensional case.

Both articles appeared in the same year, although in different journals; Beltrami only made small adaptations in the text of the first one with general references to the possibility of a more conceptual understanding of non-Euclidean geometry than looking for a "real substrate". The second article appeared as *Teoria fondamentale degli spazii di curvatura costante* (Beltrami, 1868b). The shift in interest and in outlook on the basic concepts of geometry between these two publications of Beltrami may serve as a concentrated expression for what was at stake in the change from classical geometry to modern geometry of manifolds. Beltrami lived through such a change in a couple of months, because his own line of thought already had brought him to the point of a formal generalization of Gauss's theory of surfaces, and the inherent movement was so well dynamized by Riemann's presentation.

Once Riemann's construction of manifolds was accepted, even if only in the concrete version of "number manifolds", the question of a "real substrate" for non-Euclidean geometry changed its meaning completely. To use later terminology, a differential geometric *model* of the metrically well explored (although from the axiomatic point of view still not completely elaborated) theoretical structure of non-Euclidean geometry could be given in a drastically extended framework. For the modern reader this extended conceptual framework has become so common that she may tend to overlook the hard work necessary to achieve the state of disciplinary practice and knowledge she is used to.

2.3. *First appearance of manifolds in mathematical physics*

Of course there are several semantical links of the manifold concept to physics, which could be pursued even in the 19th century. Riemann had already started to discuss such links on at least two levels. The final part and culmination of his Habilitations talk gave a sketch how in a subtle interplay between mathematical arguments and the evaluation of physical/empirical insights he proposed to come to a refined understanding of physical space. The essential bridge was an improved understanding of the miscrostructure of matter and its binding forces that should be, according to Riemann, as directly translated into differential geometric structures on manifolds as possible. But he also left the possibility open for further consideration that perhaps some time even a discrete structure of matter has to be taken into account, as it might very well be that the concepts of rigid body and light ray use their meaning in the small. Still, so Riemann argued by reference to astronomical measurements, the acceptance of a Euclidean space structure was well adapted to the physical knowledge of the time.

A second link was indicated in his famous Paris prize essay (Riemann, 1861/1876). Riemann there had modelled a three-dimensional heat flow problem in an ex ante inhomogeneous matter region and translated it into a differential geometric structure of a 3-dimensional Riemannian metric. In the result the question of a homogeneity criterion for the underlying matter could be analyzed as a question of local flatness of the metric. As is well known, that was the context in which Riemann published his most advanced results characterizing the curvature of a Riemannian manifold.[25] There should be no serious doubt, however, that Riemann was completely aware about the importance of such a connection between differential geometry and other parts of analysis or physics, although he did not, in the prize essay, elaborate explicitly on such a semantical connection, but motivated the interested reader to think along such lines by a highly interpretable reference to a Newton citation: "Et his principiis via sternitur ad majora." (Riemann, 1861/1876, p. 391)[26]

Recent historical investigations have shown how deeply connected large parts of the geometric discourse of the 19th century were to the semantics of physical space, even in parts of the discussion where, after the epistemological shift of mathematics brought about by the rise of set theory and the axiomatization movement at the turn of the century, a modern reader would not look for a direct semantical context in physical terms and would perhaps even tend to consider some parts of the debate at the end of the 19th century stricken by a surprisingly naive realism. This aspect has been discussed in detail by M. Epple in his [1997] and shall not be reproduced in this article.[27] Of long-ranging interest for the development of higher dimensional manifolds in physics were, on the other hand, the first moves for a geometrization of state spaces in mechanics. This aspect has recently studied by J. Lützen, and my short report relies completely on his results.[28]

[25] Compare among others [Reich, 1994; Laugwitz, 1996; Farwell and Knee, 1990; Scholz, 1980].

[26] Superficial and textpositivistic reading might give another picture of Riemann's intention. There are contributions to the historical literature like [Farwell and Knee, 1990] which deny the differential geometric content of Riemann's (1861/1876).

[27] Compare also M. Epple's contribution in this volume for less "naive" attempts at physical semantics of topological concepts.

[28] Cf. [Lützen, 1988, 1995].

Most important among the geometrization arguments in this problem field were the following:

(1) The subsumption of the least action principle for conservative systems under the form of a geodetical line. The state space was endowed with a physical metric of the form $\mathrm{d}s^2 = 2(H - V)\sum g_{ij}\,\mathrm{d}q_i\,\mathrm{d}q_j$, for q_i coordinates in the state space, g_{ij} the metric induced on state space by the metric of the geometric coordinate space, H total energy, and V potential energy. That had been done analytically by Jacobi in the 1830-s and geometrized in low dimension ($n = 2$) by Minding, Liouville, and Serret about the middle of the century. Geometrization for higher dimensions was apparently discussed in the 1870-s and published, e.g., by Darboux in a particularly clear way in the year 1888.

(2) Already a decade earlier Lipschitz developed a generalization of classical mechanics starting from a metric in the underlying geometrical space, which he allowed not only to be Riemannian but even Finsler (in modern terms) (Lipschitz, 1872). On that basis he developed a generalization of the term for the kinetic energy and the Hamilton–Jacobi form of mechanics. Moreover, in his discussion of conservative systems, he described the trajectories in the state space as (generalized) orthogonal to 1-codimensional submanifolds of the state space [Lützen, 1995, Section 49]. Other authors, not all of them aware of Lipschitz' research, like Thomson/Tait and Darboux, pursued similar intentions.[29]

(3) The "Liouville theorem" on the volume preservation of the time flow in the phase space of Hamiltonian mechanics with canonically conjugate coordinates q_i, p_i, $1 \leqslant i \leqslant n$, and dynamical equations

$$\frac{\mathrm{d}q_i}{\mathrm{d}t} = -\frac{\partial H}{\partial p_i}, \qquad \frac{\mathrm{d}p_i}{\mathrm{d}t} = -\frac{\partial H}{\partial q_i},$$

which was presented by Liouville only in analytical formulation in a more general context (for the first time in 1838). Jacobi transferred it to mechanics by Jacobi about the middle of the century. Geometrization appeared in works on statistical mechanics only in the late 1860-s early 1870-s by Maxwell and Boltzmann (apparently without knowing about Liouville's result).[30]

(4) Finally, Boltzmann's discussion of different types of dynamical systems to characterize his idea of entropy contains a broad range of high-dimensional arguments in configuration or phase space, although in a highly intuitive manner. These are interesting questions for a broader history of the use of advanced mathematical concepts inside late 19th century physics, which are impossible to report here.

With respect to the claim made by Felix Klein in his famous historical lectures, that Gauss's and Riemann's differential geometry has supposedly "grown from up from the soil of the Lagrangian equations" [Klein, 1926/1927, p. 146], J. Lützen has shown in his detailed historical studies of the sources that this remark distorts history highly. Klein apparently did not allow for sufficient distinction between the original historical development (as documented and accessible from the sources) and his own perception of Riemannian

[29] Lützen considers Lipschitz' generalized orthogonal trajectory discussion as the most important geometrization approach in mechanics during the 19th century [Lützen, 1995, Section 51].

[30] For more details see [Lützen, 1990, pp. 657ff.].

geometry that had formed in the late 1860-s early 1870-s, when he was a young mathematician and participated in the production of the events which he later told the history of. Explicit geometrization of configuration (state) and/or phase spaces of mechanical systems was in fact undertaken only in that relatively later period in which Klein was actively involved, and only then the language of higher dimensional manifolds became part of the discourse of theoretical physics and vice versa.

3. Steps towards a topological theory of manifolds

3.1. *The 2-dimensional case as an elementary paradigm ...*

It was also in the 1860-s and early 1870-s that several lines of thought intertwined productively and led to the first relatively well-explored segment of a theory with links to different fields of study in the mathematics of the 19th century, the combinatorial theory of polyhedra,[31] complex function theory, real projective algebraic geometry and the newly rising topological theory of manifolds. The main contributors to this subfield were A.F. Möbius (1863, 1865), C. Jordan with a series of publications through 1866, and Schläfli and Klein in their discussion on the orientability of real 2-dimensional subspaces of the projective space.[32] In geometric function theory divers authors contributed to a refined understanding of the role of topological concepts, in particular C. Neumann with his calculation of the connectivity of a Riemann surface from the winding orders of branch points,[33] Lüroth, Clebsch and Clifford with their normalized representation during the 1870-s for branched coverings of $P_1\mathbb{C}$, which represent a Riemann surface with given number of leaves, given loci and winding numbers of branch points.

Möbius and Jordan both discussed independently from each other, which "morphisms" they wanted to consider for their topological theory of surfaces. Möbius called them "elementary relationships (Elementarverwandtschaften)" and Jordan just talked about "mappings", and both circumscribed a transformation of "infinitely small elements" of one into the other, respecting neighbouring relations. They indicated that this idea could in principle be made precise by infinite series of subdivisions of the surfaces into finite surface "elements" which are one-to-one correlated, respecting the neighbouring relations.[34]

Möbius gave in his article (1865) a detailed analysis of orientation procedures in surfaces, which he decomposed in polygonal nets (a generalized representation of a triangulation). He defined orientations of the boundaries of each polygon and coherence of neighbouring polygons, if the induced orientations in the common part of the boundaries are inverse to each other. As an application he gave the famous example of a non-orientable surface: a "Möbius band" complemented by a disc to form a closed non-orientable surface homeomorphic to $P_2\mathbb{R}$ (which Möbius did not explicitly remark) (Möbius, 1865, p. 483).

His earlier publication on "elementary relationships" contained a topological classification of closed orientable surfaces embedded in $\mathbb{R}^3$ (without self-intersection). He classified singular points of a "height" function geometrically into "elliptical" and "hyperbolic"

[31] See [Lakatos, 1976].

[32] For Schläfli–Klein compare Section 2.1.

[33] Genus p of the surface given by $2p = \sum_{i=1}^{k}(m_i - 1) - 2n - 2$ for a Riemann surface with n leaves over $P_1\mathbb{C}$ and k branch points of orders $m_i - 1$ (Neumann, 1865).

[34] More details in [Pont, 1974] or [Scholz, 1980, pp. 148ff.].

points and developed what from a 20th century point of view reads as a geometric presentation of the Morse theory of differentiable closed orientable surfaces. He showed that each such surface F can be constructed from two homeomorphic ("elementary equivalent") surfaces F_1 and F_2, each with exactly n boundary components, which are pasted together at the boundary components. Möbius called n the "class" of the surface and showed that it was a classifying invariant. He did not remark, however, that his "class" and Riemann's "genus"[35] p were essentially the same, with $p = n - 1$.

While there is no indication that Möbius knew the function theoretic work of Riemann and its topological aspects, he did connect his studies to the famous debate on Euler's polyhedral formula and proved it in the general case using his invariant, $\chi(F) = 2(2-n)$. In our eyes it reads of course more naturally if rewritten with Riemann's invariant, $\chi(F) = 2 - 2p$.

Jordan classified orientable surfaces, including those with boundary, independently from Möbius. He counted the maximal number k of recurrent cuts (cuts along double-point-free pairwise disjoint closed curves c_i, $1 \leqslant i \leqslant k$), which do not dissect the surface into disconnected pieces, and the number m of boundary components. He showed that the pair (m, k) classifies the orientable (compact) surfaces uniquely (Jordan, 1866, p. 85). For the proof he used dissection of the surfaces along the recurrent cuts and additional cross cuts and topological maps of the resulting simply connected pieces.

Jordan, in contrast to Möbius, was aware of the connections between the topological theory of surfaces and complex function theory. Another aspect of his work on surfaces, the study of homotopy classes of closed paths, very likely was motivated by this context, although he did not remark so explicitly and left it to the reader to realize it. Riemann had been inspired in his topological investigations of surfaces by the behaviour of the integrals of holomorphic differential forms and thus considered a homological equivalence concept between closed paths (cycles); but of course in complex function theory the question of analytic continuation and the resulting questions of multi-valuedness played an important role (including Riemann's work, as is well known). For analytic continuation the continuous deformation of paths, or in later terminology a homotopic concept of equivalence between cycles was the proper one to study. Jordan did not explain this, but he gave a complete description of the homotopy theory of his bounded orientable surfaces, including definition of the equivalence concept, generators and relations of the fundamental group. This beautiful and surprising aspect of Jordan's work is discussed in detail in Vanden Eynde's contribution in this volume and therefore not documented in more detail here.

Here I only what to repeat that Jordan did not use explicit group terminology, as the group concept was in the middle of the 1860-s still essentially confined to substitutions. He nevertheless must have been aware of a conceptual relationship between what he did with the deformation classes of closed paths and groups, as he had been actively involved in Galois theory in the time immediately before.

Taken Riemann's, Möbius', and Jordan's work together, and perhaps adding Schläfli and Klein, it becomes clear that at the transition from the 1860-s to the 1870-s a complete topological theory, including classification, homology and homotopy aspects for compact orientable surfaces was at hand and widely accessible, and a first elaboration of questions of non-orientability had been started. Betti, moreover, had indicated how a generalization

[35] The terminology "genus" is due to Clebsch (1864).

of the homological part of the theory to higher dimensions might work; although the route he had indicated was still unexplored.

3.2. ... *and first attempts to understand higher dimensions*

In 1871 Betti published his presentation of higher numbers of connectivity. The objects of his study were n-dimensional submanifolds S_n of an $\mathbb{R}^m$, *spazi*, as he called them, in general supposed to be closed and connected.[36] The method of characterizing connectivity numbers ("Betti numbers") p_k $(1 \leqslant k \leqslant n)$ was to consider maximal systems of closed k-dimensional submanifolds, U_i $(1 \leqslant i \leqslant p)$, which "cannot form the border of a pathwise connected (sic!) $(m+1)$-dimensional part of the space" (Betti, 1871, p. 278).

Like Riemann in his (not yet published) fragment Betti argued that the maximal number p is independent of the choice of the system of submanifolds, using step-by-step substitution of the cycles. His verbal description of the boundary relation was, however, not precise enough to exclude counterarguments, which were given by Tonelli (1875) showing that a more refined symbolism for the representation of the cycles and their homology relations was needed. Moreover, Tonelli corrected the unnecessary and for the argument detrimental specification of pathwise connectedness for the bounding part of the surface. These necessary criticisms did not lessen Betti's achievement of a public presentation of the first step towards a homological theory of manifolds, which until then had lain latent in the thought and manuscripts of Riemann and some (provably at least one) of the latter's closest correspondents.

There remained the lacuna, however, that although the method was presented for n-dimensional (closed) manifolds in general, no new insights were immediately accessible by this method for higher dimensions with the exception of the simplest three-dimensional examples. Betti, e.g., discussed the connectivity of the "thickened" two-sphere and the massive and the "thickened" torus in $\mathbb{R}^3$ in letters to P. Tardy written in 1863, although published only in 1915 (Betti, 1915). It nearly remained so until Poincaré's great series on analysis situs at the turn of the century. There was, however, at least one other intermediate step of long standing significance, E. Picard's investigation of the topology of complex algebraic surfaces at the end of the 1880-s and in the early 1890-s.

Picard combined with great imagination ideas from algebraic geometry, complex analysis, early homology and homotopy to analyze the topological structure of algebraic curves. He noticed in the early 1880-s, as M. Noether had done already a decade earlier[37] that in algebraic surfaces integrals $\int_c \omega$ of meromorphic differential forms without first order poles (forms of first or second kind) over 1-dimensional cycles c are 0 "in general" (i.e. for most algebraic surfaces). Picard gave a detailed explanation of this phenomenon by an analysis of the first Betti number of a generic algebraic surface F. Starting from a singularity-free birational model of F in $P_5\mathbb{C}$ he derived a representation in projective three-space such that the resulting equation for F, $f(x, y, z) = 0$ (in inhomogeneous coordinates), leads to a 1-parameter family F_y of algebraic curves, which, with the exception of a finite set of values $Y = \{y_1, \ldots, y_k\}$, are of the same genus p. From the topological point of view Picard thus studied a fibration $F \to P_1\mathbb{C}$ with a closed oriented surface of genus p as

36 Compare Section 2.1.

37 (Noether, 1870, 1875) and (Picard, 1885, p. 282; 1886, p. 330).

generic fibre and a finite set of exceptional fibres with genus $< p$. By a beautiful blend of complex analytic and topological arguments, combining homotopy classes of closed paths in $P_1\mathbb{C} \setminus Y$, homology classes of 1-cycles in a generic fibre F_y, abelian integrals and the monodromy of the "Picard–Fuchs" differential equation describing the change of values of abelian integrals under change of y, he showed that in the generic case (for "most" F) all 1-cycles reduce homologically to only one already by monodromy constructions of boundary relations. Then by the observation that in each singular fibre F_{y_i} at least one cycle degenerates to a point, he argued convincingly that such a "vanishing cycle" is homologically trivial in F, and thus all 1-cycles are homologically zero. Picard started to use the same arsenal of methods to calculate the second Betti number of F, but did not get far in this attempt. Apparently the symbolical methods were not sufficiently elaborate to deal with this more involved situation before Poincaré entered the arena.

4. Passage to the theoretical stage

4.1. *Poincaré entering the field*

During the 1880-s Poincaré came across "manifolds" in several analytical or geometrical contexts, although he personally understood them at that time still in a rather vague way. One of these contexts arose from his work in the theory of automorphic functions that made him famous (and Klein nervous) at the beginning of the decade. The culminating problem of Poincaré's and Klein's research was the uniformization "theorem". Poincaré's points of departure were complex differential equations over algebraic curves

$$\frac{\mathrm{d}^2 \nu}{\mathrm{d}x^2} = \nu(x)\phi(x, y),$$

ϕ being a meromorphic function on an algebraic curve C given by $f(x, y) = 0$, which leads only to (finitely many) regular singularities.[38] If it could be solved by means of a pair of Fuchsian functions[39] $x(\zeta)$, $y(\zeta)$ taking as fundamental system the functions $\nu_1 = \sqrt{\mathrm{d}x/\mathrm{d}\zeta}$ and $\nu_2 = \zeta\sqrt{\mathrm{d}x/\mathrm{d}\zeta}$ (pushed down to C), Poincaré called the equation a *Fuchsian* differential equation.[40] The quotient $\nu_2/\nu_1 = \zeta$ was then the inverse of a universal covering map of the algebraic curve C, branched in the (regular) singularities of ϕ on C. Poincaré called two Fuchsian equations of the same *type* if there is a birational transformation between the underlying algebraic curves C and C' which transforms the singularities one into another such that the monodromy characteristics remain identical.[41]

For the sketch of a proof Poincaré (1884) collected all types of differential equations on an algebraic curve of given genus p and with given number k of branch points and

[38] p is *regular singular point* of the differential equation if a fundamental system of solutions can be chosen such that the quotient is a multivalued function branching over p of order k ($k \in \mathbb{N}$) or ∞.

[39] That is, ζ varies in the Poincaré half plane $\mathrm{Im}(\zeta) > 0$ and x, y are invariant under a properly discontinuous subgroup $G \subset PSL_2(\mathbb{R})$.

[40] For Fuchs's studies of the monodromy of regular singular differential equations compare [Gray, 1984; 1986, pp. 60ff.].

[41] That means, the difference of the characteristic exponents of two fundamental solutions of the equation is identical and of the form $1/k$ or 0 with $k \in \mathbb{Z} \setminus \{0\}$. In that case the quotient ζ is the inverse of a branched covering of branching order $|k|$ or ∞.

branching orders l_i $(1 \leqslant i \leqslant k)$ in a "multiplicité" M which in generic points could be characterized by $6p-6+2k$ (real) parameters. Analogously he parametrized the Fuchsian groups which lead to the proper genus p and given branching behaviour in another "multiplicité" M' (of the same dimension). Poincaré's version of the *uniformization theorem* then claimed that each type contains at least one Fuchsian differential equation.[42] To argue for the correctness of this claim he considered the map $g: M' \to M$ and showed that it is continuous and injective. The main point of the famous "continuity" proof was then to conclude the surjectivity of g from this information.

Poincaré gave a discussion which in fact spoke in favour of the surjectivity and was already sharper than Klein's, but still used highly intuitive ideas about continuous variation of images in higher dimensional spaces in a symbolically uncontrollable manner. Even the spaces themselves were not shown to be manifolds but taken as such, without further ado. For any critical reader (perhaps even including Poincaré himself) the "continuity proof" could thus be taken at least as much as an indicator for the necessity of an improved understanding of higher dimensional geometry as it was an indicator for the truth of the uniformization theorem. And in fact a clarification of the topological proof strategy was given only later by Brouwer (1911a, 1911b) who used enlarged (necessarily no longer uniquely) parametrizing spaces which indeed were manifolds and to which he could apply his domain invariance theorem for continuous injective mappings.[43]

Another context in which Poincaré gathered early experiences with higher dimensional manifolds arose from his investigation into the qualitative theory of differential equations. One of his questions was the topological classification of the singular points of a vectorfield (node, saddle point, focus, centre) and the introduction of the index as a numerical invariant. After having done so in the plane, he modelled nonlinear ordinary differential equations by the flow of a vectorfield v on (real) algebraic surfaces F and realized that the global index of the vectorfield $\mathrm{ind}(v)$ (i.e. the sum of the local or pointwise defined indexes) is equal to the Euler characteristic of the surface: $\mathrm{ind}(v) = 2 - 2p$ (Poincaré index theorem) (Poincaré, 1885).[44]

In an attempt to generalize the result to nonlinear differential equations of higher order he transformed that problem to a high dimensional system of first order equations. Then he started with a geometrization of the n-dimensional situation, although at the outset he considered geometrization as nothing more than a "useful language" (Poincaré, 1886, p. 168).

Fortunately he could build upon results of Kronecker (1869) about an analytically defined concept of index of functions systems on hypersurfaces of $\mathbb{R}^n$, which about the same time was being given a topological content by W. Dyck.[45] Dyck had shown that the Kronecker characteristic could be expressed in terms of his own purely topologically defined characteristic, which in fact was equivalent to the Euler characteristic and even equal up to sign.[46] Working with the Kronecker characteristic as a symbolical tool,[47] Poincaré was

42 In Poincaré's terminology: Each type is a "Fuchsian" type.

43 On the other hand, Brouwer made it clear that the moduli spaces used by Klein and Poincaré had singular points in curves with a nontrivial birational automorphism group, so that the argument of Poincaré and Klein turned out in fact to be unreliable in its original form.

44 Cf. [Gilain, 1991; Gray, 1992; Dahan, 1997].

45 Dyck published his first results on topological characteristics in the years 1885–1887 in the *Mitteilungen Sächsische Gesellschaft der Wissenschaften.*

46 [Scholz, 1980, pp. 249ff.].

47 Poincaré did not cite Dyck, whose publication he apparently did not know.

able to sketch the idea of a high dimensional version of the index theorem for vector fields, the later Poincaré–Hopf index theorem,[48] for the case of open submanifolds or hypersurfaces in $\mathbb{R}^n$. In this context Poincaré explicitly expressed the need for further elaboration of the methods to determine the higher orders of connectivity of Riemann and Betti (Poincaré, 1886, p. 448).[49]

Of course, there were other contexts in which Poincaré found an opportunity to come back to manifold ideas, for example in his studies of complex integration in two variables $\int f(\xi, \eta)\, \mathrm{d}\xi\, \mathrm{d}\eta$ with $(\xi, \eta) \in \mathbb{C}^2$. Poincaré showed that (in modern notation) $\mathrm{d}\omega = 0$ for $\omega = f\, \mathrm{d}\xi \wedge \mathrm{d}\eta$ and therefore the Cauchy theorem holds for two-dimensional integrals. Interestingly enough he still used the traditional language of "deformation" of one surface S into another S' through a region $A \subset \mathbb{C}^2$, in which the 2-form is analytic, although from the context it must have been clear to him that a homological concept of boundary relations was closer to the situation (Poincaré, 1887, p. 456). In his earlier experiences with high dimensional geometry Poincaré had been skeptical with regard to its usefulness, as he argued that spatial intuition would no longer be directly applicable. By the late 1880-s however, he had gathered sufficient material in different fields of his studies for him to accept that such "hypergeometrical" language of "multiplicités" are useful and perhaps even necessary for proceeding further with some of his analytical investigations.

4.2. *A constructive approach to manifolds*

The exclusion of a direct application of spatial intuitions would not exclude indirect application, mediated by a proper symbolical framework, which had been only roughly sketched by Riemann and Betti. That is what Poincaré started to pursue in the early 1890-s and continued to work on for the rest of his life, best expressed in his ground breaking series of articles on "analysis situs" (Poincaré, 1895, 1899, 1900, 1902a, 1902b, 1904). In this series Poincaré set the stage for a theoretical exploration and characterization of manifolds of any (finite) dimension which expanded so fruitfully and vastly in our century. Moreover, in the elaboration of the tools of analysis situs to make the "hypergeometry" of manifolds symbolically accessible, he brought combinatorial topology to the point where it could easily transcend the limits of manifolds and become a field of study of its own. Some traits of the theoretical and methodological achievements are outlined in the next section.

Poincaré, in accordance with his general philosophy of mathematics, did not use a formal, perhaps even axiomatic, definition of manifolds (which would moreover have been rather difficult to formulate in the 1890-s), but preferred to outline constructive procedures for the generation of manifolds. He used two main procedures to define a manifold M.

(1) In his first definition he described M as zero set $f^{-1}(0)$ of a differentiable function $f : A \to \mathbb{R}^k$, with A open subset in $\mathbb{R}^{n+k}$, defined by inequalities, and the Jacobian $\mathrm{d}f(x)$ of maximal rank for all $x \in A$. He admitted that M might have a *boundary*. More clearly than his predecessors Poincaré explicitly used the rank condition to derive local parametrizations of M. In addition he explained the morphisms under which two such representations M and M' should be considered as equivalent, as

[48] (Hopf, 1926). An intermediate step was taken by Brouwer in his work on the index of vectorfields on n-dimensional spheres (Brouwer, 1911b, pp. 107ff.); cf. [Johnson, 1987, p. 82].

[49] Ironically Poincaré got the name wrong speaking about "Brioschi" where he obviously referred to Betti's work (Poincaré, 1886, p. 448), showing that he just started to assimilate the subject.

diffeomorphisms of open neighborhoods of M, respectively M', in their embedding real space, which map M onto M' or vice versa. In his terminology he did not even indicate the possibility of a distinction between a general topological and a differential topological structure, although he used the terminology of "homéomorphismes" (Poincaré, 1895, pp. 196ff.). In fact Poincaré alluded to Klein's Erlanger program and called the groupoid of his diffeomorphisms a "group", which should define the branch of geometry called "analysis situs", as he saw it (ibid., p. 198).

(2) The second main definition allowed for a finite set (we would say atlas) of differentiable regular parametrizations of M by domains V_i in $\mathbb{R}^n$. Poincaré considered M as covered by sets U_i which all were subsets of $\mathbb{R}^m$ $(m \geqslant n)$, without considering M as a subset of $\mathbb{R}^m$: $M = \bigcup_{i \in I} U_i$ with parametrizations $\Theta_i : V_i \to U_i$ $(1 \leqslant i \leqslant l)$ and regular change of parametrization in n-dimensional components of overlaps of the U_i-s (and with a similar definition in the complex case).[50] Poincaré concentrated attention in this definition on the analytic case and used the terminology of "analytic continuation" for the description of change of parameters in overlapping regions (Poincaré, 1895, p. 200). In this case he introduced the *orientability* of M by the condition of positive functional determinant for changes of parametrization.

Of course, Poincaré did not exclusively consider these main definitions, but explained how to derive local parametrizations from the first definition, discussed diverse examples of mixed constructions, e.g., by restriction of the parameter sets V_i to lower dimensional submanifolds of the parameter space $\mathbb{R}^n$, defined by method (1). Even the operation of a finite group leaving the parametrizing submanifold invariant was included, as in his description of an image of the real projective plane $P_2\mathbb{R}$ in $\mathbb{R}^6$. [51]

Depending on the context of investigation, Poincaré later introduced additional construction procedures, which presupposed that the resulting object satisfied definitions (1) or (2).

(3) The most important of these additional constructs was the cell subdivision and the representation of M as a finite geometric *cell complex* a "polyèdre", which by definition satisfies the local manifold condition (Poincaré, 1895, pp. 270ff.). He used it inter alia for constructing manifolds with prescribed fundamental group by boundary identification rules, although restricted to the three-dimensional case, where the local manifold conditions about identified 0-cells could be controlled by a combination of symbolic representation and basic space intuition (Poincaré, 1895, pp. 229ff.).[52]

(3′) In the fifth complement (1904) Poincaré introduced even more construction procedures, a "skeleton" representation of 3-dimensional manifolds, containing some ideas of three-dimensional Morse theory (Poincaré, 1904, pp. 475ff.), and an adaptation of an idea of P. Heegard to form a closed 3-manifold M by boundary identification of two homeomorphic handle bodies V and V'.[53] He used these procedures

[50] Poincaré would not read M as literal union of the U_i. In lower dimensional components of intersection $U_i \cap U_j$ he thought in terms of a disjoint union, only in n-dimensional components he would identify the respective points of U_i and U_j; therefore he did not treat M globally as subset of $\mathbb{R}^m$. In more recent terminology, M is a manifold, while the immersion used in Poincaré's construction is not necessarily injective, and thus the image $\widetilde{M}$ no (sub-) manifold of $\mathbb{R}^m$.

[51] Poincaré used a parametrization of $P_2\mathbb{R}$ by $S^2 \subset \mathbb{R}^3$ with antipodal identification.

[52] These examples are discussed in detail in [Volkert, 1994, pp. 87ff.]; compare also [Volkert, 1997].

[53] Compare [Volkert, 1994, pp. 137ff.].

to present the famous Poincaré "dodecahedral space" M with trivial homology but fundamental group isomorphic to the (extended) dodecahedral/icosahedral group I^* (Poincaré, 1904, pp. 478ff.).

Poincaré was convinced that each manifold given by definitions (1) or (2) can be represented as a (finite) "polyèdre" as in definition (3). His arguments in favour of that conviction were, however, more founded on intuitive "optimism" than on critical evaluation of the question (Poincaré, 1899, pp. 332ff.). So Poincaré claimed to have a positive solution for what later was considered to be a basic problem for the clarification of the conceptual structure of the topological theory of manifolds, the *existence of triangulations* for differentiable or topological manifolds.[54] A similar evaluation can be given for his use of the principle that it is always possible to find a common subdivision of two given finite cell subdivisions of a manifold (Poincaré, 1895, p. 271), which later became the *Hauptvermutung* in the terminology introduced by Kneser.[55]

Already from this short presentation it may become apparent that Poincaré's constructive concept of manifold contained an arsenal of methods to build examples to enrich the understanding of the world of new geometric objects. Although he did not even attempt to give a formal analysis and unified delimitation of the concept, Poincaré's work was thus highly effective and gave a tremendous push towards a more refined understanding of the general concept outlined by Riemann and so difficult to understand in the second half of the 19th century.

4.3. *Giving a theoretical status to the topology of manifolds*

These examples of manifolds constructed and considered by Poincaré served as material for the exploration and development of methods to analyze their intrinsic *analysis situs* nature. Poincaré's work is, of course, much better known by its contribution to these methods than by the elaboration of the basic material of manifolds.[56] In fact, Poincaré presented two approaches to analyze the homology of manifolds, the first followed Riemann and Betti rather directly and was introduced in the opening work of the series (Poincaré, 1895). The second one with a presentation and elaboration of the homology of cell complexes was the subject of the first two complements (Poincaré, 1899, 1900). Moreover he introduced the fundamental group of manifolds already in (Poincaré, 1895) and constructed diverse examples of 3-dimensional manifolds with prescribed fundamental group. These more elementary examples were superseded by the elaborate case of the "dodecahedral" space in the fifth complement (Poincaré, 1904). In the two intermediate supplements he developed methods to calculate the homology of algebraic surfaces (Poincaré, 1902a, 1902b). Diverse detailed historical studies deal with different aspects in Poincaré's topological work;[57] here I only want to outline the homological part of the profile of the theory which Poincaré pro-

[54] Compare [Kuiper, 1979].

[55] Compare [Volkert, 1994, p. 164] and other contributions in this volume.

[56] For the latter aspect see [Volkert, 1994].

[57] For the homological aspects of Poincaré's work consult primarily [Bollinger, 1972] and in addition [Dieudonné, 1989, 1994] and from a semiotic point of view [Herrmann, 1996], for the homotopic aspects [Vanden Eynde, 1992], for specific construction methods of low 3-dimensional manifolds [Volkert, 1994], for a discussion of the contribution to the manifold concept and an outline of Poincaré's topological study of algebraic surfaces [Scholz, 1980]. Compare also diverse other contributions to this volume.

posed in order to make the "hypergeometry" of manifolds accessible. Poincaré's introduction of the fundamental group is discussed in the article of Vanden Eynde (this volume).

In his first approach to homology in a n-dimensional manifold M Poincaré followed Riemann's proposal to study bordance relations of *oriented* submanifolds $V_1, \ldots, V_k$ of given dimensional $m \leqslant n$. His procedure was conceptually still highly intuitive and vague, as the underlying idea supposed the study of equivalence relations on the set of all m-dimensional submanifolds and was too complicated to get hold of, with the methods available at the time. Moreover, to make the approach feasible, the "submanifolds" should admit certain "nicely behaving" singular subsets, like the topological (or smooth) "varieties" recently proposed by (Kreck, 1998). Poincaré, however, went a step further than his predecessors in the symbolical description of his objects and relations. In particular, he introduced an algebraic representation,

$$V_1 + V_2 + \cdots + V_k \sim 0,$$

for the condition that all the V_i form a complete boundary of an $(m+1)$-dimensional submanifold and transformed such homology relations by addition and multiplication of the terms with integer coefficients. In this approach the terms were of a peculiarly ambivalent semiotic nature. Basically, Poincaré interpreted terms like λV_i $(\lambda \in \mathbb{Z})$ as a collection of λ "slightly varied" copies of the (oriented) submanifold V_i; but he accepted and used a formal division of homologies,[58] as a result of which the homologies no longer *directly* had to express boundary relations.[59] In the result Poincaré got an interesting symbolical system for homologies and the calculation of Betti numbers p_i,[60] which allowed him to explore basic features of the homology of manifolds much deeper than his predecessors, in particular duality for the Betti numbers, $p_i = p_{n-i}$ for orientable closed manifolds, and the Euler–Poincaré theorem $\chi(M) = \sum_{i=0}^{n}(-1)^i p_i$.

None of these could be proven indubitably in Poincaré's approach. For the duality theorem his calculation of the intersection numbers remained highly intuitive, as the differential topology of general transversal intersections was too involved to be clarified by his means. For the generalized Euler theorem Poincaré used his principle of the existence of a common refinement of two finite cell decompositions of the manifold M (the later *Hauptvermutung*). So, from a critical point of view, both principles (Poincaré duality and Euler–Poincaré) had rather the status of well motivated conjectures than of "theorems", even in the eyes of critical contemporaries like Heegard, Dehn, Tietze et al.

After Heegard's criticism of the discussion of duality in manifolds, Poincaré established his second, much better algebraicized combinatorial method to define and calculate connectivity numbers, adding torsion numbers and coefficients to the Betti numbers (Poincaré, 1899, 1900).[61] He started from a representation of the manifold M as a geometric cell complex constituted by q-cells $a_i^{(q)}$ $(1 \leqslant i \leqslant \alpha_q$ for all dimensions $0 \leqslant q \leqslant n)$, and described

[58] He made the "division rule" explicit in the first complement answering P. Heegard's criticism of having suppressed torsion elements. In (Poincaré, 1895) it was subsumed under a sort of "metarule" for homologies: "Les homologies peuvent se combiner comme des équations ordinaires" (Poincaré, 1895, p. 207).

[59] Compare the often discussed example of the line l in $P_2\mathbb{R}$ with $2l \sim 0$ having a direct geometric interpretation as a small angular segment U between two lines, $\partial U = 2l$; whereas the result of division $l \sim 0$ had no longer direct geometric interpretation, as criticized by Heegard.

[60] To be precise, Poincaré used a slightly changed definition of Betti numbers $P_i := p_i + 1$, if p_i is the maximal number of homologically independent i-cycles ("with division", i.e. calculating with integer coefficients in $\mathbb{Q}$).

[61] Compare [Bollinger, 1972].

boundary identifications as "congruences", $a_i^{(q)} \equiv \sum_j \varepsilon_{ij}^{(q)} a_j^{(q-1)}$, codified by the matrices $E^{(q)} = (\varepsilon_{ij}^{(q)})$, and reduced the consideration of cycles and their boundary relations to those expressible in linear combinations of cells. That allowed him, of course, to avoid the difficult problems arising from investigation of all submanifolds and led to the well-known approach of combinatorial topology. Poincaré could thus very well show that the most evident difficulties arising from his first approach resulted geometrically from nonorientability of the manifold M and algebraically from the ambivalence between homology "with division" (we would say calculating the homology over $\mathbb{Q}$) and "without division" (over $\mathbb{Z}$).

Poincaré (1900) presented a new definition and a calculus for the calculation of Betti numbers and torsion from the incidence matrices $E^{(q)}$ of a cell decomposition of M. The method used diagonalization of incidence matrices by elementary transformations to matrices $T^{(q)}$. Expressed in slightly more structural terms Poincaré developed a calculus to choose generators of the $\mathbb{Z}$-module C_q of cellular q-chains such that all boundary operators $\partial_q : C_q \to C_{q-1}$ are diagonalized. That allowed him to read off immediately the Betti numbers and torsion coefficients and the distinction between manifolds "with" or "without" torsion from the diagonalized matrices $T^{(q)}$ (Poincaré, 1900, p. 369).

Poincaré was sure that his second, the combinatorial, method led to the same homological invariants (Betti numbers and torsion coefficients) as the first, bordance of submanifolds, method. He first showed that a subdivision of the "polyédre" does not change the combinatorial invariants (1899, pp. 303ff.). Considering now a set of submanifolds, arising in the representation of cycles and homologies of the first method from the "principle" of the existence of cell subdivision for each of them and the assumed possibility of constructing a common subdivision ("Hauptvermutung"), he concluded without any hesitation that the "old" and the new (combinatorially defined) homological invariants are identical. This part of the "proof" needed only *six lines* in his presentation (Poincaré, 1899, p. 309). Although he thus got new problems he could not solve or even realize, he achieved on the other hand a proof of duality for orientable manifolds and the generalized Euler theorem in the symbolically clear framework of the new approach (1899, pp. 302f.; 313ff.).

In the end Poincaré had achieved a lot for a homological theory of (differentiable compact) manifolds about the turn of the century. He had introduced the old invariants (Betti numbers) in a new, much clearer symbolical framework, had introduced new ones (torsion coefficients), developed a well algebraicized calculus to compute them, calculated them in a great variety of cases, and proven two basic theorems (duality, Euler–Poincaré). Moreover he had introduced and given a basic analysis of the topological importance of the fundamental group, which is put into the context of the development of homotopy ideas in the contribution of R. Vanden Eynde in this volume. Thus, even taken into consideration that Poincaré took basic principles to be valid without any hesitation (triangulability, Hauptvermutung), that turned out to contain serious problem potential for the future clarification of basic structures of the topology of manifolds during the century to come, there can be no doubt that he was the main initiator of a topological theory of manifolds of wide range.[62] Moreover, the elaboration of his second (combinatorial) approach to homology opened the path towards a homological theory of more general topological spaces.

62 This advancement tends to be suppressed in Dieudonné's discussions of Poincaré, as he looks at the latter rather with the eyes of a "modern" mathematician in the sense of the 20th century than with those of a historian.

5. Elaboration of a logical frame for the modern manifold concept

5.1. *Early axiomatic attempts for two-dimensional manifolds*

Topological spaces on different levels of generalization were analyzed in different approaches and with varying degrees of precision in the rise of modern mathematics in the early 20th century. During the last three decades of the 19th century Cantor had developed his theory of point sets in $\mathbb{R}^n$ in the framework of general set theory. He himself was shocked to realize that bijective maps between real continua of different dimensions can be conceived, and even Dedekind's comforting conviction that more specific maps, in this case bijective and (bi-)continuous ones, would respect the *invariance of dimension* left the problem to prove (or disprove) such a conjectured invariance. Naive assumptions from space intuition were particularly deceptive in this field; that became even clearer about 1890 when Peano published his example of a "spacefilling" curve with the surprising effect, that the lack of injectivity would even for continuous maps not necessarily lead to a decrease of dimension (or keep it at most invariant), but could as well *increase* it. Early attempts by Lüroth, Thomae, Netto, and Cantor himself, to prove the invariance of dimension under bijective continuous maps, turned out to contain unclosable gaps and again (as in the case of the continuity proof for uniformization) it was only Brouwer who surmounted the difficulties and indeed proved the correctness of Dedekind's suggestion (Brouwer, 1911a).[63]
About the turn of the century two methodological strategies for clarifying the concept of manifold were formed and sketched, an axiomatic one proposed by Hilbert, taken up by Weyl (about 1913), Hausdorff, H. Kneser, and Veblen/Whitehead, and a constructive one proposed by Poincaré, taken up by Dehn/Heegard, Tietze, Steinitz, Brouwer, Weyl (after 1920), Vietoris, van Kampen and others.

The first attempts at an axiomatic formulation of manifolds by Hilbert and by Weyl (1913) were limited to dimension 2 by contextual considerations. They contained a blend of early ideas of general topology and postulates for regular coordinate systems as specific manifold structures. Hilbert's approach (1902a, 1902b) arose from the context of the foundations of geometry and had as its main goal the erection of an axiomatic framework for the concept of a (simply connected) two-dimensional continuous manifold which should serve as a starting point for a group theoretic characterization of the principles of Euclidean geometry.

Hilbert supposed the *plane* E to be topologized by a sufficiently rich system of *neighbourhoods* ("Umgebungen") $\mathcal{U}_p$ of each point $p \in E$, formed by sets $U \subset E$ containing p and each complemented by at least one coordinate bijection $\psi : U \to V$ onto a Jordan domain $V \subset \mathbb{R}^2$, such that the four following conditions hold:

(1) For each Jordan domain $V' \subset V$ containing $\psi(p)$ the counterimage $\psi^{-1}(V')$ is also a neighbourhood of p.
(2) For two coordinate bijections ψ and ψ' of the same neighbourhood U onto V and V' the coordinate change $\psi'\psi^{-1} : V \to V'$ is bijective and continuous.
(3) A neighbourhood V of $p \in E$, containing a point q, is also a neighbourhood of q.
(4) Each two neighbourhoods V, V' of p contain another one $V'' \subset V \cap V'$.
(5) To any two points $p, q \in E$ there exists a common neighbourhood V.

[63] For the history of invariance of dimension see [Johnson, 1979/1981, 1987] and for an outline of Brouwer's proof [Koetsier and van Mill, 1997].

Hilbert commented that his postulates contain, as he thought, the "precise definition of the concept, which was called *multiply extended manifold* by Riemann and Helmholtz and *number manifold* by Lie" (Hilbert, 1902a, p. 233). This remark of Hilbert was, like so many others in the foundations of mathematics, a bit rash but showed a promising way to proceed.

In fact, Hilbert's sketch of an axiom system for two-dimensional manifolds containing all the conceptual components for the later refinement of both, the characterization of general topological spaces, by what would later be called a neighbourhood basis as formulated by Hausdorff (1914)[64] and the axiomatic definition of manifolds by coordinate systems and a regular atlas as elaborated by Veblen and Whitehead (1931). Hilbert dealt, however, with both aspects in a simplified form justified by his restricted context. The later Hausdorff separability was indirectly implied by his last axiom of "big" coordinate neighbourhoods to any two points $p, q \in E$ and their separability in the coordinate plane by Jordan regions (in addition to axiom (1)).

Weyl, in his *Idee der Riemannschen Fläche*, could already build upon Brouwer's result of the invariance of dimension under bijective continuous maps between open sets in $\mathbb{R}^n$. That may have given him the confidence that a slightly more "intrinsic" characterization (than Hilbert's) of a two-dimensional manifold was possible.

Like Hilbert he characterized the structure of a two-dimensional manifold F by a system of neighbourhoods $\mathcal{U}_p$, each of which, $U \subset F$ would contain p and be supplemented by a bijective map $\psi : U \to V \in \mathbb{C}$, with V an open disk with center $\psi(p)$. The totality of neighbourhoods was used by Weyl as a neighbourhood basis for the topology of F in the modern sense. He demanded that they satisfy two conditions. The first one amounted to what would be expressed in more recent terminology as (i) the *open map* condition for the coordinate map ψ with respect to the topology on F induced by the neighbourhood basis. The second one was: (ii) for any neighbourhood U of a point $p \in F$ with coordinate map $\psi : U \to V$ and a small disk $V' \subset V$ of center $\psi(q)$ $(q \in U)$, there is a neighbourhood U' of q such that $\psi(U') \subset V'$.

The second postulate had a double function in Weyl's argument; it made sure that coordinate maps were continuous and it secured the existence of sufficiently many "neighbourhoods" to constitute a neighbourhood basis (from our point of view). Essentially Weyl characterized a manifold F as a topological space by the assignment of a neighbourhood basis $\mathcal{U}$ in F, postulating that all assigned neighbourhoods $U \in \mathcal{U}$ are homeomorphic to open balls in $\mathbb{R}^2$. That was, of course, a remarkable contribution to the clarification of what is essential for an axiomatic characterization of manifolds. Weyl left, however, a gap, which was not surprising for the time. He dropped Hilbert's axiom (5) to achieve stronger localization than his former teacher; but he did not realize that separability of points by neighbourhoods was thus lost. So it was left to Hausdorff, the more acute thinker with respect to logical clarification of concepts, to pinpoint the necessity of such an additional postulate in his axiomatization of topological spaces (Hausdorff, 1914, p. 213).[65]

64 Compare [Scholz, 1996; Aull and Lowen, 1997].

65 Weyl was relatively slow to accept the necessity of a separability axiom for manifolds. He did *not* supplement or change his axiomatics of two-dimensional manifold in the second edition of (Weyl, 1913) in 1923 and did so only for the third edition in 1955. In the middle of the 1920-s he had accepted the importance of this Hausdorffian specification (Weyl, 1925/1988, p. 3). I owe R. Remmert the hint at the this gap and the relatively late correction in Weyl's approach.

At the end of his book on the foundation of analysis *Das Kontinuum* Weyl experimented with a modified axiomatization of the concept of two-dimensional manifolds from a constructive perspective. Now he worked with a restricted real continuum, the *Weylian reals* $\mathcal{W}$, constructed by only those Dedekind cuts in $\mathbb{Q}$ that are definable in a semiformalized arithmetical language (essentially using first order predicate logic and recursive definitions over $\mathbb{N}$) (Weyl, 1918, pp. 80ff.).[66] He postulated a "somehow" constructively given countable base of nested neighbourhoods $U_{p,n}$ $(n \in \mathbb{N})$ with $U_{p,n+1} \subset U_{p,n}$ for a countable "dense" net of points $p \in X \subset F$, each $U_{p,n}$ bijectively bicontinuous with an open disk in the Weylian number plane $\mathcal{W}^2$. But for philosophical reasons he was discontented with his new approach just from the beginning.[67] A little later he turned towards Brouwer's approach in the foundation of mathematics, even if only for a while, and essentially became an adherent of a constructive (combinatorial) approach to manifolds.

5.2. *The rise of the combinatorial and piecewise linear approach*

Other mathematicians had already started to pursue such another, more constructive approach to a modern formulation of the manifold concept, following Poincaré's decomposition of manifolds into geometric cell complexes ("polyèdres"). Already Dehn and Heegard in their article on *Analysis Situs* for the *Encylopädie der Mathematischen Wissenschaften* emphasized the combinatorial construction of manifolds, which was intended as a *definition*, not as a reconstruction of an object that had already been given in another way. In consequence they explicitly introduced the idea that morphisms of these objects should be defined by combinatorial equivalence[68] rather than by (bicontinuous) homeomorphism. Such an approach was also chosen by H. Tietze in his Habilitationsschrift in which he studied manifolds as n-dimensional cell complexes up to combinatorial equivalence. To specify manifolds among more general cell complexes he postulated that the star of each m-dimensional cell C^m, i.e. the union of all higher dimensional cells that intersect the boundary of C^m be *simply connected*, by which he understood that it is combinatorially equivalent to a sphere S^{n-m-1} (Tietze, 1908, p. 24).[69] He left open, however, how such an equivalence could be identified.

As a great advantage of this approach Tietze observed that it would lead to a foundation of analysis situs independent of the consideration of infinite sets and their inherent logical difficulties and methodological subtleties (1908, p. 2).[70] As a contribution to such subtleties (at least from the point of view of Poincaré) he discussed cell decompositions of the same manifold M with infinitely many components of the intersection of cells. Thus he showed that Poincaré's conviction that each two (finite) cell decompositions have a common subdivision was too rash to be accepted. He admitted that a proof of the existence of

66 [Feferman, 1988; Coleman and Korte, 1998].

67 Compare [Scholz, 1998].

68 Dehn and Heegard used Möbius' terminology of "elementary relationship (Elementarverwandtschaften)" (Dehn and Heegard, 1907, pp. 159f.).

69 Tietze used the terminology of "homeomorphism" to S^{n-m-1}, but made it clear that he understood in this context "homeomorphism" in the sense of combinatorial equivalence (Tietze, 1908, p. 13).

70 At the time of publication of Tietze's studies the principle of choice, which had been used by Zermelo (1904) a little earlier and explicitly introduced as an axiom of set theory the same year (Zermelo, 1908a, 1908b), led to intense debate and controversian reactions among mathematicians in France and Germany. Cf. [Moore, 1978, 1982].

such a common subdivision "might be relatively simple in the case of two dimensions", but that it "waits for a deeper investigation (harrt einer eingehenderen Erledigung) in the general case of higher dimensions (Tietze, 1908, p. 14). So he openly posed the question, whether two homeomorphic manifolds M and M' are in fact combinatorially equivalent, as an important problem of the theory. In the 1920-s H. Kneser emphasized the methodologically central role of this conjecture even more strongly and gave it the famous name of *Hauptvermutung* for the combinatorial theory of manifolds (Kneser, 1926, p. 6).

In another publication of the same year E. Steinitz attempted an axiomatic foundation of combinatorial topology by a set of postulates for the incidence structure of abstract finite cell complexes constituted by a finite set of "elements" a graded by "dimension" $\dim a = [a]$ $(0 \leqslant [a] \leqslant n)$ and with prescribed incidence relations. After the introduction of six axioms to regulate the concept of an abstract combinatorial polyhedron Steinitz added three more to specify what he considered as "combinatorial manifolds". Strangely enough he only postulated connectedness (axiom 7), existence of bounding cells for cells of intermediate dimension $(2 \leqslant [a] \leqslant n-2)$ and connectness of the boundary set (axiom 8) and existence of incident cells $[c]$ of each intermediate dimension $[a] < [c] < [b]$ to each two incident cells a and b of dimensional difference at least 3 (axiom 9) (Steinitz, 1908, pp. 37f.).[71] Of course, Steinitz also introduced a combinatorial concept of equivalence for his abstract cell complexes (and "manifolds"); but although his axiomatization broke new ground for an abstract approach towards combinatorial topology in general, his characterization of manifolds was much too weak to be accepted or of broader influence for future research. So it was in fact Brouwer's highly influential introduction of a "mized approach" of combinatorial and continuity methods, in which manifolds were defined by simplicial methods, that marked the next remarkable leap for a constructive underpinning of the manifold concept. It also pointed out in which direction one had to go if manifolds should be selected among the more general objects of abstract combinatorial complexes.[72]

Brouwer seems to have detected the importance of simplicial decomposition of manifolds, of simplicial approximation, and of mapping degree for the investigation of long standing problems in the topology of manifolds early in 1910.[73] He introduced his new tools of simplicial approximations and the mapping degree of continuous maps between manifolds in his famous publication (Brouwer, 1911b). There he defined manifolds in a manner adapted to his context of the simplicial methodology. He explained an n-dimensional manifold M to be a (possibly) infinite[74] *geometric simplicial complex* of dimension n such that:

(1) two intersecting n-simplexes share a p-dimensional face $(1 \leqslant p \leqslant n-1)$ and with it all lower dimensional faces of the latter,
(2) for each vertex the collection of incident simplexes is homeomorphic to an n-ball (Brouwer, 1911b, p. 97).

[71] Compare [Volkert, 1994, pp. 173ff.].

[72] For an outline of how Brouwer's intuitionism and his topological constructivism went in hand see [Koetsier and van Mill, 1997].

[73] See Freudenthal's evaluation of an unpublished notebook of Brouwer in (Brouwer, 1976, pp. 422–425); compare also [Johnson, 1987, pp. 81ff.].

[74] In the case of a finite simplicial decomposition he called the manifold "closed" (in our terminology compact), in the infinite case "open".

If all n-simplexes of an n-dimensional manifold M are represented by $n + 1$ homogeneous coordinates of a standard simplex in the $(n + 1)$-dimensional "number space",[75] such that lower dimensional simplexes carry identical coordinates from each n-simplex to which they belong, Brouwer called the manifold "measured", in more recent terminology M carries a *piecewise linear* (abbreviated PL-) structure. By a recursion procedure over the dimension he showed that a manifold in his sense can always be "measured" (given a PL-structure). That allowed him to characterize orientability and orientation of his PL-manifolds, barycentric subdivisions, simplicial approximation of continuous maps, and the mapping degree of continuous maps between manifolds. That served as the basis for his investigation of the index of vectorfields on spheres, his fixedpoint theorem, the proof of the invariance of dimension, etc.[76] Thus he introduced a new approach for a constructive characterization of manifolds besides the less standardized representations as geometric cell complexes in the sense of Poincaré, Tietze et al.

Brouwer's approach to manifolds combined a constructive representation of the global structure by "measured" simplicial complexes with a criterion of local simplicity, which still referred to pointset topological properties of "numberspaces" and did not even attempt to transform the latter to combinatorial criteria. In this respect Tietze (and Steinitz) had been more consequential in their attempt to avoid the fallacies of pointset topology. They had successors to elaborate more in detail, what Tietze had left open in his all-inclusive characterization of "simple connectedness" of neighbourhoods of k-cells. In the early 1920-s Veblen and Weyl pushed this characterization a step further, although they were not completely successful in their search for a convincing and operative characterization.

5.3. *Manifolds in the methodological "battles" of the 1920-s*

Veblen followed in his *Analysis Situs* (1922) the combinatorial approach to manifolds and complemented it by ingredients from Brouwer's simplicial constructs. Like Tietze he explicitly tried to avoid pointset topological considerations as far as possible.[77] He modified Tietze's combinatorial definition by a rather pragmatic reduction of the combinatorial problem to characterize "simply connected" stars of k-cells in an n-dimensional complex.[78] After giving three procedures to build an n-complex combinatorially equivalent to an n-cell,[79]

---

[75] I deliberately use Brouwer's original terminology and do not write $\mathbb{R}^n$, as Brouwer's terminology leaves the interpretation of the number continuum open. It can be interpreted by classical real numbers, Brouwer's inituitionistic real continuum, or even (later) by Weylian reals $\mathcal{W}$ of 1918.

[76] Cf. [Johnson, 1987; Koetsier and van Mill, 1997].

[77] "... we leave out of consideration all the work that has been done on the point-set problems of analysis situs and on its foundation in terms of axioms or definitions other than those actually used in the text." (Veblen, 1922, p. vii). In consequence Hausdorff did not start to read Veblen's and other mathematicians' work on combinatorial topology before the late 1920-s when Alexandroff's approach to homology via "nerves" of open coverings allowed an algebraization of homology which was directly applicable to topological spaces independent of a combinatorial structure.

[78] In addition Veblen choose to give the explanation of a "neighbourhood" of a k-cell $a^{(k)}$ in an n-dimensional complex C_n a surprising shift towards pointset theoretic criteria. He characterized such a "neighbourhood" as any set S of nonsingular cells of C_n such that all point sets $M \subset C_n$ with a limit point on $a^{(k)}$ have points in S (Veblen, 1922, Chapter III.4).

[79] The simplest of these three procedures was of course the following: Two nonintersection n-cells $a_1^{(n)}$, $a_2^{(n)}$ incident with exactly one $(n - 1)$-cell $a^{(n-1)}$ constitute an n-cell.

Veblen defined an *n-dimensional manifold* to be a closed (finite) regular cell complex of dimension n, in which each k-cell has a "simply connected" star, where "simple connectedness" was defined by a combination of the three construction procedures of k-cells given before.[80]

H. Weyl was at that time highly impressed by Brouwer's ideas of "free choice sequences" to characterize continuum ideas mathematically without reference to the conceptual framework of transfinite sets. So he experimented at the end of his polemical article on the *new foundational crisis of mathematics* with a characterization of point localization in a two-dimensional continuum by "free choice sequences" of nested stars in an infinite series of barycentric subdivisions of a two-dimensional Brouwerian manifold (Weyl, 1921, p. 177f.). He tried to come to a genetic definition of points in a two-dimensional combinatorial continuum and rejected the idea that "points" might be presupposed as ideal "atomistic" local determinations in advance.[81] In that respect the "purely" combinatorial approach to manifolds appeared to him of high importance for the foundations of mathematics, the more so as he could not be sure that Brouwer's intuitionistic continuum (which in the early 1920-s was not yet well elaborated in technical details) and his own ideas on an "infinitesimal continuum" would conceptually coincide after sufficient symbolical elaboration. In any case, it would be logically preferable to free the combinatorial approach to continuum concepts from the direct link to number concepts, which was presupposed in Brouwer's "mixed" approach to manifolds.

Thus Weyl took the opportunity of his visit to Madrid and Barcelona in early 1922 not only to elaborate his ideas on the "analysis of the space problem" from the new viewpoint of his infinitesimal geometric approach but also to give an exposition of his view of combinatorial topology for the *Revista Matematica Hispano–Americana* (Weyl, 1923, 1924). He developed his own approach to a characterization of a combinatorial n-sphere (a "Zyklus" as Weyl said) by two groups of axioms. As there was neither a semantically complete axiomatic characterization of combinatorial spheres, nor a complete set of construction procedures for the latter in sight, Weyl proposed for the time being a provisional axiomatic characterization of structural properties of combinatorial spheres "from above" in a first group of axioms, and in addition a second group of axioms, which gave a collection of genetic procedures for the construction of combinatorial k-spheres "from below". He hoped for a step by step completion of the axiom system in future research, which in the end might lead to a coextensive characterization of combinatorial k-spheres by any of the (extended) two groups of axioms, and thus of manifolds.[82]

The first group of axioms for a Weylian combinatorial n-sphere Z^n were the postulates that Z^n be connected (axiom 1), that to each k-cell $a^{(k)}$ in Z^n $(0 \leqslant k < n)$ the collection of all higher dimensional cells $b^{(j)}$ bounding directly or indirectly $(k < j \leqslant n,\ a^{(k)} \subset \bar{b}^{(j)})$ carries the combinatorial structure of a Weylian combinatorial sphere Z^{n-k-1} (axiom 2),[83] that it be orientable (axiom 3), and homologically trivial in dimensions less than n (axiom 4).[84]

[80] (Veblen, 1922, Chapter III.24).

[81] For more details on Weyl's philosophical motivation and the context of his rejection of transfinite set theory as a background in which to model "continuum" ideas, compare [Scholz, 1998].

[82] Cf. (Weyl, 1924, pp. 416f., 419) and also (Weyl, 1925/1988, p. 10).

[83] Weyl called this (modified) Tietzean property of a cell complex to be "unbranched".

[84] Weyl called this a "plain (schlicht–de una hoja)" complex (Weyl, 1923, p. 403).

For the "genetic" characterization of combinatorial n-spheres Weyl characterized the 0-sphere as two points (axiom 0), generation of n-spheres from n-spheres by subdivision of a rather general kind[85] up and down (axioms A, B) and two constructions of higher dimensional spheres from lower dimensional (axioms C, D). Weyl proved several results for combinatorial manifolds, in particular Poincaré duality for closed orientable manifolds. But his approach was probably too involved in foundational considerations and technically too sophisticated to be taken up as a convincing strategy for the elaboration of a more broadly accepted genetic concept of manifold, which stood up to the standards of modern mathematics.[86] So the research strategy proposed by Weyl was not taken up by other mathematicians, but at the best selectively adapted to other methodological views.

About the middle of the 1920-s all ingredients for a satisfying formulation of the manifold concept, taking up the knowhow on axiomatization and on genetic characterizations of manifolds were at hand. There is no point in repeating here the interesting history of the preparation and elaboration of the general concept of topological space.[87] After Hausdorff, Fréchet, and Riesz opened this new field of investigation, it found particularly active supporters in the newly rising mathematical groups in Poland, the Soviet Union, and the United States, and it also left its imprint on the modern reframing of mathematics in Germany and Austria. The first attempt for balance between the different approaches to the manifold concept was given by the young Hellmuth Kneser, who had written his dissertation with Hilbert in 1921 and got a professorship in Greifswald in 1925, in an article for the *Jahresbericht der DMV* (Kneser, 1926).

Kneser discussed both approaches, an axiomatic one based on Hausdorff's set theoretic foundations for topology, and a combinatorial one referring to, but deviating from Weyl's approach. For the axiomatic characterization of manifolds he limited himself to the topological case, without any discussion of differentiable structures. He thus characterized a *topological manifold* M by Hausdorff's axioms for a neighbourhood basis (of a Hausdorff space) including the second countability axiom for a neighbourhood base of all points in M (thus of all open sets in M) and added just one postulate: Each point $p \in M$ has a neighbourhood which is topologically equivalent to an open ball in the "n-dimensional numberspace", by which he obviously referred to the $\mathbb{R}^n$ (Kneser, 1926, pp. 1–3). A "closed" (in our terminology compact) manifold was characterized by the Heine–Borel criterion for open coverings of M.

After the introduction of a combinatorial decomposition of a "closed" manifold as a finite cell complex Kneser introduced the *Hauptvermutung* as fundamental for the combinatorial theory of manifolds and proposed a characterization of a *combinatorial manifold* by the simultaneously inductive definition of the concepts of n-dimensional cell complex C^n, n-dimensional combinatorial sphere S^n, n-dimensional cell E^n, the boundary of E^n, and internal transformations (allowed subdivisions of cells of S^n-s). He first defined by

---

[85] For subdivision of an n-cell E^n Weyl used any combinatorial n-sphere Z^n punctuated it and substituted it for the E^n.

[86] H. Kneser (1926) referred to Weyl, and gave a little later van Kampen the hint that the latter's cell complexes satisfy Weyl's axioms (van Kampen, 1929, p. 3). On the other hand Kneser criticized Weyl's axioms as too complicated, as the consistency and completeness of the axioms were left open (1926, pp. 12f.). The former aspect (consistency) had in fact been discussed by Weyl (1924, pp. 416ff.), while completeness had been marked as a severe problem by the latter (1924, p. 419).

[87] See diverse articles in the recent *Handbook* [Aull and Lowen, 1997] and the announced next volume(s) with several contributions on the history of set theoretic ("general") topology.

induction over dimension n, what might be called a regular cell complex[88] C^n, where essentially each n-cell Z^n in C^n has a combinatorial sphere S^{n-1} as boundary, and the $(n-1)$-subcomplex is also regular. Moreover for all k-cells Z^k in C^n $(1 \leqslant k \leqslant n)$ internal transformations by elementary cell subdivisions (or the inverse operation) are defined.[89] Starting from standard combinatorial n-spheres[90] S_N^n Kneser allowed all those combinatorial schemes of spheres, which can be constructed by internal transformations in his sense. Thus he was proud to achieve a simpler characterization of combinatorial manifolds M^n than Weyl, by the condition that the neighbourhood complex of each 0-cell Z^0 has an S^{n-1} as boundary.[91] But his approach was not only built on the unproven (and unprovable as we know) Hauptvermutung, but did even not allow the proof of Poincaré duality for orientable manifolds by combinatorial means. So it was in the end doubtful whether his approach had a real advantage in comparison with Weyl's, although he had achieved a much simpler framework of postulates.

In the late 1920-s several mathematicians in different international groups, relatively independent from each other, turned towards a more pragmatic approach with respect to combinatorial manifolds. They turned the question upside down[92] and looked for combinatorially accessible conditions that an orientable cell complex satisfies "Poincaré" duality. J.W. Alexander, L.S. Pontrjagin,[93] L.F. Vietoris (1928), and E.R. van Kampen in his Leiden dissertation (1929) chose similar strategies to get rid of the unanswerable question under which conditions a combinatorial complex is a "real" (i.e. topological) manifold. The essential common point of their approaches was the idea of weakening the sphere condition for the boundaries of neighbourhood complexes from combinatorial to purely homological ones. In this sense Vietoris took up Brouwer's constructive definition of a manifold and modified it by a homology criterion for the local simplicity property, defined inductively over dimension. More precisely, he defined an *h-manifold* as a simplicial complex M in which the star of each vertex e_0 is bounded by an "h-sphere". An $(n-1)$-dimensional *h-sphere*, on the other hand, is defined as an orientable h-manifold of dimension $n-1$ with the same Betti numbers as a sphere: $p_0 = p_{n-1} = 1$, $p_i = 0$ for $1 < i < n-1$; with the inductive definition anchored in the obvious stipulation that a 0-dimensional h-sphere is a pair of points (Vietoris, 1928, p. 170). That allowed him to establish Poincaré duality for orientable closed h-manifolds by the construction of dual complexes and the use of Poincaré's argument. In fact, in the introduction of his paper he stated frankly that his proposal of a modified concept of h-manifolds arose from a proof analysis as a result of which he did not try to fill the gap in the original argumentation, but preferred to adapt the conceptual frame to Poincaré's original proof structure.[94]

88 Kneser used the terminology "cell building (Zellgebäude)".

89 More precisely the boundary of each Z^k which is a S^{k-1} is divided into two standard $(k-1)$-cells with common boundary S^{k-2}. Then the substitution of Z^k by two k-cells Z_1^k, Z_2^k, and a $(k-1)$-cell Z^{k-1} which are bounded by the subdivided parts of the S^{k-1} and inherit the boundary relations of the large cell Z^k is an *internal transformation* in the sense of Kneser (1926, p. 8).

90 S_N^n is defined by two cells in each dimension $0, \ldots, n$, each of which is bounded by all cells of less dimension.

91 Kneser claimed that by use of the internal transformations the same holds for "all the other points".

92 From the point of view of the manifold concept one should perhaps say that they turned the question "downside up".

93 Alexander and Pontrjagin's in unpublished notes, as was reported by van der Waerden (1930, p. 125); compare also [Dieudonné, 1989, p. 50].

94 "We shall not fill this gap (of the original proof referring to manifolds, ES) but define a manifold concept for which we can fill it, while the remaining proof of Poincaré can be transferred without change." (Vietoris, 1928,

Lefschetz generalized this approach in terms of relative homology with respect to a subcomplex, thus documenting that the combinatorial strategy to work out the manifold concept was deeply influenced and even transformed by the advent of algebraic topology (Lefschetz, 1920, pp. 119ff.).

Van Kampen followed an approach closer to Veblen's and Weyl's recursive definition of manifolds. During his doctoral research he was in contact with B.L. van der Waerden and informed by the latter about the different strategies for coming to a formally satisfying definition of the concept. Van Kampen choose to add to the basic structure of a Brouwerian simplicial complex the structure of what he called a star-complex, where the concept of *star* and *star complex* had a common recursive definition.[95] Equality of star-complexes STC^n and STC'^n was defined by him as a combinatorial equivalence of the underlying simplicial complexes, which leads to a bijection of the stars. Thus the incidence structure of the stars (of all dimensions) gives complete information about the structure of a star-complex,[96] and allowed him to define a dual star-complex STC^{n*} to a given star-complex STC^n with the same underlying simplicial complex and dualization $k' = n - k$ of the dimensions k, k' of dual stars.[97] Then the incidence matrices of a star-complex and its dual arise from interchanging order of the columns and transposition and behave like Poincaré's incidence matrices in the proof of Poincaré duality.

Van Kampen had thus won a recursively defined normalization of simplicial complexes to which he added a postulate with a dual combinatorial criterion of local simplicity for defining a *combinatorial manifold*: (1) Each k-star is homologically trivial in dimensions $1 < j < k$; and (1′) the same holds for the stars of the dual star-complex (van Kampen, 1929, p. 13). The approach was chosen to derive different duality theorems (Poincaré-, Alexander-, etc.) in a purely combinatorial and thus finite manner. Moreover the combinatorial manifolds satisfy Weyl's axioms, as van Kampen remarked with reference to H. Kneser, but without any discussion of Weyl's original goal to sharpen increasingly the combinatorial postulates until they are coextensive with an axiomatic characterization of continuous manifolds.[98]

The next year B.L. van der Waerden gave a talk at the annual meeting of the *Deutsche Mathematiker-Vereinigung*, in which he presented and discussed the different proposals for the definition of a topological manifold on what he called the "battlefield of different methods" in combinatorial topology (van der Waerden, 1930, p. 121). He counted 5 different possibilities, an axiomatic one (Kneser, 1926), two purely combinatorial ones, of which he presented one as methodologically unsatisfying (Dehn and Heegard, 1907, Tietze, 1908) and the other, homologically oriented one, as more sophisticated (Vietoris, 1928; van Kampen, 1929), and "two" mixed approaches (Poincaré, 1899, 1900; Brouwer, 1911b). Van der

p. 165). Here we have, to put it in Lakatos' terminology, a beautiful case of a completely conscious concept modification generated by proof analysis.

95 A star of dimension 0 is a point; a 0-dimensional star-complex is a finite set of stars. An *n-dimensional star* is a (simplicial) projection of an $(n-1)$-dimensional star-complex from a point (the *centre* of the star). An *n-dimensional star-complex* STC^n is produced from an $(n-1)$-dimensional star-complex STC^{n-1} by adding n-stars, generated by projection of star-subcomplexes of STC^{n-1}, such that each star from the latter is part of the border of at least one of the n-stars (van Kampen, 1929, pp. 6f.).

96 (van Kampen, 1929, Theorem 2a,b).

97 A k-dimensional star a_i^k of STC^n is dualized by collecting all $(n-k)$-dimensional stars b_j^{n-k} which meet the centre of a_i^k, but no other vertex.

98 (van Kampen, 1929, p. 3).

Waerden discussed the relative merits and disadvantages of all these approaches.[99] The phase of open exploration for the topological manifold concept had more or less come to a conclusion; the axiomatic characterization and a constructive (purely combinatorial) one were the outcome of differing methodological approaches, Brouwer's "mixed" approach gave the most promising bridge, and Weyl's original intentions were close to forgotten.[100]

5.4. *Finally the "modern" axiomatic concept*

There was, of course, still another line of research, more closely linked to differential geometry, where manifolds played an essential role, and purely topological aspects (independently of whether continuous, combinatorial, or homological ones) did not suffice and still needed elaboration. In North America Oswald Veblen and his students formed an active center in both fields of topology and modern geometry. Veblen and his student J.H.C. Whitehead, coming from (and going back to) Oxford, brought the axiomatization of the manifold concept to a stage which stood up to the standards of modern mathematics in the sense of the 20th century (Veblen and Whitehead, 1931, 1932). Veblen was an admirer of the Göttingen tradition of mathematics, in particular, F. Klein and D. Hilbert, and cooperated closely with H. Weyl, the broadest representative of his own generation from the Klein and Hilbert tradition. Veblen and J.H.C. Whitehead combined a view of the central importance of structure groups for geometry (generalizing the Erlanger program) with Hilbert's embryonic characterization of manifolds by coordinate systems; and they took care that the topologization of the underlying set would satisfy Hausdorff's axioms for a topological space.

They characterized the *structure* of a manifold by the specification of a regular groupoid G ("pseudogroup") of transformations of open sets ("regions") in $\mathbb{R}^n$, allowing as main examples C^i-transformations of open sets ($i = 0, \ldots, \infty$, or $i = \omega$). The n-dimensional manifold of structure G in the sense of Veblen and Whitehead consists in a set M and a system of *admissible coordinate systems* $\varphi : U \to V$ with bijective maps φ onto regions $V \subset \mathbb{R}^n$, defined for $U \in \mathcal{U} \subset \mathcal{P}(M)$, such that three groups of axioms hold:

(A) *Basic axioms for admissible coordinate systems*:
Changes of coordinates are given by maps from the structure groupoid G and each coordinate map may be changed by a transformation from G (axioms A_1, A_2). Moreover, to each coordinate map $\varphi : U \to V$ a restriction to $U' \subset V$ such that $\varphi(U') = V'$ is an n-cell V' in $\mathbb{R}^n$ is also an admissible coordinate system (A_3). U' is called an n-cell in the manifold.

(B) *Union of compatible coordinate systems*:
If for a collection of admissible coordinate systems $\varphi : U_i \to V_i$ $(i \in I)$, with n-cells as coordinate images V_i, the coordinate maps coincide on overlaps ($U_i \cap U_j \neq \emptyset$), then the "union" of coordinate systems defined in the obvious way, $\varphi : \bigcup_i U_i \to \bigcup_i V_i$, is

[99] The references between the 5 approaches to authors were not all made explicit by van der Waerden but presented on a purely methodological level.

[100] Weyl's contribution appeared in van der Waerden's bibliography, but was not discussed by him. He thus indirectly took part in the methodological "battle" of combinatorial topology, although he probably did not realize it.

also admissible (axiom B_1). Each admissible coordinate system can be represented as such a union (B_2).[101]

(C) *Topological axioms*:
For intersecting n-cells U, U' in M with $p \in U \cap U'$ there is an n-cell $U'' \subset U \cap U'$ containing p (axiom C_1). For each two different points $p, q \in M$ there exist nonintersecting coordinate neighbourhoods U_p, U_q of p and q, respectively (C_2). Finally, M contains at least two different points (C_3).

Taking n-cells in M, containing p, as neighbourhoods of p the axioms of Veblen and Whitehead give a structure of a Hausdorff space on M (without second countability axiom) (Veblen and Whitehead, 1931, p. 95; 1932, p. 79).

Whitehead and Veblen presented their axiomatic characterization of manifolds of class G first in a research article in the *Annals of Mathematics* (Veblen and Whitehead, 1931) and in the final form in their tract on the *Foundations of Differential Geometry* (Veblen and Whitehead, 1932). Their book contributed effectively to a conceptual standardization of modern differential geometry, including not only the basic concepts of continuous and differentiable manifolds of different classes, but also the "modern" reconstruction of the *differentials* $\mathrm{d}x = (\mathrm{d}x_1, \ldots, \mathrm{d}x_n)$ as objects in tangent spaces to M.[102] Basic concepts like Riemannian metric, affine connection, holonomy group, covering manifolds, etc. followed in a formal and symbolic precision that even from the strict logical standards of the 1930-s there remained no doubt about the wellfoundedness of differential geometry in manifolds. Moreover they made the whole subject conceptually accessible to anybody acquainted with the language and symbolic practices of modern mathematics.

5.5. *And first successes in unification*

The clear definition and mutual delimitation of continuous, differentiable and analytic structure of manifolds by Whitehead and Veblen improved the framework for a more detailed study of the basic questions of triangulation, Hauptvermutung and thus the questions which were at stake with the competing strategies of a genetic/constructive characterization of manifolds versus an axiomatic one. They had been posed at first for topological manifolds, but could as well be fruitfully transferred to the differentiable case.

Already at the turn of the thirties, i.e. before the Veblen and Whitehead axioms had been formulated, first positive results on the connection between the two large strategies had been achieved. In 1925 T. Radó had shown that two-dimensional manifolds can be triangulated and thus that in this respect Tietze had been right. During the following decade the higher dimensional case could only be dealt with under structurally specifying conditions. Several authors contributed to the proof that a real analytical manifold admits triangulation: Van der Waerden (1929) clarified the triangulability for algebraic manifolds, Lefschetz (1920, Chapter VIII) sketched the outline for a general proof in the case of a general analytic manifolds, and Koopman and Brown (1932) elaborated a complete proof. Only a few years later S.S. Cairns, a former student of M. Morse, proved the existence of triangulations for n-dimensional differentiable manifolds (Cairns, 1934). In 1940 J.H.C. Whitehead considered and showed the existence of differentiable triangulations of a C^1-manifold of

[101] Neither here nor elsewhere did Veblen and Whitehead postulate a countability restriction for the coordinate neighbourhoods.

[102] They still used the pre-Bourbakian terminology of "contravariant" vector for the objects in the tangent space.

any dimension and proved that in this structure even the (differentiable) Hauptvermutung is true.[103] Thus the combinatorial and the axiomatic approach had turned out by 1940 to be complementary aspects of completely coextensive characterizations for *differentiable* manifolds.

So far Poincaré's intuition had been vindicated and put on a solid logical basis, although the elaboration of "purely" (continuity) topological structure had shifted the question to a deeper conceptual level than Poincaré ever would have considered.[104] And even the topological case seemed at first rather promising, at least in low dimensions. Radó's success for dimension 2 was extended in the early 1950-s, when E.E. Moise proved that each 3-dimensional continuous manifold admits a triangulation (Moise, 1952). At the turn of the 1950-s one might thus have hoped that the different specifications of the manifold concept had led to difficult and challenging technical problems for modern mathematics, but that they could perhaps be solved positively by increasingly sophisticated methods and an interplay between the different structural level and methods. Why should they not lead to a unified frame for the topology and geometry of manifolds in a rather straightforward manner?

6. Outlook on more recent developments

6.1. *Growing diversity...*

Even the conceptual unity one might have hoped for in the early 1950-s was, however, not at all a narrow one. Already Riemann had indicated the possibility of investigating manifolds from different methodological views and had considered this differentiation as an important feature for adapting the general concept to diverse scientific contexts. Such a differentiation had developed on a technically much more refined level during the first half of the 20th century in a broader range. Besides the distinction between the combinatorial or PL- and axiomatic approaches to the topological manifold concept and its differentiation according to smoothness levels (C^i, $0 \leqslant i \leqslant \infty$ or $i = \omega$), other contexts had given reasons for developing the concepts of a complex analytic manifold and of algebraic birational variety. These, as well as the diverse differential geometric structural specifications on differentiable manifolds, would have to be considered for a broader picture of the growing diversity of manifolds in our century, but remain outside the range of this article.

To keep closer to the core of our subject, we have to face the surprising diversity in the topological and differential structures on manifolds of dimension $n \geqslant 4$, which became apparent by and by starting in the late 1950-s. After J. Milnor detected nonstandard differentiable structures on the 7-sphere (Milnor, 1956), an increasing number of unexpected insights into the differentiable structure of higher dimensional manifolds came to the fore. Among them were E. Brieskorn's and others' study of exotic spheres, which arose relatively "naturally" in investigations of singularities of algebraic geometry, and in the 1960-s M.H. Freedman's and S.K. Donaldson's broad investigations of differentiable structures on 4-manifolds. During the 1980-s the tremendous range of effects, from a number of unexpected differentiable structures on supposedly well known manifolds, like higher dimensional spheres and the $\mathbb{R}^4$, to the fact that certain topological 4-manifolds do not admit

[103] (Whitehead, 1940, Theorem 8, p. 822).

[104] Poincaré considered his manifolds always as differentiable, in times even as analytic, which he defined by an approximation argument of analytic maps by differentiable ones.

a differentiable structure at all, became known. They would have given sufficient reason for a Poincaré to deplore again, and now on another much more sophisticated level, the turn of mathematics towards an "artificiality which alienates the whole world", as he had proclaimed in his talk to the second International Congress of Mathematicians with respect to the rise of modern mathematics (Poincaré, 1902c). He was particularly struck by the results of the logical analysis of continuous nondifferentiable functions.[105] The preparation of such unexpected symbolical phenomena was nevertheless an important part of the achievements of the high phase of modern mathematics and characteristic for its spirit. They are discussed more in detail and with much more expertise in other contributions to this volume.

Similar evaluations might be drawn on the fate of the triangulation problem and the Hauptvermutung for topological manifolds of dimension $n \geqslant 5$. J. Milnor's early example of manifolds in dimension 8, with different combinatorial structures (Milnor, 1961) paid tribute to but finished the hope for a too simply conceived positive end of the program outlined in the first third of the century. The work by R.C. Kirby and L.C. Siebenmann (1969) with the characterization of exact obstruction criteria, given by cohomology classes of the manifold in question, allowed their successors to determine manifolds for all dimensions $n \geqslant 5$, in which the Hauptvermutung does not hold, and to characterize topological manifolds without any PL-structure.

Thus from the late 1960-s onward the hope for a conceptually unified framework for all of modern mathematics has undergone a deep transformation,[106] forced upon the research community by the growing complexity of material, methods and results.[107] There is a growing and perhaps ever increasing trend towards diversification, and differentiation even, to a certain degree, between different subbranches or aspects in the mathematics of such a relatively well delimited field as the topology and differential topology of manifolds.

6.2. ... *but still a unifying perspective on mathematical practice by overarching concepts*

Some observers even tend to see a loss of connection between different branches of mathematics as a whole and identify such a loss of unity, growing pluralism of methods, structures, and approaches from a specific cultural perspective as a "postmodernist" dynamics of mathematics, which has speeded up from the late 1960-s onward. No doubt, modern mathematics, and maybe with it, modern culture has reached a mature, probably even "late" stage, at least in comparison with its expansionary "high" phase from the late 19th to the middle of the 20th century.[108] But history has always been an open process, and Riemann's and other persons' vision of the cognitive strength and productivity of conceptual unification has neither lost its fruitfulness nor its convincing power.

The vision of a strictly unified and structurally predetermined symbolical universe of mathematics, which seems to have been the dream of many of the protagonists of the high

[105] The famous citation of the monster functions which he abhorred is in (Poincaré, 1908); H. Mehrtens describes this as Poincaré's "antimodernist" view of mathematics [Mehrtens, 1990, Chapter 3.3].

[106] First glances of such an ongoing shift could probably be seen already in the late 1950-s by very sensitive observers.

[107] Compare also [Corry, 1996] with respect to the fate of the structural "image" in recent algebra.

[108] I. James has called this phase of modernity as the "classical" one in his Nice talk.

phase of mathematical modernity,[109] has become obsolete through the progress of mathematical work itself. But we are not at all obliged to understand mathematical concepts in a rigid, mainly technical sense; we may also conceive them as organizing centers of cognition, which act in a "dialectical" interplay between their role of cognitive orientation and the symbolical and technical specification they impart on the practice of mathematics. There is no good argument to reduce, or even to proclaim the end for, the unifying role of concepts in today's and future mathematics and human knowledge more general. We could just as well draw the opposite conclusion and insist on an increasing importance of their unificatory role as a counterbalance to cultural and cognitive diversification. Thus the history of the manifold concept may be taken as paradigmatic for a symbolical world of increasing diversity and richness in which we live, work, and orient ourselves.

Bibliography

Historical literature

Aull, C.E. and Lowen, R. (1997), *Handbook of the History of General Topology*, Vol. 1, Kluwer, Dordrecht.

Bollinger, M. (1972), *Geschichtliche Entwicklung des Homologiebegriffs*, Archive for History of Exact Sciences **9**, 94–170.

Bottazzini, U. (1977), *Riemanns Einfluss auf Betti und Casorati*, Archive for the History of Exact Sciences **18**, 27–37.

Bottazzini, U. (1985), *Dinis Arbeiten auf dem Gebiet der Analysis*, Mathemata. Festschrift für Helmuth Gericke, M. Folkerts and U. Lindgren, eds, Franz Steiner Verlag, Stuttgart, 591–605.

Bottazzini, U. (1986), *The Higher Calculus: A History of Real and Complex Analysis from Euler to Weierstrass*, Springer, Berlin/Heidelberg/New York/Tokyo.

Coleman, R.A. and Korte, H. (1998), *The scientific achievements of Hermann Weyl*, Manuscript, to appear in Hermann Weyl's Space–Time–Matter ..., Proceedings DMV-Seminar 1992, Birkhäuser, Basel.

Corry, L. (1996), *Modern Algebra and the Rise of Mathematical Structures*, Birkhäuser, Basel.

Dahan, A. (1997), *Le difficile héritage de Henri Poincaré en systèmes dynamiques*, Henri Poincaré. Science et Philosophie ..., J.-L. Greffe, G. Heinzmann and K. Lorenz, eds, Congrèss International Nancy, France 1994, Akademie Verlag, Berlin; Blanchard, Paris, 13–34.

Dieudonné, J. (1974), *Cours de Géomètrie Algébrique*, Vol. 1, Presses Universitaires de France, Paris.

Dieudonné, J. (1989), *A History of Algebraic and Differential Topology, 1900–1960*, Birkhäuser, Basel.

Dieudonné, J. (1994), *Une brève histoire de la topologie*, Development of Mathematics, 1900–1950, J.-P. Pier, ed., Birkhäuser, Basel, 35–156.

Epple, M. (1997), *Topology, space, and matter. Topological notions in 19-th century natural philosophy*, Manuscript, Mainz, to appear in Archive for History of Exact Sciences.

Farwell, R. and Knee, C. (1990), *The missing link: Riemann's "Commentatio", differential geometry and topology*, Historia Mathematica **17**, 223–255.

Feferman, S. (1988), *Weyl vindicated: "Das Kontinuum" 70 years later*, Atti del Congresso Temi e Prospettive della Logica e dela Filosofia della Scienza Contemporanee, Vol. 1, Cesena 7–10 gennaio 1987, CLUEB, Bologna, 59–93.

Ferreirós, J. (1993), *El Nacimiento de la Teoría de Conjuntos en Alemania, 1854–1908*, Publicaciones de la Universidad Autónoma, Madrid.

Ferreirós, J. (1996), *Traditional logic and the early history of sets, 1854–1908*, Archive for History of Exact Sciences **50**, 5–71.

Ferreirós, J. (1999), *Labyrinth of Reason. A History of Set Theory and its Role in Modern Mathematics*, Birkhäuser, Basel.

[109] Such protagonists were, among others, D. Hilbert and N. Bourbaki. Deviating visions existed, like Hausdorff's or, from a completely different philosophical background, Brouwer's and Weyl's, but were of a much more restricted influence in the scientific world during the phase of "high" modernity.

Gilain, C. (1991), *La théorie qualitative de Poincaré et le problème de l'intégration des équation différentielles*, Cahiers d'Histoire de Philosophie des Sciences **34**, 215–242.

Gilain, C. (o.D), *La théorie géométrique des équations différentielles de Poincaré et l'histoire de l'analyse*, Thèse de Doctorat 3ème cycle, Universitè Paris I.

Gray, J.J. (1979), *Ideas of Space. Euclidean, Non-Euclidean and Relativistic*, Clarendon, Oxford; 2nd ed. 1989.

Gray, J.J. (1984), *Fuchs and the theory of differential equations*, Bull. of Amer. Math. Soc. **10**, 1–26.

Gray, J.J. (1986), *Linear Differential Equations and Group Theory from Riemann to Poincaré*, Birkhäuser, Basel.

Gray, J.J. (1992), *Poincaré, topological dynamics, and the stability of the solar system*, An Investigation of Difficult Things. Essays on Newton and the History of Exact Sciences, P.M. Harman and A.E. Shapiro, eds, 503–524.

Greffe, J.-L., Heinzmann, G. and Lorenz, K. (1997), *Henri Poincaré Science et Philosophie ...*, Congrès International Nancy, France, 1994, Akademie-Verlag, Berlin; Blanchard, Paris.

Herrmann, A. (1996), *Eléments d'histoire sémiotique de l'homologie*, Thèse de Doctorat, Université Paris VII.

Johnson, D. (1979/1981), *The problem of the invariance of dimension in the growth of modern topology, I, II*, Archive for History of Exact Sciences **20** (1979), 97–188; **25** (1981), 85–167.

Johnson, D. (1987), *L.E.J. Brouwer's coming of age as a topologist*, Studies in the History of Mathematics, E.R. Phillips, ed., 61–97.

Klein, F. (1926/1927), *Vorlesungen über die Entwicklung der Mathematik im 19. Jahrhundert*, 2 Bde: **1** (1926), **2** (1927), Springer, Berlin. Nachdruck: Chelsea, New York, 1967; Springer, Berlin, 1979.

Koetsier, T. and van Mill, J. (1997), *General topology, in particular dimension theory, in the Netherlands: The decisive influence of Brouwer's intuitionism*, Handbook of the History of General Topology, Vol. 1, C.E. Aull and R. Lowen, eds, Kluwer, Dordrecht, 135–180.

Kolmogorov, A.N. and Yushkevitch, A.P. (eds) (1996), *Mathematics of the 19th Century. Geometry, Analytic Function Theory*, Birkhäuser, Basel.

Kuiper, N. (1979), *A short history of triangulation and related matters*, Mathematical Centre Tracts **100**, 61–79.

Lakatos, I. (1976), *Proofs and Refutations*, The Logic of Mathematical Discovery, Cambridge Univ. Press, London.

Laugwitz, D. (1996), *Bernhard Riemann 1826–1866. Wendepunkte in der Auffassung der Mathematik*, Birkhäuser, Basel.

Lützen, J. (1988), *The geometrization of analytical mechanics. A pioneering contribution by* J. Liouville (ca. 1850), Preprint 28, Københavns Universitet Matematisk Institut.

Lützen, J. (1990), *Joseph Liouville, 1809–1882. Master of Pure and Applied Mathematics*, Studies in the History of Mathematics and Physical Sciences, Springer, New York.

Lützen, J. (1995), *Interactions between mechanics and differential geometry in the 19th century*, Archive for History of Exact Sciences **49**, 1–72.

Mehrtens, H. (1990), *Mathematik – Modern – Sprache. Die mathematische Moderne und ihre Gegner*, Suhrkamp, Frankfurt/Main.

Moore, G. (1978), *The origins of Zermelo's axiomatization of set theory*, Journal of Philosophical Logic **7**, 307–329.

Moore, G. (1982), *Zermelo's Axiom of Choice. Its Origins, Development, and Influence*, Studies in the History of Mathematics and Physical Sciences, Vol. 8, Springer, New York.

Pont, J.-C. (1974), *La Topologie Algébrique des Origines à Poincaré*, Presses Universitaires de France, Paris.

Reich, K. (1994), *Die Entwicklung des Tensorkalküls. Vom Absoluten Differentialkalkül zur Relativitätstheorie*, Birkhäuser, Basel.

Scholz, E. (1980), *Geschichte des Mannigfaltigkeitsbegriffs von Riemann bis Poincaré*, Birkhäuser, Basel/Boston/Stuttgart.

Scholz, E. (1982a), *Herbart's influence on Bernhard Riemann*, Historia Mathematica **9**, 423–440.

Scholz, E. (1982b), *Riemanns frühe Notizen zum Mannigfaltigkeitsbegriff und zu den Grundlagen der Geometrie*, Archive for History of Exact Science **27**, 213–282.

Scholz, E. (1995), *Hermann Weyl's purely "infinitesimal geometry"*, Proceedings International Congress of Mathematicians, Zürich 1994, Birkhäuser, Basel, 1592–1603.

Scholz, E. (1996), *Logische Ordnungen im Chaos: Hausdorffs frühe Beiträge zur Mengenlehre*, Felix Hausdorff zum Gedächtnis, Vol. 1, Aspekte seines Werkes, E. Brieskorn, ed., Vieweg, Wiesbaden, 107–134.

Scholz, E. (1998), *Hermann Weyls Differentialgeometrie in Kritik und Fortsetzung der Riemannschen Tradition*, Manuskript, Wuppertal, to appear in: Hermann Weyl's Space–Time–Matter ..., Proceedings DMV-Seminar 1992, Birkhäuser, Basel.

Tòth, I. (1972), *Die nichteuklidische Geometrie in der Phänomenologie des Geistes*, Wissenschaftstheoretische Betrachtungen zur Entwicklungsgeschichte der Mathematik, Horst Heiderhoff Verlag, Frankfurt.

Tòth, I. (1980), *Wann und von wem wurde die nichteuklidische Geometrie begründet*? Archive Internationales d'Histoires des Sciences **30**, 192–205.

Vanden Eynde, R. (1992), *Historical evolution of the concept of homotopic paths*, Archive for History of Exact Sciences **29**, 127–188.

Volkert, K. (1994), *Das Homöomorphieproblem insbesondere der 3-Mannigfaltigkeiten in der Topologie 1892–1935*, Habilitationsschrift, Heidelberg.

Volkert, K. (1997), *The early history of the Poincaré conjecture*, Henri Poincaré. Science et Philosophie . . . , J.-L. Greffe, G. Heinzmann and K. Lorenz, eds, Congrèss International Nancy, France 1994, Akademie-Verlag, Berlin; Blanchard, Paris, 241–250.

Sources

Beltrami, E. (1868a), *Saggio di interpretazione della geometria non-euclidea*, Giornale di Matematiche **6**, 284–312. Opere Matematiche, Vol. 1, Milano, 1902, 262–280.

Beltrami, E. (1868b), *Teoria fundamentale degli spazii di curvatura costante*, Annali di Matematica (2) **2**, 232–255. Opere Matematiche, Vol. 1, Milano, 1902, 262–280.

Betti, E. (1871), *Sopra gli spazi di un numero qualunque di dimensioni*, Annali di Matematica (2) **4**, 140–158. Opere Mathematiche, Vol. 2, Milano, 1913, 273–290.

Betti, E. (1913), *Opere Matematiche*, Vol. 2, Milano.

Betti, E. (1915), *Correspondence with Placido Tardy*, Atti Accademia dei Lincei (5) **24**, 517–519.

Brouwer, L.E.J. (1911a), *Beweis der Invarianz der Dimensionszahl*, Mathematische Annalen **69**, 169–175.

Brouwer, L.E.J. (1911b), *Über Abbildung von Mannigfaltigkeiten*, Mathematische Annalen **71**, 97–115.

Brouwer, L.E.J. (1976), *Collected Works*, Vol. 2, Amsterdam.

Brown, A.R. and Koopman, B.O. (1932), *On the covering of analytic loci by complexes*, Transactions AMS **34**, 231–251.

Cairns, S. (1934), *On the triangulation of regular loci*, Annals of Mathematics **35**, 379–587.

Cauchy, A. (1847), *Mémoire sur les lieux analytiques*, Comptes Rendus **24**, 885. Œuvres (1) **10** (1891), 292–295.

Cayley, A. (1843), *Chapters in the analytical geometry of n dimensions*, Cambridge Mathematical Journal **4**, 119–127. Collected Mathematical Papers **1**, 317–326.

Clebsch, A. (1864), *Über die Anwendung der Abelschen Funktionen in der Geometrie*, Journal für Mathematik **63**, 189–243.

Dehn, M. and Heegard, P. (1907), *Analysis Situs*, Enzyklopädie der Mathematischen Wissenschaften, Teil III, Bd. 1.1, 153–220.

von Dyck, W. (1888), *Beiträge zur Analysis Situs I*, Mathematische Annalen **32**, 457–512.

von Dyck, W. (1890), *Beiträge zur Analysis Situs II*, Mathematische Annalen **37**, 273–316.

Gauß, C.F. (1851/1917), *Bestimmung des kleinsten Werts der Summe* $x_1^2+x^2+\cdots+x_n^2 = R^2$ *für m gegebene Ungleichungen* $u \geqslant 0$. Aus der Vorlesung: Über die Methode der kleinsten Quadrate, WS 1850/81, Mitschrift A. Ritter. Werke **10.1**, Leipzig 1917, 473–481. Reprint Hildesheim etc. 1973.

Grassmann, H.G. (1844), *Die lineale Ausdehnungslehre, ein neuer Zweig der Mathematik, dargestellt und durch neue Anwendungen auf die übrigen Zweige der Mathematik, die Lehre vom Magnetismus und die Krystallonomie erläutert*, Leipzig. Werke **1.1**, Leipzig 1894.

Hausdorff, F. (1914), *Grundzüge der Mengenlehre*, Veit, Leipzig. Reprint: Chelsea, New York, 1949, 1965, 1978.

von Helmholtz, H. (1868), *Über die Tatsachen, die der Geometrie zu Grunde liegen*, Göttinger Nachrichten 1868. Wissenschaftliche Abhandlungen **2**, Leipzig 1883, 618–639.

Hilbert, D. (1902a), *Über die Grundlagen der Geometrie*, Nachrichten Gesellschaft der Wissenschaften Göttingen, 233–241. Also in Anhang IV of Grundlagen der Geometrie, 2nd ed., 1903, 121ff.; 7th ed. 1930, 178–230.

Hilbert, D. (1902b), *Über die Grundlagen der Geometrie*, Mathematische Annalen **56**, 381–422. [Abbreviated version of (1902a).]

Hopf, H. (1926), *Vektorfelder in n-dimensionalen Mannigfaltigkeiten*, Mathematische Annalen **96**, 225–250.

Jacobi, C.G.J. (1834), *De binis quibuslibet functionibus homogeneis secundi ordinis per substitutiones lineares in alias binas transformandis quae solis quadratis variabilium constant; . . .* Journal für die Reine und Angewandte Mathematik **12**, 1–69. Gesammelte Werke **3**, Berlin 1884, 191–268.

Jordan, C. (1866), *Sur la déformation des surfaces*, Journal de Mathématique (2) **11**, 105–109. *Œuvres* 4, 85–90.
Jordan, C. (1893), *Cours d'Analyse*, 2nd ed., Paris.
van Kampen, E.R. (1929), *Die kombinatorische Topologie und die Dualtitätssätze*, Dissertation, Leiden, Publiziert: Den Haag: van Stockum.
Kirby, R.C. and Siebenman, L.C. (1969), *On the triangulation of manifolds and the Hauptvermutung*, Bull. Amer. Math. Soc. **75**, 742–749.
Klein, F. (1871), *Über die sogenannte Nicht-Euklidische Geometrie*, Mathematische Annalen **4**, 573–625. [GMA 1, 244–253].
Klein, F. (1872), *Vergleichende Betrachtungen über neuere geometrische Forschungen. Erlangen*, Mathematische Annalen **43** (1893). [GMA 1 (1921), 460–497].
Klein, F. (1873), *Über die Flächen dritter Ordnung*, Mathematische Annalen **6**. [GMA 2, 11–44].
Klein, F. (1874–1876), *Bemerkungen über den Zusammenhang der Flächen* (zwei Aufsätze aus den Jahren 1874 und 1875/76), Mathematische Annalen **7, 9**. [GMA 2, 63–77].
Klein, F. (1921–1923), *Gesammelte Mathematische Abhandlungen*, 3 Bde, Springer, Berlin. Reprint: Springer, Berlin, 1973.
Kneser, H. (1926), *Die Topologie der Mannigfaltigkeiten*, Jahresbericht DMV **34**, 1–14.
Kreck, M. (1998), *(Co-)Homology via Topological Varieties, following Riemann and Poincaré I*, Preprint University Mainz.
Kronecker, L. (1869), *Ueber Systeme von Functionen mehrer Variablen*, Monatsberichte Berliner Akademie der Wissenschaften, 159–193. Werke **1**, 175–212.
Lagrange, J.-L. (1788), *Mécanique Analitique*, 2 vols, Paris. *Œuvres* 11, 12, Paris 1888.
Lagrange, J.-L. (1797), *Théorie des Fonctions Analytiques*, Paris; 2nd ed. 1813.
Lefschetz, S. (1920), *Topology*, American Mathematical Society, New York.
Lipschitz, R. (1872), *Untersuchung eines Problems der Variationsrechnung, in welchem das Problem der Mechanik enthalten ist*, Journal für Reine und Angewandte Mathematik **74**, 116–149.
Milnor, J. (1956), *On manifolds homeomorphic to the 7-sphere*, Annals of Mathematics (2) **64**, 399–405.
Milnor, J. (1961), *Two complexes which are homeomorphic but combinatorially distinct*, Annals of Mathematics (2) **74**, 575–590.
Möbius, A.F. (1863), *Theorie der elementaren Verwandtschaften*, Abhandlungen Sächsische Gesellschaft der Wissenschaften 15. [GW 2, 433–471].
Möbius, A.F. (1865), *Über die Bestimmung des Inhalts eines Polyeders*, Abhandlungen Sächsische Gesellschaft der Wissenschaften **17**. [GW 2, 473–512].
Möbius, A.F. (1886), *Gesammelte Werke*, Bd. **2**, Leipzig.
Moise, E.E. (1952), *Affine structures on 3-manifolds*, Annals of Mathematics (2) **56**, 96–114.
Neumann, C. (1865), *Vorlesungen über Riemanns Theorie der Abelschen Integrale*, Leipzig; 2nd ed. 1884.
Noether, M. (1870), *Zur Theorie des eindeutigen Entsprechens algebraischer Gebilde von beliebig vielen Dimensionen*, Mathematische Annalen **2**, 293–321.
Noether, M. (1875), *Zur Theorie den eindeutigen Entsprechens algebraischer Gebilde, 2. Aufsatz*, Mathematische Annalen **8**, 495–533.
Peano, G. and Genocchi (1884), *Calcolo Differenziale*, Torino.
Picard, E. (1885), *Sur les intégrales de différentielles totales algébriques de première espèce*, Journal de Mathématique (4) **1**, 281–346.
Picard, E. (1886), *Sur les intégrales des différentielles totales de séconde espèce*, Journal de Mathématique (2) **2**, 329–346.
Plücker, J. (1846), *System der Geometrie des Raumes in Neuer Analytischer Behandlungsweise, Insbesondere die Flächen Zweiter Ordnung und Klasse Enthaltend*, Düsseldorf; 2nd ed. 1852.
Poincaré, H. (1884), *Sur les groupes des équations linéaires*, Acta Mathematica **4**, 201–311. [*Œuvres* 2, 300–401].
Poincaré, H. (1885), *Sur les courbes définies par les équations différentielles*, Journal de Mathématique (4) **1**, 167–244. [*Œuvres* 1, 90–161].
Poincaré, H. (1886), *Sur les courbes définies par les équations différentielles*, Journal de Mathématique (4) **2**, 151–217. [*Œuvres* 1, 167–221].
Poincaré, H. (1887), *Sur les résidus des intégrales doubles*, Acta Mathematica **9**, 321–380. [*Œuvres* 3, 493–539].
Poincaré, H. (1895), *Analysis situs*, Journal Ecole Polytechnique **1**, 1–121. [*Œuvres* 6, 193–288].
Poincaré, H. (1899), *Complément à l'Analysis situs*, Rendiconti del Circolo Matematico Palermo **13**, 285–343. [*Œuvres* 6, 290–337].

Poincaré, H. (1900), *Second complément à l'Analysis situs*, Proceedings London Mathematical Society **32**, 277–388. [*Œuvres* 6, 338–370].

Poincaré, H. (1902a), *Sur certaines surfaces algébriques, troisième complèment à l'Analysis situs*, Bulletin Société Mathématique de France **30**. [*Œuvres* 6, 373–392].

Poincaré, H. (1902b), *Sur les cycles des surfaces algébriques, quatrième complèment à l'Analysis situs*, Journal de Mathématique **8**. [*Œuvres* 6, 397–434].

Poincaré, H. (1902c), *Du rôle de l'intuition et de la logique en mathématiques*, Compte Rendu du Deuxième Congrès International de Mathématique, Gauthier-Villars, Paris, 115–130. Reprinted as first chapter in: *La Valeur de la Science*, Flammarion, Paris, 1905.

Poincaré, H. (1904), *Cinquième complément à l'Analysis situs*, Rendiconti del Circolo Matematico Palermo **18**, 45–110. [*Œuvres* 6, 435–498].

Poincaré, H. (1908), *Science et Méthode*, Flammarion, Paris.

Poincaré, H. (1928–1956), *Œuvres*, 11 Vols, Paris.

Radó, T. (1925), *Über den Begriff der Riemannschen Fläche*, Acta Literarum Scientarum Universitas Szeged **2**, 101–121.

Riemann, B. (1851), *Grundlagen für eine allgemeine Theorie der Functionen einer veränderlichen complexen Grösse*, Inauguraldissertation, Göttingen. [GMW 3–45].

Riemann, B. (1854), *Über die Hypothesen, welche der Geometrie zu Grunde liegen, Habilitationsvortrag*, Göttingen, Göttinger Abhandlungen **13** (1867). [GMW 272–287].

Riemann, B. (1857), *Theorie der Abel'schen Functionen*, Journal für Mathematik **54**. [GMW 86–144].

Riemann, B. (1857/1876), *Zwei allgemeine Lehrsätze über lineare Differentialgleichungen mit algebraischen Koeffizienten* [Nachlaß, datiert 30. 2. 1857]. [GMW 379–390].

Riemann, B. (1861/1876), *Commentatio mathematica, qua respondere tentatur quaestioni ab illustrissima Academia Parisiensi propositae: "Trouver quel doit être l'état calorifique d'un corps solide homogène indéfini ..."* [GMW 391–404].

Riemann, B. (1876a), *Fragment aus der Analysis Situs*, H. Weber, ed., Gesammelte Mathematische Werke und Wissenschaftlicher Nachlaß, B. Riemann, Leipzig, 1876, 479–482.

Riemann, B. (1876b), *Gesammelte Mathematische Werke und Wissenschaftlicher Nachlaß*, Leipzig; 2nd ed. 1892. Reprinted: Dover, New York, 1995; Sändig, Nendeln, 1978; R. Narasimhan, ed., Springer, Berlin, 1990.

Schläfli, L. (1851/1901), *Theorie der vielfachen Kontinuität*, Neue Denkschrift d. Allgemeinen Schweizer Gesellschaft Naturwissenschaften **38**. [GMA 1, 167–387].

Schläfli, L. (1872), *Quand'é che dalla superficie generale di terzo ordine si stacca una parte che non sia realmente segata da ogni piano reale*? Annali di Matematica (2) **5**. [GMA 3, 229–237].

Schläfli, L. (1902/1909), *Gesammelte Mathematische Abhandlungen*, 2 Bde, Berlin.

Steinitz, E. (1908), *Beiträge zur Analysis situs*, Sitzungsberichte Berliner Mathematische Gesellschaft **7**, 28–49.

Tietze, H. (1908), *Über die topologischen Invarianten mehrdimensionaler Mannigfaltigkeiten*, Monatshefte für Mathematik und Physik **19**, 1–118.

Tonelli (1875), *Osservazioni sulla teoria della connessione*, Atti della Reale Accademia d. Lincei (2) **2**, 594–601.

Veblen, O. and Whitehead, J.H.C. (1931), *A set of axioms for differential geometry*, Proceedings National Academy of Sciences **17**, 551–561. Whitehead Mathematical Works **1**, 93–104.

Veblen, O. and Whitehead, J.H.C. (1932), *The Foundations of Differential Geometry*, Cambridge University Press, London.

Veblen, O. (1922), *Analysis Situs*, American Mathematical Society, New York; 2nd ed. 1931.

Vietoris, L. (1928), *Über die Symmetrie in den Zusammenhangszahlen kombinatorischer Mannigfaltigkeiten*, Monatshefte für Mathematik und Physik **35**, 165–174.

van der Waerden, B.L. (1929), *Topologische Begründung des Kalküls der abzählenden Geometrie, Anhang I*, Mathematische Annalen **102**, 360–361.

van der Waerden, B.L. (1930), *Kombinatorische Topologie*, Jahresbericht DMV **39**, 121–139.

Weyl, H. (1913), *Die Idee der Riemannschen Fläche*, Teubner, Leipzig/Berlin; 2nd ed. 1923. Also: Chelsea, New York; 3rd ed., 1955.

Weyl, H. (1918), *Das Kontinuum, Kritische Untersuchungen über die Grundlagen der Analysis*, Leipzig. Reprint: New York, 1960.

Weyl, H. (1921), *Über die neuere Grundlagenkrise der Mathematik*, Math. Zeitschrift **10**, 39–79. Selecta, 211–247 (Nachtrag Juni 1955, pp. 247f.). [GA **2**, 143–180].

Weyl, H. (1923), *Analysis situs combinatorio*, Revista Matematica Hispano–Americana **5**, 43. [GA **2**, 390–415].

Weyl, H. (1924), *Analysis situs combinatorio (continuaciòn)*, Revista Matematica Hispano–Americana **6**, 1–9, 33–41. [GA **2**, 416–432].

Weyl, H. (1925/1988), *Riemanns Geometrische Ideen, ihre Auswirkungen und ihre Verknüpfung mit der Gruppentheorie*, K. Chandrasekharan, ed., Springer, Berlin.

Whitehead, J.H.C. (1940), *On C^1-complexes*, Annals of Mathematics **41**, 809–824. Mathematical Works **2**, Oxford, 1962, 207–222.

Zermelo, E. (1904), *Beweis, daß jede Menge wohlgeordnet werden kann. (Aus einem an Herrn Hilbert gerichteten Briefe.)* Mathematische Annalen **65**, 514–516.

Zermelo, E. (1908a), *Untersuchungen über die Grundlagen der Mengenlehre I*, Mathematische Annalen **65**, 261–281.

Zermelo, E. (1908b), *Neuer Beweis für die Möglichkeit einer Wohlordnung*, Mathematische Annalen **65**, 107–128.

CHAPTER 3

Development of the Concept of Homotopy

Ria Vanden Eynde

Prinses Lydialaan 46, B-3001 Heverlee, Belgium

1. Introduction

Homotopy is concerned with the identification of geometric objects (at first, paths) which can be continuously deformed into each other, these are then considered equivalent. The formal expression of this type of intuitive equivalence concept in terms of reflexivity, symmetry and transitivity was given in the late 1920's ([81, pp. 233, 234]) after a century-long history.

The origins of the homotopy concept for paths can be found within analysis where it was used as a visual tool to decide whether two paths with the same endpoints would lead to the same result for integration, or analytic continuation of a multi-valued function. The mathematization of the intuitive equivalence concept thus depended upon the objectives of the mathematicians who used it. We will see how this caused an ambiguity around the homotopy concept practically from the moment it originated: it got confused with other kinds of equivalences which were not immediately recognized as being different. Gradually, certain descriptions of the homotopy concept came to the front: constrained deformation (in which one or both endpoints are fixed) became favoured compared to free deformation. Although the latter is a more intuitive concept, the former will prove to be more interesting: it will allow for the introduction of a group structure.

This paper describes the history of the concept of homotopy of paths and its after-effects in, e.g., higher homotopy groups.

2. Implicit occurrences

2.1. *Implicit homotopy in the calculus of variations*

To tackle questions about the calculus of variations in the plane, the problem was often reduced to the determination of a curve represented by $y = \varphi(x)$ going from the point (x_1, y_1) to the point (x_2, y_2) which maximizes or minimizes an integral of the form $\int_{x_1}^{x_2} f(x, y, y')\,\mathrm{d}x$. Usually authors replaced the curve by a polygon $Q_1Q_2 \ldots Q_n$. For three consecutive points $Q_{k-1}Q_kQ_{k+1}$ they varied the ordinate y_k of the point Q_k such

HISTORY OF TOPOLOGY
Edited by I.M. James

that the new polygon satisfies the requirements. This provided them with equations in the increments Δy. By introducing new points such that the difference Δx approaches zero, they ultimately obtained a differential equation to determine the solution $\varphi(x)$.

Lagrange's [56, 57] method for the solution of the problem was to allow the whole curve $y = \varphi(x)$ to vary, which he described by introducing a function $y = \varphi(x, i)$ such that $y = \varphi(x)$ when $i = 0$. The variation of y, δy then corresponds to $\partial\varphi/\partial i(x, 0)$ and the problem is solved by substituting $y = \varphi(x, i)$ in the integral and demanding that the resulting function of i has derivative zero for $i = 0$. Of course, Lagrange assumed that the function $y = \varphi(x, i)$ could be represented by a power series in both x and i so that the function φ has derivatives of higher order with respect to i. Two intermediate curves, corresponding to two values of i, are continuously deformable into one another. So, when Dieudonné [27] treats Lagrange's method he was right to call it a first occurrence (albeit implicit and not in the most general form) of what was later to be homotopy of paths ("une première idée de ce qui sera plus tard l'homotopie"), see [27].

Jacobi and Weierstrass also handled the variation of curves with fixed endpoints. Jacobi refers to the curves of a family as "unendlich nahe Curven". Weierstrass uses the term "benachbarte Curven" (see [44, 90]).

2.2. *Integration of a complex function of a complex variable*

Since the integration of analytic functions along homotopic paths in the complex plane from which the singularities are removed gives the same value for the integral, one would expect this subject to provide the appropriate ground for the concept of homotopy to appear in, but these appearances often remained implicit. Most authors failed to recognize that there was a new concept worth mentioning and used continuous deformation of paths as a means to an end, a tool to describe certain situations.

2.2.1. *Cauchy's work on integration.* Before Cauchy introduced mobility of paths in 1825, an important step was made by Gauss [33] and Poisson [70]. Using the identification of complex numbers with points in the plane, both men noted that in an integral $\int_{x_0+\mathrm{i}y_0}^{X+\mathrm{i}Y} f(z)\,\mathrm{d}z$ different paths of integration may lead to different values. In [12], Cauchy fixes the meaning of the expression

$$I = \int_{x_0+\mathrm{i}y_0}^{X+\mathrm{i}Y} f(z)\,\mathrm{d}z$$

and examines how the value of the integral depends on the choice of the curve joining $x_0 + \mathrm{i}y_0$ to $X + \mathrm{i}Y$. Combining the identification of complex numbers to points in the plane and variational techniques he borrowed from Lagrange, Cauchy was able to formulate his results in a visual way using a "mobile" curve. He sets $x = \phi(t)$, $y = \chi(t)$ where ϕ and χ are monotone functions of the real variable t. He rewrites the integral as

$$A + \mathrm{i}B = \int_{t_0}^{T} f(\phi(t) + \mathrm{i}\chi(t))\big[\phi'(t) + \mathrm{i}\chi'(t)\big]\,\mathrm{d}t$$

and shows that the result is independent of the choice of ϕ and χ if the function remains finite and continuous for $x_0 \leqslant x \leqslant X$, $y_0 \leqslant y \leqslant Y$. In fact, as J. Grabiner, M. Kline, E.B. Jourdain and U. Bottazzini have already pointed out, he uses not only the existence but also the continuity of $f'(x)$. He also assumes ϕ, χ to be continuously differentiable. He says:

> If one wants to pass from one curve to another, which is not infinitely near the first, one can imagine a third mobile curve, which is variable in its shape, and have it coincide successively and at different instances with both fixed curves.

In his later work [13], we recognize what we now call a contractible loop and a contractible domain. Compared to 1825, the papers of 1846 and 1851 illustrate an evolution in Cauchy's formulations. Whereas in 1825 Cauchy starts with two curves joining two given points and expounds his results in terms of a third curve varying from the first to the second, in 1846 and 1851 he starts with closed curves enclosing a domain and considers variations of those curves. Moreover this domain may be on a curved surface.

Although it is obvious that a new concept enters the picture, Cauchy's approach remains phenomenological: though a certain property introduces itself, Cauchy does not think it is necessary to precisely define this property. He only describes an obvious situation. His purpose is only to define rigorously the meaning of the integral $\int_{x_0+iy_0}^{x+iy} f(z)\,dz$ in order to apply this theory to questions which are of interest to him. He calculates the value of definite integrals and integrates differential equations, he develops a function in a series assuming the coefficients can be represented by definite integrals evaluated over circles. These can be extended or contracted without changing the value of these coefficients. Here, also implicitly, the winding number occurs. This last aspect proved to be of special interest in astronomy where he used this method to develop the perturbating function.

2.2.2. *Riemann's doctoral thesis.* In his doctoral thesis of 1851 [74], Riemann introduces surfaces which cover a domain A in the complex plane. Where more than one layer ("Flächenteil") lies over A, Riemann defines a branch point of order $m - 1$ ("Windungspunkt $(m - 1)$-ter Ordnung") as a point where m layers of the surface are connected in such a way that after having made m circuits around the point, passing continuously from one layer to the next, one returns to the initial point. In 1857 [75], these surfaces are more explicitly constructed as a representation of multi-valued functions. In [74], Riemann defines the connectivity number of such surfaces as follows: A surface is called simply connected if every cross-cut (a cross-cut is a line which runs through the interior of the surface without selfintersections and joins one boundary point to another) divides the surface. A surface has connectivity number n if $n - 1$ cross-cuts turn it into a simply connected surface. These definitions assume the existence of a boundary. Closed surfaces are first provided with a boundary through a perforation and then treated as above. This way, the torus has connectivity number 3, a sphere with g handles has connectivity number $2g+1$. A function defined on such a surface and continuous (Riemann assumes analyticity) except perhaps in isolated points or lines can then be seen as a function of the variables x, y in the underlying complex plane. Riemann then considers two functions X, Y defined on a surface T with boundary, covering the domain A in the plane and proves that

$$\int_T \left(\frac{\partial X}{\partial x} + \frac{\partial Y}{\partial y}\right) dT = -\int_{\partial T} (X\cos\chi + Y\cos\eta)\,ds,$$

where χ is the angle between the interior normal to the boundary and the positive x-axis, η is the angle between the interior normal and the positive y-axis and s is the arc length

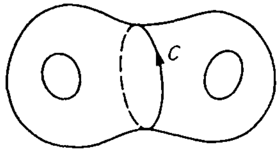


Fig. 1. C is null homologous but not null homotopic.

measured along the boundary of T. Here, Riemann implicitly assumes that the interior normal to the boundary can indeed be drawn, a result associated with the orientability of the surface he considers. If the integrand $\partial X/\partial x + \partial Y/\partial y$ vanishes, Riemann obtains the following results, p. 15:

> II. The value of the integral $\int (X(\partial x/\partial p)+Y(\partial y/\partial p))\,\mathrm{d}s$ calculated along the boundary curve of a surface covering A, remains constant as this surface is expanded or reduced, as long as no parts of the surface are included or removed by this operation, for which the assumptions of the previous theorem cease to hold.

His discussion here also shows a resemblance to Cauchy's paper of 1846 [14] (About the origins of Riemann's work see [5, pp. 221, 222]). If by expansion or reduction ("Erweiterung oder Verengerung"), Riemann means that one can omit parts of the surface with chosen boundaries (as is suggested in Satz I, p. 15), then the first part of Satz II refers to homology [77]. Cauchy however, varies the domain T by homotopic deformation of the boundary. If Riemann had seen Cauchy's work, it is possible he knew of this difference. In any case he does not mention this aspect and we cannot say anything about it. Later work shows that this attitude is at least confusing. An example of a closed curve which is null homologous but not null homotopic (see Figure 1) can be found in F. Klein's papers [48] as G. Hirsch [27] already pointed out. Klein however did not take up the opportunity to distinguish between null homologous and null homotopic closed curves. The difference between both concepts will explicitly be shown by H. Poincaré.

If a surface T contains singularities ("Unstetigkeitsstellen"), they are removed from T and the remaining surface is reduced, by making cuts, to a simply connected one. The integral $\int_{O_0}^{O} (Y(\partial x/\partial s) - X(\partial y/\partial s))\,\mathrm{d}s$ is then a function on T which changes its value along a curve corresponding to a cut which joins two branch points. At the end of his thesis, Riemann uses the Dirichlet principle to prove that every simply connected plane region with a boundary which contains more than one point (he also includes simply connected regions on a surface) can be mapped in a one-to-one conformal way onto the unit disk. Later this will lead authors to describe a contractible loop as a loop which can be "spanned" by a singular disk [82].

2.3. *Algebraic functions*

2.3.1. *Abel and Jacobi about elliptic functions, Cauchy's work on multi-valued functions.* Both Abel [1, 2] and Jacobi [43] considered the elliptic functions associated with the equation $w^2 = (1 - x^2)(1 - k^2x^2)$. The Riemann surface which belongs to this algebraic function is a torus. The integration of a rational function f of (x, w) on the torus from a point P_0 to a point P can give different values if carried out along different paths. Let the fundamental group of the torus from which the singularities of f are removed have base point O and generators a, b, $\{r_i\}_{i=1}^n$; a is a parallel loop, b a meridian loop and each of the r_i a loop which goes once round one of the singularities of f. Let c be a path from O to p_0. For the different paths γ, γ' joining P_0 to P the loop $c\,\gamma\gamma'^{-1}c^{-1}$ is homotopic to a combination of the generators a, b and $\{r_i\}_{i=1}^n$. Since the integration of closed differentials along homotopic paths give the same result,

$$\int_{c\gamma\gamma'^{-1}c^{-1}} f = k\int_a f + l\int_b f + \sum_{i=1}^{n} p_i \int_{r_i} f,$$

where $k, l, p_i \in \mathbb{Z}$. From this

$$\int_\gamma f = \int_{\gamma'} f + k\int_a f + l\int_b f + \sum_{i=1}^{n} p_i \int_{r_i} f\,.$$

The integrals $\int_a f$, $\int_b f$, $\int_{r_i} f$ are called periods of the integral $\int f$ and this shows where they come from. In this way homotopy can play a decisive role in the integration of multi-valued functions. On the other hand, so can homology, since the integral $\int_C \omega$ (where ω is a closed differential on a Riemann surface and C is a closed curve), depends only upon the homology class of C. We shall see below that Puiseux uses continuous deformation, while Riemann reasons in terms of homology. Neither Jacobi nor Abel were able to explain the double periodicity of the elliptic functions or the occurrence of the genus p basically because they did not have a method to handle multi-valued functions and their integrals. For the case of elliptic functions Cauchy and Puiseux clarified matters. For the general case, the solution was given by Riemann.

2.3.2. *Puiseux's paper on algebraic functions.* At the beginning of his paper of 1850 [72], Puiseux emphasizes that if a function u is given by an algebraic equation $f(u,z) = 0$, it may be multi-valued. If in the algebraic equation $f(u,z) = 0$ a value for z is chosen, the corresponding value for u is not always uniquely determined. To each value of $z = x + \mathrm{i}y$, Puiseux associates a point Z with rectangular coordinates x and y. Then in order to avoid the ambiguity mentioned above Puiseux takes an initial value c for z and chooses one of the roots u_i of $f(u,c) = 0$. Then the value of this root will change continuously as the point Z varies along a certain path from the point C, corresponding to $z = c$, to the point K, corresponding to $z = k$, as long as values of z are avoided for which equation $f(u,z) = 0$ has multiple roots for u, or in which u is infinite. Under these conditions, all the values of u which satisfy $f(u,c) = 0$ are continuous functions of z; Puiseux calls these solutions for u; $u_1, u_2, \ldots, u_n$ throughout the paper. The final value of u_i at the point K will be determined by the path followed. Puiseux explicitly emphasizes that this path may be a straight line, a curved line or a polygonal line. The only condition he imposes on the path is that it must form an uninterrupted line ("un trait non interrompu") between the points C and K. His description of a path thus remains intuitive. If z varies from the point C to the point K along a line CMK which does not pass through a point where u becomes infinite or equal to another root of $f(u,z) = 0$, then as z reaches k, u will attain a value h which is a root of $f(u,k) = 0$. Puiseux proves that this value h will be the same if the line CMK changes into a line $CM'K$ infinitely near CMK ("si la ligne CMK vient à se changer dans la ligne infiniment voisine $CM'K''$). He emphasizes: "c'est là une proposition fondamentale dans notre théorie". He then proposes to gradually alter ("altérer graduellement") the line CMK and formulates the following proposition (deformation always means continuous deformation), p. 370:

> If the point Z varies from C to K, along the path CMK, or along the path CNK, the function u_1, which in c equalled b_1, will yield the same value h_1 in both cases, assuming one can let the path CMK coincide with the path CNK after a deformation of the first without crossing a point for which the function u_1 becomes infinite or equal to another root of the equation $f(u, z) = 0$.

After explaining a process that we now call analytic continuation, Puiseux explains how to develop the functions u satisfying $f(u, z) = 0$ in the neighbourhood of a branch point. He also describes the behaviour of these solutions when z describes a contour going round a branch point; they form circular systems ("des systèmes circulaires") within which they permute. Puiseux then considers the following situation. He chooses a point $z = c$ where the equation $f(u, c) = 0$ does not have equal roots and joins this point to the points A, A', A'', etc., where the equation has multiple roots for the corresponding values $a, a', a'', \ldots$ for z. He assumes $f(u, a) = 0$ to have p roots equal to b. He joins C to A by a line which does not pass through any of the points A', A'', A''', etc. He then chooses a point D infinitely near A and draws an infinitely small closed contour $DNPD$ enclosing A and going round A once. Puiseux calls elementary contour ("contour élémentaire", see Figure 2) a contour consisting of the line CD, the contour $DNPD$ and the line DC, such that z running through this contour follows twice the line CD but in opposite directions. Poincaré will later call such a contour "un lacet". On this "contour élémentaire" the solutions $u_1, u_2, \ldots, u_p$ which become equal to b as Z goes from C to A behave in the same way as they did on the contour $DNPD$. They form a circular system. The other solutions $u_{p+1}, \ldots, u_n$ return to their original values. If Z starting from C describes a closed path around A which can be "reduced" to the elementary contour $CDNPDC$ without passing through any of the points A, A', A'', etc. the solutions will behave in exactly the same way as on the elementary contour. Puiseux's terminology remains intuitive: he uses the term "reduce to" without giving a more rigorous definition. Puiseux now draws elementary contours around every point A, A', A'', etc. where the equation has multiple roots. He then asserts that every closed curve which runs through C can be reduced to a sequence of "contours élémentaires" by deformation such that successive positions of the curve do not pass through any of the points A, A', A'', etc. and such that C remains fixed; p. 413:

> Join the point C to the different points A, A', A'', etc., by way of lines CDA, $CD'A'$, $CD''A''$, etc., Figure 12 which can be drawn arbitrarily except for the condition that none of these lines intersect; let D, D', D'', etc., be points on these lines which are chosen infinitely near the points A, A', A'', etc.; surround the latter points by infinitely small closed contours $DNPD$, $D'N'P'D'$, $D''N''P''D''$, etc., and consider the elementary contours $CDNPDC$, $CD'N'P'D'C$, $CD''N''P''D''C$, etc., which we will denote by (A), (A'), (A''), etc. Now, given any closed contour passing through the point C, it will always be possible to reduce this contour to a sequence of elementary contours, without passing through any of the points A, A', A'', etc., and such that the point C remains fixed during the deformation. This assertion does not need to be proved and a little attention is sufficient in order to perceive its exactness.

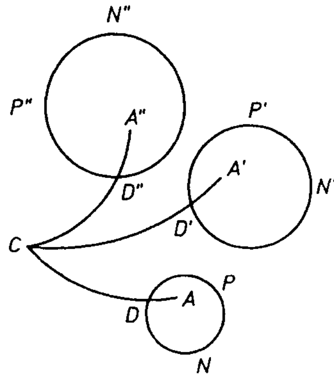


Fig. 2. Puiseux's elementary contours.

Puiseux emphasizes that it is important to specify the orientation of the curve, by introducing the notation $(+A)$, $(-A)$. He gives a method to find the corresponding sequence of elementary contours for a curve $CLMC$ if the reduction of the curve is given when it is run through in the opposite direction. From then on, he assumes that every contour can be represented by its corresponding sequence of elementary contours, which he calls "la caractéristique du contour". For closed contours reducible to the single point C he introduces the notation (o) for its "caractéristique", which can be added or omitted as often as one wants. Puiseux then states that, to every closed contour corresponds just one "caractéristique". Two contours, having the same "caractéristique" can be reduced to one another and conversely if two contours have different "caractéristiques" they cannot be reduced to one another. He implicitly assumes that a sequence $(+A)(-A)$ is reducible to the point C, otherwise the representation of a closed contour would not be unique. Although Puiseux's description is very near our understanding of the fundamental group of the complex plane from which A, A' etc. are removed, it would be wrong to interpret Puiseux's results in terms of implicit group theoretic thinking. Though Galois' paper [32] of January 16th 1831, which contains the germ of group theory had been published by Liouville in 1846, the group concept still had a long evolution to go through before reaching the abstract level needed here [93]. For a long time after Galois' paper, the group concept was restricted to the case of a set of permutations with the composition of these permutations as the law of composition. That is why Cayley's [15, 16] first attempt in 1854 (published again in 1878 [17]) to generalize the group concept to a set of symbols closed under a product found no response. Permutation groups were the only groups under investigation and so no generalization was needed. So it is obvious that Puiseux couldn't recognize the underlying group structure because there is no interpretation of the elements of the group (the closed curves based in C) as permutations working on something. It is exactly this association of a permutation to a closed curve based in C which will be used by Poincaré in 1895 to define the fundamental group. Even long after Puiseux's time groups were thought of as sets of permutations, and later more general transformations, closed under composition. Therefore it would be premature to interpret Puiseux's results, $(+A)(-A) = 0$ and $-(-A)(+A')(-A'') = (+A'')(-A')(+A)$ in terms of a law of composition. What is more, Puiseux does not consider the "caractéristique" of the sequence of two curves. Puiseux only associates a "symbol" to a given curve and his symbolic notation allows him to integrate functions along this curve in a simple way. His notation may even be inspired by this aspect as we shall see below. In this sense, using Wussing's terminology, Puiseux's paper has not made a causal contribution to the evolution towards the abstract group concept. Puiseux's paper is also typical of the way authors worked with homotopy and more generally with notions that now belong to topology. They reasoned intuitively, e.g., in terms of continuous deformations, also they did not give rigorous definitions of curves, lines, contours.

Although Puiseux's treatment of the behaviour of an algebraic function in the neighbourhood of a branch point is well-known [55, p. 641], no mention is made of his explicit attention to the continuous deformation of curves through a given point. Still, Puiseux's attention to the concept is meant to be an introduction to the use of continuous deformation in analysis. It is a means to reduce curves to a canonical decomposition along which the value of an algebraic function is calculated or along which integration is stepwise carried out. The homotopy concept is still linked to questions in analysis. First, Puiseux shows how the characteristic of a closed curve can help to obtain the final value of an algebraic function when the point Z runs through that curve. He substitutes the characteristic for the curve and follows the changes in the function as Z successively describes the elementary contours of the characteristic. Then he illustrates how the value of the function changes if the point Z follows a given path from a point C to a point K. In modern notation, Puiseux finds it intuitively clear that $\alpha \simeq (\alpha\beta^{-1})\beta$, which refers to his underlying assumption that a path followed by the same path in opposite direction is reducible to a point. This is analogous to the same implicit assumption for elementary contours; $(+A)(-A) = (0)$. Finally in the last part of his paper, Puiseux applies his results to the integration of algebraic functions. He considers the integral $\int_c^k u_1 \, \mathrm{d}z$, where u_1 is a root of $f(u, z) = 0$ and emphasizes that the value is determined only if besides the values c and k, the path between the corresponding points C to K is given. Moreover, as long as the path is continuously deformed without crossing any of the points A, A', A'', etc., where the equation has equal or infinite roots for u, the value of the integral does not change. (Weierstrass [89] also treats this question for polygonal paths. He does not use deformation terminology.) After introducing the notation $A_{+n}^{(i)}$, $A_{-n}^{(i)}$ (elementary integrals, "intégrales élémentaires") for the values of the integral $\int u_n \, \mathrm{d}z$ taken along the paths $(+A^{(i)})$, $(-A^{(i)})$ he calculates the value of the integral taken along an arbitrary closed path $CLMC$. He substitutes for the contour the corresponding characteristic and adds the values of the integrals along the elementary contours appearing in it, taking into account the changes of the algebraic function.

Puiseux then investigates whether relations exist between the elementary integrals, which brings him to questions referring to linear independent periods. Thereby he rediscovers the double periodicity of elliptic functions, first discovered by Abel and Jacobi.

2.3.3. *Riemann's paper of 1857 on Abelian functions and his "Fragment aus der Analysis Situs"* [75, 76]. At the beginning of his paper "Theorie der Abel'schen Functionen", Riemann explains how to construct a surface that can represent a multi-valued function. He had already used such surfaces in his thesis of 1851 but at that time he introduced them in terms of covering surfaces. In his paper of 1857, Riemann considers a function w of a complex variable $z = x + \mathrm{i}y$ which is given in a part of the z-plane ("in einem Theile der (x, y)-Ebene") and which satisfies the equation $\mathrm{i}(\partial w/\partial x) = \partial w/\partial y$. He states that the function w can be continued continuously outside the given domain in a unique way (if such a continuation is possible). According to the nature of the function w, he then distinguishes between two possibilities. Either the continuation of the function w will lead to the same value for w at a given value for z whichever the path may be along which the function is continued, or, different paths may lead to different values for w. In the first case, Riemann calls the function single-valued ("einwerthig") in the second case the function is called multi-valued ("mehrwerthig"). If the function is multi-valued, certain points exist in the (x, y)-plane such that if the function is continued around them, it will assume another set of values. (It is possible that Riemann knew Puiseux's work, but he does not mention it [5].) The different continuations ("Fortsetzungen") of one function for the same part of the z-plane are called branches of the function ("Zweige dieser Function"). A point around which one branch of a function goes over into another is called a branch point ("Verzweigungstelle dieser Function"). Where no ramification takes place the function is called "Einändrig" or "monodrom". Finally he shows how a function with branch points can be visualized geometrically by what we now call a Riemann surface.

For functions X, Y on a Riemann surface T which satisfy the equations $\partial X/\partial y = \partial Y/\partial x$ in local coordinates, he repeats his previous results, p. 92:

> The integral $\int(X\,\mathrm{d}x + Y\,\mathrm{d}y)$, evaluated along different paths joining the same endpoints, takes on the same value, if these both paths taken together form the boundary curve of a part of the surface T. If every closed curve situated in the interior of T forms the boundary of a part of T, then the value of the integral evaluated from a fixed initial point to a fixed endpoint remains constant and defines a continuous function of the upper limit of the integral, a function which is independent of the path of integration.

This means that, in modern terminology, for such functions X, Y the integral depends only upon the homology class of the path of integration. As mentioned earlier, Riemann had not noted the difference between null homologous and null homotopic closed curves. The above mentioned result leads Riemann (as in 1851) to distinguish between simply connected and multiply connected surfaces, p. 92:

> This leads to a distinction between simply connected surfaces, in which any closed curve is the boundary of a part of the surface – like for instance a circle – and multiply connected surfaces, for which this property is not valid, – like for instance the surface bounded by two concentric circles.

Riemann's definition of simple connectivity is in terms of homology, while our modern definition uses homotopy. For surfaces, both definitions are equivalent but not for higher dimensional manifolds. This was first shown by Poincaré [65] (see 3.1.2 below). The definition of the connectivity number of a surface Riemann gave in 1857 is not the same as the one he gave in 1851. The new definition perhaps derives from Riemann's interest in the integration of functions on a Riemann surface; the integrals of closed differentials on a Riemann surface depend only upon the homology class of the paths of integration, p. 92:

> If in a surface F closed curves $a_1, a_2, \ldots, a_n$ can be drawn, which neither separately, nor taken together form the boundary of a part of the surface F, but which, if taken together with any other closed curve do form the boundary of a part of the surface F, then this surface is called $(n+1)$-fold connected. This number associated to the surface is independent of the choice of the system of closed curves $a_1, a_2, \ldots, a_n$, since any other system of n closed curves $b_1, b_2, \ldots, b_n$, which can not bound a part of the surface, will, taken together with any other closed curve, form the boundary of a part of F.

As Maja Bollinger [4] points out, the curves a_i can be considered to be singular one-cycles without orientation. The system of those curves then forms a basis for the first homology group modulo 2 of the surface.

In order to repeat his results of 1851, used to integrate functions on a Riemann surface, Riemann describes how a surface of connectivity number $n + 1$ can be cut into a surface of connectivity number n. Here Riemann first thought of (orientable) surfaces with a boundary. If such a surface has, in modern terminology, genus g and r boundary curves, its connectivity number is $2g + r$. For (orientable) surfaces with no boundary, Riemann first punctures the surface. It is not always clear whether Riemann works with the punctured surface or with the closed surface. Riemann never specifies this, probably because the connectivity number as defined by him does not change if one punctures a closed surface. In this context, for closed surfaces, the cutting of the surface into a simply connected one has to be modified in such a way that the first cross-cut is a closed curve. Riemann then considers integrals of a function f on a closed surface T which has isolated singularities in points $\varepsilon_1, \varepsilon_2, \ldots$ on T. He reduces the surface T to a simply connected surface T' by making $2n$ cross-cuts. (These cuts correspond to what we now call generators of the fundamental group of T.) Then from a boundary point of T' Riemann makes cuts along lines l_i to each of the singularities ε_i such that no two lines l_i and l_j cut each other. On the new surface T'' the integral of the function f, $\int_{p_0}^{p} f$ is single-valued. The values of the integral of the function f taken along the cross-cuts are called "Periodicitätsmoduln". Each time a curve on T crosses a cross-cut or one of the lines l_i the value of the integral changes by a constant which can be calculated as $\pm \int_{a_i} f$ or $\pm \int_{b_i} f$, where a_i, b_i represent the cross-cuts, or as $\pm \int_{l_i} f$. The value of the integral $\int_{p_0}^{p} f$ on T is then determined up to the $2n$ constants associated with the cross-cuts and as many constants as there are lines l_i.

We can say that Riemann's work is directed towards homology theory (rather than homotopy theory): his definitions and his reasoning clearly refer to curves which form the boundary of a domain on the surface he considers. Also, to study integration on the surface T, Riemann reduces T to a simply connected surface T'. Then, he uses his result that the integral of an analytic function f taken along the boundary of a domain on T' is zero. Here, Riemann thinks of a simply connected surface as a surface on which every closed contour bounds a domain. So, the integral $\int_{p_0}^{p} f$ considered on T' is single-valued but becomes multi-valued on T. Because Riemann reasons along these lines and concentrates on the simply connected surface T' derived from the original surface T, we cannot decide at this point whether Riemann ever distinguishes between curves which are continuously deformable into one another and curves which bound a domain. In other words, we cannot say whether he was unaware of the difference between null homotopic closed curves and null homologous closed curves, or just neglected it.

In modern terminology, Riemann thinks of a simply connected surface as a surface with trivial first homology group mod 2. For orientable 2-dimensional surfaces, the triviality of the first homology group implies that its genus is zero, so that the surface is either a sphere or, if there is one boundary curve or one boundary point, the surface is a disk (as Riemann proved in 1851 using the Dirichlet principle), or the whole plane. These surfaces have trivial fundamental groups and all closed curves on them are both null homotopic and null homologous. This can explain why we did not find a difference between null homotopic and null homologous closed curves on surfaces in Riemann's papers.

At any rate, if we consider his "Fragment aus der Analysis Situs" (which we cannot date precisely) where Riemann aims to generalize the connectivity numbers of surfaces to higher dimensional manifolds ("Mannigfaltigkeiten"), it is obvious that he did not see the difference between null homotopic and null homologous closed curves. He defines the connectivity numbers for higher dimensional manifolds as follows, pp. 479, 480:

> If in the interior of a manifold each closed n-dimensional part of this manifold ("n-Streck"), taken together with m fixed of these "n-Strecke", which on their own do not form the boundary of a part of the manifold, forms a boundary , then this manifold has connectivity $(m+1)$ of dimension n. A connected manifold is called simply connected, if the connectivity number of each dimension equals 1.

The terminology seems to show that Riemann intends to use homology. Riemann's definition of simple connectivity does not correspond to the one used today. These definitions follow some statements about "Einstrecke". (The term "Einstrecke" is not clearly defined here. From the context it is clear that it can be interpreted as a one-dimensional complex. It is not clear whether such an "Einstreck" is homeomorphic to a line segment or whether it can have multiple points). From p. 479:

> Two "Einstrecke" are considered to belong to the same group or to distinct groups as they can be continuously deformed into one another or not.

The definition clearly suggests homotopy and even a partition into homotopy classes. But Riemann continues as follows, p. 479:

> Each pair of "Einstrecke", bounded by the same pair of points, together form a connected, closed "Einstreck" which bounds a "Zweistreck" or not, according to whether they belong to the same group or not. An interior connected closed "Einstreck" can , or cannot, form the boundary of an interior "Zweistreck".

(Here the term "Zweistreck" is to be interpreted as a 2-dimensional "part" of the manifold under consideration. On p. 481 Riemann says that "A connected "n-Streck" can or cannot be divided into separate pieces by cutting along any $(n-1)$-dimensional part of it." (... along any $(n-1)$-Streckigen Querschnitt".) This means that the "Zweistrecke" need not be simply connected). This shows that at that time Riemann overlooked the difference between null homotopic and null homologous closed curves. In [87] we can see that Listing knew that a closed curve can also be the boundary of a surface which is not simply connected. F. Klein was aware of this difference as can be seen in his work of 1882 [48]. In his work of 1905, as we shall see below, Poincaré proves that the two ways to partition closed curves do not correspond. Riemann's confusing attitude is again obvious in his definition on p. 479:

> An "n-Streck" A is deformable into another "n-Streck" B, if A and parts of B form the boundary of an interior "$(n+1)$-Streck".

The definition we first quoted immediately follows this. There Riemann appears to be using homology, while the above definition suggests he could be thinking of continuous deformation if we interpret the term deformable ("veränderlich") as "continuously deformable in". But if we interpret "veränderlich" in that way, then Riemann's definition is not correct. M. Bollinger [4] shows that the term "in einander veränderlich" is not synonymous with homologous. The same example shows that the term does not mean the same as "continuously deformable into" either.

From this we can conclude that at this time the distinction between homotopy and homology was still quite vague and remained so until Poincaré's work of 1905. Looking back we can say that in V. Puiseux's work the homotopy concept became a more explicit and clear-cut notion. But, in his description of "Riemann surfaces", Riemann introduced another new concept, the homology concept, without making explicit that this is not the same as the homotopy concept. Since Riemann was interested in the behaviour of functions on simply connected surfaces this ambiguity did not have much consequence for his results in that context. (On simply connected surfaces every closed curve is both null homotopic and null homologous.) It might even explain why he did not make note of the difference. But, Riemann's attitude, in view of his reputation, probably contributed to a confusion between homotopy and homology that lasted until Poincaré's work clearly distinguished between null homotopic and null homologous closed curves in 1905. In this way, Riemann's work means a regression in the development of the rigorous definition of the homotopy concept.

Riemann's ideas are continued by Betti in a paper of 1871 [3] (see [6]). Probably, Betti adopted Riemann's ideas through their conversations when Riemann stayed in Italy. His results also refer to homology rather than homotopy. Betti uses them to generalize results of the theory of integration to n-dimensional manifolds. He also considers manifolds we would now describe as having the same homotopy type.

2.3.4. *Jordan's work on the integration of algebraic functions* [47]. We cannot interpret Jordan's work on the integration of algebraic functions correctly, without referring to a paper [45] he published in 1866 on closed curves on surfaces. This paper does not refer to any problem in analysis and can thus be considered as a paper in topology. In his paper of 1866, Jordan starts off with the following definition, p. 91:

> Any two closed contours, drawn on a given surface, are called reducible into one another, if one can pass from one to the other by a progressive deformation.

And he remarks that, p. 91:

> Any two contours drawn in the plane are reducible to one another; however this is not true on any surface: for instance, on a torus a meridian and a parallel are two irreducible contours.

Jordan never mentions that all the surfaces (with or without a boundary or boundaries) he considers are orientable. Jordan then takes such a surface S with m boundary curves $A, A_1, \ldots, A_{m-1}$. He lets n be the maximum number of non-intersecting closed curves without multiple points on the surface which do not divide the surface into discon-

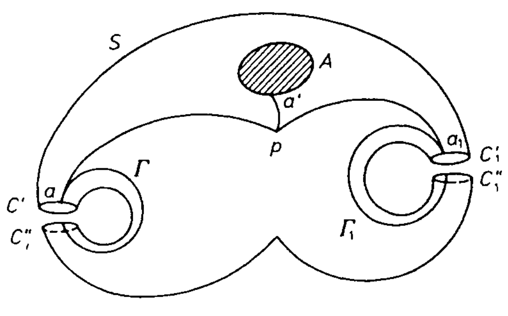


Fig. 3. Jordan's elementary contours.

nected pieces (n is now called the genus of the surface). He denotes these curves by $C, C_1, \ldots, C_{n-1}$ and cuts the surface along them. The resulting surface stays arcwise connected ("d'une seule pièce"). The two "sides" of the same cut C_i are denotes by C_i' and C_i''. For every cut C_i, Jordan draws a closed curve Γ_i through a point a_i on C_i. These curves Γ_i are drawn without multiple points and so that they do not intersect each other. The curve Γ_i meets C_i in the unique point a_i. Jordan then cuts the surface along these curves as well. See Figure 3 which illustrates Jordan's way of thinking. The points $a, a_i, \ldots, a_{n-1}$ are joined to a point p on the surface S by lines $pa, pa_1, \ldots, pa_{n-1}$ which do not intersect. He also joins p to points $a', a_1', \ldots, a_{m-1}'$ chosen on the boundary curves $A, A_1, \ldots, A_{m-1}$. The lines $pa', \ldots, pa_{m-1}'$ are also drawn so that they do not intersect. All the lines are drawn without multiple points. Jordan denotes the closed curves $pa_1 C_i a_i p$, $pa_i' A_i a_i' p$ and $pa_i \Gamma_i a_i p$ with the symbols $[C_i]$, $[A_i]$ and $[\Gamma_i]$. (Jordan does not specify the orientation of the curves C_i, A_i, Γ_i.) The curves $[C_i]$, $[A_i]$ and $[\Gamma_i]$ are called elementary contours ("contours élémentaires"; Jordan may have borrowed this terminology from V. Puiseux). Jordan also introduces the notation C^{-1} for the curve C traversed in the opposite direction.

Jordan's aim is to prove that every closed contour on S is reducible to a unique sequence of elementary contours. (In fact, since Jordan considers free deformations, the sequence is unique up to cyclic permutations of the elementary contours in the sequence.) In Jordan's paper, the difference between (constrained) homotopy and free homotopy is not considered. In general Jordan uses free homotopy. Therefore, when he says that two contours S and S' are irreducible to one another if their corresponding sequences of elementary contours are not identical, he should allow cyclic permutations of the elementary contours in the sequences. If he cannot pass from one sequence to the other by cyclic permutations of the elementary contours in the sequences, then the contours S and S' are not reducible to one another.

Interpreted with hindsight, Jordan obtains a set of generators for the fundamental group of an orientable surface of genus n and with m boundary curves, satisfying the relation

$$[A][A_1]\cdots[A_{m-1}][C][\Gamma][C]^{-1}[\Gamma]^{-1}\cdots[C_{n-1}][\Gamma_{n-1}][C_{n-1}]^{-1}[\Gamma_{n-1}]^{-1}\simeq 1.$$

In fact, in this paper, Jordan is very near to the abstract group concept. He had material at hand to describe an abstract group. Moreover, his notation is entirely adequate: he writes $C \cdot \Gamma$ for the contour determined by traversing C first and then Γ (although he uses the term sum ("somme") to denote such a product); he uses the notation $[\Gamma]^x$ to denote that the contour $[\Gamma]$ is traversed x times, $[\Gamma]^0$ means that $[\Gamma]$ is not traversed and $[\Gamma]^{-1}$ denotes the elementary contour $[\Gamma]$ traversed in the opposite direction. Throughout the paper Jordan applies the rule which says that two contours $[\Gamma]$ and $[\Gamma]^{-1}$ neutralize each other ("$[\Gamma]$ et $[\Gamma]^{-1}$ se détruisent"), such a product $[\Gamma]\cdot[\Gamma]^{-1}$ can be added or omitted as often as needed. He also implicitly assumes associativity. What is still missing to define the fundamental group is continuous deformation of based loops leaving the base point fixed and identification of homotopic loops. Jordan does not require that the contours pass through the point p, except for the elementary contours, nor does he emphasize the constrained deformation. But this is not the essential reason why Jordan does not recognize the underlying group structure. At this time groups were still thought of as permutation groups. The groups Jordan worked with were groups of permutations (which he called groups derived from a set of generators [46]) and later groups of motions. He used the product notation for the composition law. But in Jordan's paper of 1866, the contours he considered cannot be interpreted as operations working on objects; there is no set of permutations at hand and therefore Jordan could not interpret the material as belonging to a group structure [93, 71, 77].

We will now discuss Jordan's work on algebraic functions. In the second edition of the first volume of the Cours d'analyse, Jordan investigates the behaviour of a given algebraic function along different paths which join two arbitrary different points z_0 and ξ in the complex plane and defines the term equivalent ("équivalent"); pp. 224, 225. A "ligne" (or "chemin") is defined by the equations $x = \varphi(t)$ and $y = \psi(t)$ where φ and ψ are continuous.

> Suppose that z, instead of following the line L, follows another line L', also joining z_0 to ζ. The roots of the equation $f(u, z) = 0$, will vary continuously from the initial values $u_1^0, \ldots, u_n^0$ to the final values $v_1', \ldots, v_n'$. These new values are, as $v_1, \ldots, v_n$, roots of the equation $f(u, z) = 0$. They will thus coincide with these last roots except perhaps for their order of succession. If this order is the same, the two paths L and L' are called equivalent.

With this definition of Jordan, the equivalence of paths depends on the given equation $f(u, z) = 0$ and depends on the permutations of the n roots for u of the equation $f(u, z) = 0$. Saying that two paths are equivalent is not the same as saying that they are reducible to one another in Jordan's own terminology of 1866. Since the permutation of the roots $u_1 \ldots u_n$ does not change under continuous deformation, two paths which are reducible to one another are equivalent but not vice versa.

Jordan uses this concept of equivalence to reduce an arbitrary path joining the points z_0 and ξ to a "standard path" which is equivalent to it. The implicit assumption of the uniqueness of this "combinaison de lacets" is not correct because a closed contour can be equivalent to more than one sequence of elementary contours. Jordan's conviction that the reduction gives a unique sequence probably is the result of a confusion between the terms "équivalent à" and "reducible to" (as he uses this term in 1866) [87].

The concept of a Riemann surface as introduced by Riemann in 1851 was not easy to understand for mathematicians at that time. Riemann was aware of this himself. Later authors tried to clarify the concept and tried to fill up gaps left open by Riemann (often because he felt things were intuitively clear) [21, 58, 19, 42]. In [19] and [58], the authors both implicitly use the fact that homotopic loops around the branch points do not change the Riemann surface if it is constructed according to those loops. Then an efficient choice of these loops will simplify the construction of the surface. The above mentioned papers also provide us with more material to show how homotopy was implicitly used in analysis for integration and analytic continuation of functions.

2.3.5. *F. Klein's discussion of Riemann surfaces in connection with homotopy* [48]. Klein's discussion lies in the context of the study of flows of electric currents on surface of genus p where $p > 0$. Klein feels that Riemann might have been inspired by the same context when he treated complex functions on surfaces [5].

Klein's discussion can be represented as follows: u represents a non-constant everywhere regular potential function associated with a flow on the surface (regular means here that u is the real part of a complex potential function $u + iv$ which is finite and differentiable at each point). The function u must be multi-valued since a single-valued everywhere regular potential function on a closed surface is a constant. The change in potential along a closed curve C (streamline) is denoted by $\int_C du$ and is called the period of u on C. Two

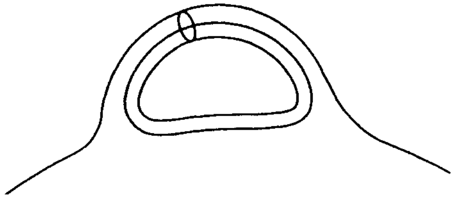


Fig. 4. Klein's "Querschnitte".

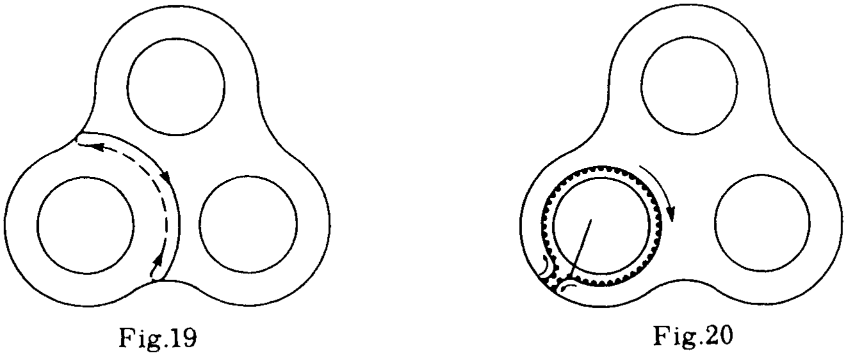


Fig. 5. Klein's Figures 19 and 20; a null homologous curve which is not null homotopic.

curves C and C' are considered equivalent if $\int_C du = \int_{C'} du$. Klein introduces a concept of equivalence which seems to be sufficient for studying flows on surfaces. It seems likely that the term equivalent ("äquivalent") should be interpreted as follows: two closed curves C and C' on the surface s are equivalent if and only if $\int_C du = \int_{C'} du$, where u is an arbitrary everywhere regular potential function on s. Hence, the term refers to homology. Klein, however, does not define this term precisely. From his paper it is clear that a sufficient condition for the equivalence of closed curves is that they are continuously deformable into one another. Klein also says that closed curves which go round ("völlig umschliessen" see Figure 19) one or more handles of the surface are "equivalent to zero" thus making it clear to us that homotopy is not a necessary condition for equivalence. In modern terminology, Klein finds that null homologous and null homotopic closed curves have zero periods (are equivalent to zero). For the curve in Figure 20 he says: "... this curve, like one which can be shrunk into a point, divides the surface into two separate pieces." Klein realized that null homology implied equivalence to zero and that null homotopic closed curves are null homologous. Indeed, application of a continuous deformation to a curve C does not affect the homology class of C even if the deformation is not constrained. This can easily be explained if we use the fact that the homology group of dimension 1 is the quotient group of the fundamental group by its commutator subgroup (see [83]). However, the exact relation between these two concepts (null homologous and null homotopic closed curves) is not made clear by Klein: both are considered as the boundary of a part of the surface. If we consider the Figures 19, 20 he draws, it seems likely that Klein "saw" the difference but had no way of turning it into a rigorous theory. (In 1905 Poincaré showed that the two ways to partition closed curves, homology and homotopy, are not the same.)

As in the case of integration of complex functions in the plane, the subject of Riemann surfaces as it is treated in works [37], reflects how homotopy remained implicit and intuitive. In these works, the concept is used in the context of continuation of algebraic functions along paths in the plane as described by Puiseux. For integration on Riemann surfaces, the authors follow Riemann's lines of thought and are, as we know now, directed to homology theory. In [37], we find the equivalence relation defined by homology explicitly emphasized and proved! Weyl [91] eschews the use of the homotopy concept by using covering surfaces in his definition of simply connected surfaces. Generally, the authors themselves however were not really aware that they in fact handled two different concepts, homotopy and homology. This mostly unconscious confusion can be seen in Klein's and Riemann's work as discussed above. Continuous deformation was a handy tool to decide whether two paths lead to the same value for an integral and since this theory provided mathematicians with an abelian structure (see [51, p. 675]), the unconscious confusion with homology becomes apparent to us. This situation resulted in an ambiguous "split" between mathematical methods used in the integration of functions on Riemann surfaces (continuous deformation) and the theoretically exact concept (homology) which is needed to discuss the theory of integration. The behaviour of algebraic or more general multi-valued functions themselves indeed requires homotopy since for instance, the group of permutations of the roots of an algebraic equation is in general not abelian. It seems likely that the study of multi-valued functions as in the context of uniformization and the solution of differential equations led to distinguishing between the concepts of homotopy and homology.

2.4. *Uniformization*

2.4.1. *Klein's work on uniformization.* Klein wrote several papers [49–51] where he discussed the possibility of representing the points of a given Riemann surface with equation $f(w, z) = 0$ using functions of a complex variable η, which reproduce themselves if η undergoes a transformation of the form $\eta' = (a\eta + b)/(c\eta + d)$, where $a, b, c, d \in \mathbb{C}$ and $ab - bc \neq 0$ (i.e. a linear transformation). For a discussion of these papers see [35]. Klein considers a fundamental domain ("Fundamentalbereich") N in the complex η-plane, the complete description of which is given in a paper of 1891 [52, p. 711]:

> The basic idea I used in Abh. CIII, is that for the definition of an algebraic surface one can use not only a closed Riemann surface, either given in space or given as a multiple-sheeted covering surface lying over a given surface, but one can also use a plane region, the boundary curves of which are to be identified in pairs according to a certain rule. I denoted such a plane region by fundamental region.

In what follows, Klein usually assumes that the sides of the "Fundamentalbereich" are circular arcs which are perpendicular to a given circle. This is an example of the concept of "Fundamentalpolygon" [53], which corresponds to our idea of a fundamental domain [29] of a group of linear transformations. A fundamental domain with respect to a given group of transformations is a region on the sphere, connected or not, of which no two points can be mapped onto each other by a transformation of the given group and such that the neighbourhood of any point of its boundary contains points which are the image of points of the given region by a transformation of the group. Obviously, a fundamental domain can thus be multiply connected.

Klein identifies this fundamental domain N with the Riemann surface S it represents. On this surface he considers a given single-valued analytic function F. He then reproduces this fundamental domain by applying the linear transformations φ_i in the η-plane which map those sides of the fundamental domain of each other which have to be identified to obtain the surface S (these are called congruent in [29]). If φ_i maps the domain N onto $\varphi_i N$ such that $\eta' = \varphi_i(\eta)$, Klein considers a function F' such that $F'(\eta) = F(\eta)$ on the side A of the fundamental domain N, where N and $\varphi_i N$ abut and such that $F'(\varphi_i(\eta)) = F(\eta)$. Because the function F coincides with F' on the side A, F and F' are identical. All the "copies" or "translates" of N, obtained by applying all the transformations of the group generated by the transformations φ_i, will, according to Klein, fill out a simply connected domain in the η-plane without overlaps.

Klein considers η as a function on the Riemann surface S. This function is multi-valued on S: to each point on S corresponds different η-values which are mapped onto each other by linear transformations in the η-plane (now called automorphisms). When a point P on S runs along a closed curve on S, η changes into the value $\eta' = (a\eta + b)/(c\eta + d)$. Klein reduces the surface S to a simply connected surface through $2p$ cuts along the curves A_i and B_i $(i = 1, \ldots, p)$ [48]. He joins these cuts to an arbitrary point O on the surface through curves c_i which join O to the point of intersection of A_i and B_i. Later he suppresses these curves c_i such that the curves A_i and B_i all pass through O. The function η is single-valued on the surface S' he thus obtained and which he knows is simply connected. When a point runs along a closed curve on the original surface S, it will in general pass over one or more of the curves A_i, B_i which correspond to the cuts. Each time this happens, the value η is transformed into η' by a linear transformation denote by S_i, T_i. These transformations are generators ("Erzeugenden") for all the transformations the value η can undergo, since a closed path on the surface can be replaced by a sequence of the curves A_i and B_i. The resulting group of transformations is, in general, not Abelian. As seen above, Klein remarks himself in [51, p. 675] that there is a difference between the theory of integration on Riemann surfaces and the discussion of the behaviour of the multi-valued function. In modern terminology, the fact that, in the context of this paper, in a sequence of paths the order can in general not be changed, implies that homology will not be sufficient to discuss the transformations η undergoes if a point P moves along a closed path on the surface. Klein also obtains relations ("Relationen") which exist between the transformations S_i and T_i associated with the cuts along A_i and B_i. He surrounds the given point O by a closed path of which he implicitly assumes that it has no multiple points. All the curves A_i and B_i will be crossed and η will reproduce itself at the end of the circumvolution. This gives the "Fundamentalrelation":

$$S_1^{-1}T_1^{-1}S_1T_1 \cdots S_p^{-1}T_p^{-1}S_pT_p = 1.$$

Klein then studies the fundamental domain of the group of transformations generated by the transformations S_i and T_i and reproduces it by making copies of which he proves, p. 684, that they will fill up a disk if $p > 1$. He also proves that any other relation $r = 1$ between the generating transformations must be an aggregate of transforms of the "Fundamentalrelation" ("If $R = 1$ is a relation, π any element of the group, then $\pi R \pi^{-1} = 1$ is called the transformed relation"). He reasons as follows: such a relation $R = 1$ consists of an aggregate of the transformations S_i, T_i say $R = V_1^{\varepsilon_1} V_2^{\varepsilon_2} \cdots V_r^{\varepsilon_r}$, where each V_i is a generating transformation and $\varepsilon_1 = \pm 1$. Then the copies of the fundamental domain (denoted by 1) are marked in the plane beginning with 1, $V_1^{\varepsilon_r}(1)$, $V_{r-1}^{\varepsilon_{r-1}} V_r^{\varepsilon_r}(1), \ldots$. Each of these copies has a common side with the previous one and since $R = 1$, the sequence of copies is "closed". Instead of handling this sequence of copies Klein uses a curve γ that runs through ("der Reihe nach") them. It is not clear whether γ may have multiple points. According to Klein, this curve can be deformed continuously in the domain formed by the copies from which the vertices are removed:

> Now we can shift this curve, without changing its meaning, with as the only condition that we can not pass over any of the vertices of the fundamental region. ... If we choose a point in the interior of our curve, we can replace the curve by a sequence of loops, which start from the chosen point, run around each one of the vertices of the fundamental region and then return to the initial point. In other words, our curve is equivalent to the successive revolution around certain of the vertices.

Here, the term equivalent ("äquivalent") does not have the same meaning as in Klein's previous paper [48] where it referred to homology. Klein remarks that the above relation between the transformations S_i and T_i corresponds to a circumvolution of the vertex O in the original fundamental domain. The circumvolution of any other vertex, which is the π-image of a vertex of the fundamental domain, leads to a relation which is the "transform" of the original relation; $R = 1$ can be written as an aggregate of such transformed relations and hence provides no new relation. Actually, if we interpret these results with hindsight, Klein obtains the isomorphism between the group of automorphisms of the universal covering surface of a surface S and the fundamental group of that surface.

2.4.2. *Poincaré's paper of 1883 on uniformization.* In the paper [63] Poincaré also proposed a uniformization theorem:

> Let y be an analytic, multi-valued function of x. Then one can always find a variable z such that x and y are single valued functions of z.

In this paper, Poincaré defines a surface S which will later be called the universal covering surface using analytic continuation of multi-valued functions on an underlying surface T. His definition of the surface S (universal covering surface) refers to the closed paths on the surface T: if a path is null homotopic on T, its lifting on S will be closed. The fact that Poincaré considers the behaviour of multi-valued functions on the surface T, and not the integrals of such functions, forces him to use continuous deformation, and not homology.

It seems likely that Poincaré was inspired by his work on automorphic functions. An analytic function F in a region G on the sphere is called an automorphic function relative to a group Γ of conformal homeomorphisms g of G onto itself if $F(g(w)) = F(w)$ for all $g \in \Gamma$. Both Klein and Poincaré published papers on automorphic functions. From their correspondence [54, 35] about the subject, it is clear that Poincaré knew that to a given Riemann surface corresponds a group of linear transformations in the plane. If the Riemann surface has genus $p > 1$ these automorphisms can be interpreted as isometries of the non-Euclidean plane. The collection of the copies of a fundamental domain of the group can be seen as the surface S he defines in his paper of 1883.

3. Continuous deformation as a topological concept

As seen above, continuous deformation plays an important role in the discussion of the behaviour of multi-valued functions. Gradually, mathematicians began to realize that there was a difference between this subject and the theory of integration. While the theory of integration provides an abelian structure (if on a manifold V, α and β are two closed paths based in a point P, $\alpha \cdot \beta$ is the path obtained by first traversing α and then β, and if f is a given function then the integral $\int_{\alpha\cdot\beta} f$ equals $\int_{\beta\cdot\alpha} f$ since $\int_{\alpha\cdot\beta} f = \int_\alpha f + \int_\beta f$), the permutations which the values of a multi-valued function undergo however need not be the same along the paths $\alpha \cdot \beta$ and $\beta \cdot \alpha$. To take into account the behaviour of multi-valued functions, mathematicians used the group of permutations their values undergo. At the basis of this group lie the topological properties of the paths in the space under consideration. This is a subject which belongs to analysis situs. So it seems natural that at this stage the attention shifted from the context in which mathematicians used continuous deformation to these deformations themselves. Gradually this evolution will result in an adequate terminology to work with. In the papers of Jordan, Riemann, Betti and Klein which we already discussed in the first part of this text, the interest in questions belonging to analysis situs is also apparent. In these papers, with the exception of Riemann's notes [76] (but we must take into account that these are only fragmentary) and Jordan's paper [45] of 1866, the discussion of these questions is still put in the context of analysis (theory of integration).

We shall now discuss other papers in which it became clear that the concept of continuous deformation actually belongs to analysis situs. Its use in analysis will become a secondary aspect. Continuous deformation gives information about the space in which it is considered. It even, in some cases, becomes a tool (finer than homology) to distinguish between spaces. For orientable surfaces, this classification problem was solved by Jordan in 1866 [45], the problem had also been treated by Möbius in 1863 [59]. In modern terminology, Jordan proved that two orientable surfaces with boundaries are homeomorphic ("applicable, l'une à l'autre sans déchirure ni duplicature") if and only if they have the same genus and the same number of boundary curves. In 1892 [64], Poincaré generalized the question, p. 187:

> The question remains whether the Betti numbers suffice to characterize a closed surface in the context of analysis situs, i.e. given two closed surfaces with the same Betti numbers, is it possible to pass from one to the other by way of a continuous deformation? This is true in 3-dimensional space and one might be led to believe that this remains so in an arbitrary space. This is not the case however.

Poincaré uses the term "surface" for higher dimensional manifolds ("variétés"). As we will see immediately below, Poincaré tackled the problem with the use of the fundamental group and thus provided algebraic methods to treat topological questions [38].

3.1. *Poincaré's work in Analysis situs*

3.1.1. *Sur l'Analysis situs* [64]. In this paper Poincaré introduces what he calls the group of a manifold. He considers unbranched multi-valued functions F_i on a manifold determined by an equation

$$f(x_1, x_2, \ldots, x_{n+1}) = 0.$$

If the functions F_i are continued along a loop they undergo a permutation. All the permutations which correspond to the possible closed paths on the manifold form a group. Poincaré emphasizes that this group depends upon the functions F_i. To take away this dependency he considers all such possible functions. The resulting group G can then be used to describe the manifold in the following way (p. 190):

> The group G can thus characterize the shape of the surface and be denoted as the group of the surface. It is clear that if two surfaces can be transformed into one another by way of a continuous deformation, their groups are isomorphic. The inverse, though less evident remains valid, for closed surfaces, so that the group defines the closed surface from the viewpoint of analysis situs.

For 2-dimensional manifolds, this result is correct. But, as Hirsch explains in the "Abrégé d'histoire des mathématiques" [39] the example of lens spaces considered by Alexander in 1919 illustrates that this last conjecture is false in general.

3.1.2. *Analysis situs*·[65]. In 1895 Poincaré digresses on the ideas of the above discussed paper. As an illustration for the functions F_i he considered in 1892, he uses the solutions of a given differential equation with analytic coefficients on the manifold ("variété") V under consideration. V is given by equations $f_\alpha(x_1, x_2, \ldots, x_n) = 0$ and inequalities $\varphi_\beta(x_1, \ldots, x_n) > 0$. In this paper, Poincaré introduces the term "lacet" (loop), p. 240:

> If the point M describes an infinitely small contour on the manifold V, the functions F will return to their initial values. This remains true if the point M describes a loop on V, that is to say, if it varies from M_0 to M_1 following an arbitrary path M_0BM_1, then describes an infinitely small contour and returns from M_1 to M_0 traversing the same path M_1BM_0.

In modern terminology, such a "lacet" is null homotopic. Poincaré uses the notation

$$M_0BM_0 \equiv 0 \quad \text{if } M_0BM_0 \text{ reduces to a loop.}$$

"To reduce to" is not explained, from what follows we can see that Poincaré interprets this as: along M_0BM_0 all possible functions F_i return to their original value.

The sequence of two paths $M_0AM_1BM_0$ and $M_0BM_1CM_0$ is written as $M_0AM_1BM_0 + M_0BM_1CM_0$ and the relation $M_0AM_1CM_0 \equiv M_0AM_1BM_0 + M_0BM_1CM_0$ reflects that all possible functions F_i behave the same way along the two paths in the two members which differ only in a path run through twice in opposite directions. Poincaré emphasizes that the path $M_0AM_1CM_0$ is not the same as the path $M_0CM_1AM_0$ and that the order of the terms in the above sum cannot be changed. The sum notation, tacitly introduced here, later becomes an explicit law of composition for paths [69, p. 523]. The path $-ECBFD$ stands for the path $ECBFD$ traversed in the opposite direction. Since the product of closed paths is not abelian in most manifolds, this sum notation is not very adequate. Compared to Jordan's paper of 1866 we discussed above, and where Jordan uses a product notation, Poincaré's notation constitutes a regression.

In his paper of 1895, Poincaré says (p. 241):

> $M_0BM_0 \equiv 0$, if the closed contour M_0BM_0 constitutes the complete boundary of a 2-dimensional manifold contained in V; and, in fact, this closed contour can then be decomposed into a very large number of loops.

This is not correct, the surface of which M_0BM_0 is the boundary should be simply connected. Poincaré adds (p. 241):

> This way we have to take into consideration relations of the form
>
> $$k_1C_1 + k_2C_2 \equiv k_3C_3 + k_4C_4,$$
>
> where the k are integers and the C closed contours drawn on V and starting in M_0. These relations, which I will call equivalences, resemble the above homologies. They differ from these:
> 1. Since, for homologies, the contours can start from an arbitrary initial point;
> 2. Since, for homologies, one can change the order of the terms in a sum.

In the fifth complement [68] which we will discuss below, Poincaré reformulates this as (p. 450):

> This way for homologies, the terms are composed according to the rules of ordinary addition; for equivalences, the terms are composed according to the same rules as the substitutions in a group; that is why the set of equivalences can be symbolized by a group which is the fundamental group of the manifold.

Even though he emphasizes this difference in working with the homologies and the equivalences there is no difference between his above definition for "$M_0BM_0 \equiv 0$" and for a "homologie" (p. 207):

> Consider a p-dimensional manifold V; let W be a q-dimensional manifold ($q \leqslant p$) contained in V. Suppose that the boundary of W consists of λ $(q-1)$-dimensional manifolds $v_1, v_2, \ldots, v_\lambda$. We will denote this situation by the notation $v_1 + v_2 + \cdots + v_\lambda \sim 0$.

To me it seems that Poincaré in this one paper uses the term equivalence for two different concepts (homotopy and homology). In [68] the meaning of "$K \equiv 0 \pmod V$"(K is a closed path (cycle) in the manifold V) is correctly defined as (p. 490):

> This means that there is a simply connected region in V, the boundary of which is formed by the cycle K.

For now, the fundamental group of the manifold V is defined as follows [65, p. 242]:

> This way, one can imagine a group G satisfying the following conditions:
> 1. For each closed contour M_0BM_0 there is a corresponding substitution S of the group.
> 2. S reduces to the identical substitution if and only if $M_0BM_0 \equiv 0$;
> 3. If S and S' correspond to the contours C and C' and if $C'' = C + C'$, the substitution corresponding to C'' will be SS'.
>
> The group G will be called the fundamental group of the manifold V.

The second condition will ensure that equivalent paths will lead to the same group element. For Poincaré this was probably intuitively clear. In modern terminology this means that homotopic closed paths lead to the same group element. Poincaré says nothing about the role of the chosen point M_0 probably because it was clear to him that any other point would lead to the "same" group since the manifolds he considers are arcwise connected. Nowadays, the fundamental group is defined in a more direct way. For two loops α and β based in the same point, the product $\alpha \cdot \beta$ is the loop obtained by first running along α and then along β. If we go over to the homotopy classes of such loops and define the product of two classes as the class of the product of two representing loops, this product does not depend on the choice of the representing loops and the group structure is ensured. Why Poincaré makes a detour along permutations to define the fundamental group is explained by Hirsch [38, 39]. Using Wussing's [93] results he says that the only groups mathematicians worked with at that time were groups of permutations, as for instance in Galois's and Jordan's papers, or groups of transformations, as in Klein's, Lie's and Jordan's work. As already mentioned earlier this shows why mathematicians did not readily use Cayley's abstract definition of a group (published first in 1854 and again in 1878), a definition to which Cayley himself added the result that every group is in fact a permutation group. To us now it is very easy to recognize a group structure in, for instance, Puiseux's and Jordan's papers, because we see a set of elements with a law of composition. At that time however, such a direct recognition was actually inconceivable because group elements had to operate on something as permutations or transformations do. The "product" is then the law of composition and associativity is ensured. As we shall see below for instance in Tietze's, Dehn's and Gieseking's work, the fundamental group will be introduced by these authors by means of generators and relations. It will gradually lose its characterization as a group of permutations.

After defining the fundamental group Poincaré explains how to calculate it for a given manifold V which is obtained form a polyhedron P_1 of which the faces are to be identified in pairs in a given manner. Poincaré says that the fundamental group will be derived (this is Jordan's terminology) from a set of principal permutations S_i ("substitutions principales") associated to closed contours C_i which he calls fundamental contours ("contours fermés fondamentaux"). Any other closed contour will be equivalent to a combination of these fundamental contours. The fundamental contours may satisfy a relation of the form $k_1C_1 + k_2C_2 + k_1'C_1 + k_3C_3 \equiv 0$ which Poincaré interprets as follows, p. 243:

> This means that the substitution $S_1^{k_1} S_2^{k_2} S_1^{k_1'} S_3^{k_3}$ reduces to the identical substitution.
>
> It is clear that we obtain the fundamental contours as follows. Let M_0 be a point interior to P_1, A a point on one of the faces of P_1, A' the corresponding point on the conjugated face. One will pass from M_0 to A, then from A' to M_0 without leaving P_1; the corresponding path on the manifold V will be closed. This way there are as many fundamental contours as there are pairs of faces. In order to form the fundamental equivalences: Consider a cycle of edges. Let, for instance, an edge be the intersection of the faces F_1 and F'_μ, which I therefore will call the edge $F_1F'_\mu$; let F'_1 be the conjugated face of F_1 and $F_2F'_1$ the conjugated edge of $F_1F'_\mu$ on this face; let F'_2 be the conjugated face of F_2 and $F_3F'_2$ the conjugated edge of $F_2F'_1$ on this face; etc. until we return to the face F'_μ and the edge $F_1F'_\mu$. Note that while performing this operation we can return, several times, to the same face. Let A_i be a point of F_i and let A'_i be the corresponding point on F'_i; let C_i be the fundamental contour $M_0A_i + A'_iM_0$. We will have the fundamental equivalence $C_1 + C_2 + \cdots + C_\mu \equiv 0$. This way there will be as many fundamental equivalences as there are cycles of edges. Once we have formed the fundamental equivalences in this way, we can deduce the fundamental homologies differing from them by the fact that the order of the terms is irrelevant. From these homologies the determination of the Betti number P_i will follow.

As follows from the above, Poincaré faulty definition of the equivalence $M_0BM_0 \equiv 0$ does not affect his calculation because he uses another criterion to decide whether a group element is the identity. It is clear that Poincaré got his inspiration from his work on automorphic functions (see [65, p. 247]; [61, 62]). In modern terminology he works with closed paths in the universal covering space. These correspond to null homotopic closed paths in the underlying space. Poincaré inattention may even be the result of his method since in the universal covering space any closed curve is both null homotopic and null homologous.

Poincaré illustrates his method by examples. These show that Poincaré works with what we now call homology with rational coefficients and this implies that at this stage the torsion coefficients escape his notice [4]. As Bollinger points out, Poincaré says in the second complement [67, p. 339]: "We will combine ... the homologies using addition, subtraction, multiplication and sometimes division." Thus Poincaré recognizes the difference between homology (with coefficients in $\mathbb{Z}$) and homology with rational coefficients. After a comment made by P. Heegaard on the duality theorem [65], Poincaré also points out that his definition of the Betti numbers differs from the one Betti gave. In modern terminology, whereas Betti works with homology mod 2, Poincaré uses homology with coefficients in $\mathbb{Z}$. One of the examples is a 3-manifold with non-trivial fundamental group but with the same Betti numbers as the 3-sphere, which causes Poincaré to redefine the term simple connectivity. Whereas Riemann and Betti had defined connectivity in terms of boundaries (i.e. for us homology) and simple connected manifolds as manifolds for which all connectivity numbers are 1, Poincaré reserves the term simply connected for manifolds with trivial fundamental group.

3.1.3. *Cinquième complément à l'Analysis situs* [68]. In the fifth complement (1904), Poincaré comes back to the "équivalences" he considered in 1895. He again relates the fundamental group of an orientable surface of genus p to a group of transformations in the plane, which as we now know is the group of automorphisms of the universal covering surface. In this paper, Poincaré explicitly tries to find a cycle on such a surface which is "homologue à zéro" but not "équivalent à zéro". The example corresponds to a contour as drawn by Klein in 1882 [48]. In this paper (see pp. 465, 466) we also find the explicit distinction between what we now call free and constrained homotopy ("équivalence impropre" and "équivalence propre") and the characterization of these using the universal covering surface. In the case of constrained null homotopy, the lifting of the curve to the universal covering space will be closed. In the case of free homotopy, if the liftings of the curves C, C' are MPM' and M_1QM_1' then C and C' will be freely homotopic if the automorphism mapping M to M' will also map M_1 to M_1'. Further on in this paper, Poincaré gives an example of a 3-manifold which is not simply connected although it has the same Betti numbers and the same torsion coefficients as the 3-sphere. It is the spherical dodecahedron space. He thus shows, in a dramatic way, that the concepts of homotopy and homology do not coincide. The question whether the triviality of the fundamental group of a closed 3-manifold implies that it is homeomorphic to the 3-sphere remains open. Poincaré's papers thus provide a rigorous basis for the discussion of homotopy. A missing aspect is the formal terminology to describe the continuous deformation. In 1912, Brouwer will fill up this gap. It seems likely that this delay results from the tendency towards combinatorial methods (Poincaré, Tietze, Heegaard, Dehn), while the basis to describe continuous deformation actually lies in the context of "general topology", the study of which began in papers by, among others Fréchet [30, 31], Schoenflies [78, 79] and Hausdorff [36].

3.2. *M. Dehn and P. Heegaard on Analysis situs [26]*

Whereas in the above papers, the discussion of topological questions tacitly assumes that the manifolds are embedded in a given Euclidean space the properties of which play no role, Dehn and Heegaard give a stepwise construction of manifolds ("Mannigfaltigkeiten") which are the union of abstract complexes. In the second part of their article, Dehn and Heegaard give an overview of results about complexes obtained thus far and put them in this new context (for instance results concerning homology, Poincaré's work and normal forms of surfaces). A discussion can be found in [4].

It is also in this article that we find the term "Homotopie" for the first time. We shall see below that Dehn and Heegaard's definition does not exactly cover our concept of homotopy. To explain this further we first analyse their discussion of transformations of complexes. Dehn and Heegaard distinguish between internal and external transformations of complexes. An internal transformation for a "Streckenkomplex" (one-dimensional complex) introduces a new point Q_0 on a "Strecke" (P_0^1, P_0^2) and replaces this "Strecke" by the two "Strecken" $(P_0^1, Q_0), (Q_0, P_0^2)$. An internal transformation for a "Flächenkomplex" (2-dimensional complex) adds a new "Strecke" $Q_1 = (P_0^1, P_0^2)$ where the points P_0^1 and P_0^2 belong to a circle which bounds a surface ("Flächenstück") in the complex and subdivides this "Flächenstück" in two "Flächenstücke". For a 3-dimensional complex, such a transformation comes down to a subdivision of the 3-dimensional elements. These transformations refer to what we now call subdivisions of complexes. Two complexes C_n and C_n' are called homeomorphic ("homöomorph" or "elementarverwandt") if they are identical maybe after a change in denomination of their elements or if they can be made identical after a sequence of internal transformations. To describe an external transformation, Dehn and Heegaard consider an n-dimensional manifold M_n $(n > 1)$ and a one-dimensional complex C_1 belonging to M_n. If for every point P_0^i of C_1 there exists a point Q_0^i in M_n (they tacitly assume that Q_0^i is different from P_0^i and that to different points P_0^i and P_0^k correspond different points Q_0^i and Q_0^k) and if for every "Strecke" (P_0^i, P_0^k) there exists a unique "Strecke" (Q_0^i, Q_0^k) in M_n joining the points Q_0^i and Q_0^k such that the circle $\pi_1 = \{(P_0^i, P_0^k), (P_0^k, Q_0^k), (Q_0^i, Q_0^k), (P_0^i, Q_0^i)\}$ bounds an elementary manifold in M_n ("Elementarmannigfaltigkeit": for dimensions 1 to 3 these are a "Strecke", the 2-sphere, the disk and the 3-ball), then the transition of the complex C_1 with points P_0^i to the complex C_1' with points Q_0^i is called an external transformation in M_n from C_1 to C_1. Although they did not explicitly require Q_0^i to be different from P_0^i they assume that they can join these points by a "Strecke". This shows an oversight of the case in which some of the points P_0^i coincide with some of the points Q_0^i. An analogous definition of an external transformation is given for two dimensional complexes. Here Dehn and Heegaard require M_n $(n > 2)$ to be an elementary manifold. They assume that, for every point P_0^i of a 2-dimensional complex C_2 there exists a point Q_0^i in M_n which again is tacitly assumed to be different from P_0^i. They again tacitly assume that to different points P_0^i and Q_0^k correspond different points Q_0^i and Q_0^k. If for every "Strecke" (P_0^i, P_0^k) there exists a "Strecke" (Q_0^i, Q_0^k) in M_n joining the two points Q_0^i and Q_0^k; for each "Flächenstück" determined by the circle π_1 there exists a "Flächenstück" in M_n determined by the points which correspond to the points on the circle π_1 such that the circle $\{(P_0^i, P_0^k), (P_0^k, Q_0^k), (Q_0^i, Q_0^k), (P_0^i, Q_0^i)\}$ is the boundary of an elementary surface ("Elementarflächenstück") E_2 in M_n; if moreover, two corresponding "Flächenstücke" can be used as elements on the boundary of a 3-dimensional "Elementarmannigfaltigkeit" E_3 in M_n then the complexes of the points P_0^i, respectively the points Q_0^i are said to go over in each other by an external transformation. These conditions could be satisfied even if M_n is not an elementary manifold so why they require M_n to be an elementary manifold is not clear. They say, p. 165:

> If two complexes transform into one another by way of a sequence of external transformations, possibly preceded by internal transformations, we call them homotopic.[1] We have the theorem: Two homotopic complexes are homeomorphic. For each complex one can find a homotopic one with given location. If in a given M_n all closed circles are the boundary of "Elementarflächenstücke", then one can find a homotopic complex to each complex in M_n (of dimension less than n) with given location for the "Strecken", etc. We thus obtain the following theorem: Two homeomorphic $(n - m)$-dimensional complexes, with or without singularities in a E_n are homotopic.

Bollinger concludes in her paper [4, p. 146]: "This way the concept of continuous deformations is put in combinatoric terminology." But the transformation described by Dehn and Heegaard is not the most general continuous deformation because for them complexes which are "homotop" are "homöomorph". Two lemniscates in the plane with the same orientation are "homotop". But a simple closed curve C_1' and a lemniscate C_1 cannot be called "homotop" because, on the lemniscate there are four "Strecken" with P as a vertex which cannot occur on the curve C_1' whichever point is chosen to correspond to P. Later in the paper, Dehn and Heegaard use the term "homotop" in the same meaning, as Jordan's term "reducible to". But Jordan's concept allows for more general continuous deformation than Dehn and Heegaard's which is nearer to our concept of isotopy [82, p. 14]. Dehn and Heegaard also introduce a concept of isotopy as follows. According to them, the external transformations are the composition of more simple transformations called elementary transformations ("Elementartransformationen"). If for integers m, n, where $m \leqslant n$, M_n is a given manifold and E_{n-m} is an elementary manifold on a closed manifold M_{n-m} in M_n; if E'_{n-m} is an elementary manifold on M_n which together with E_{n-m} constitutes the boundary of an elementary manifold E_{n-m+1} on M_n, then the replacement of E_{n-m} by E'_{n-m} is called an elementary transformation of M_{n-m} on M_n. For instance if $m = 1$, $n = 2$, the replacement of E_1 by E_1' in our figure (Figure 6) is an elementary transformation. Dehn and Heegaard say that such a transformation is an external transformation, but the endpoints of E_1 remain fixed and this is not allowed for external transformations as they defined them because of their implicit requirement that P_0^i differs from Q_0^i. For the same reason, these transformations cannot generate any external transformation. It seems likely that while Dehn and Heegaard described external transformations they had something else in mind which would allow for fixed points. For two manifolds M_{n-m} and M'_{n-m} without multiple points, multiple "Strecken", etc., isotopy is then defined as follows. If no interior point of the manifolds E'_{n-m} of E_{n-m+1} considered above belongs to M_{n-m}, then the elementary transformation is called a "spezielle externe Transformation" and two manifolds M_{n-m} and M'_{n-m} which go over into each other by such transformations preceded if necessary by internal transformations are called "isotop". It is clear that they think of

[1] The concept of homotopy in E_3 is the most often considered concept in analysis situs and is mostly called: "equivalence in the context of analysis situs" ("Äquivalenz im Sinne der Analysis situs"). Two homotopic curves on a surface are called reducible by Jordan [45, p. 100]. Two homotopic complexes in E_n are called homeomorphic by Poincaré [65, p. 1]. Because of the above fundamental theorem, these definitions are consistent with ours.

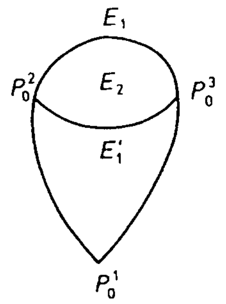


Fig. 6. $M_1 = P_0^1, P_0^2, P_0^3; (P_0^1, P_0^2), E_1 = (P_0^2, P_0^3), (P_0^3, P_0^1)$; M_1 is situated in the plane M_2.

E_{n-m}, E'_{n-m} and E_{n-m+1} as sets of points, an assumption which they explicitly make in the "Anschauungssubstrat". For curves for instance, this concept almost coincides with isotopy as we now use the term. It would be the same if the elementary manifolds E_1 and E'_1 were only to be included in the boundary of an elementary two dimensional manifold E_2, such that their boundary points would not have to coincide. In the "Anschauungssubstrat", Dehn and Heegaard say, p. 169: (domains are n-dimensional manifolds which are part of another n-dimensional manifold).

> Two complexes on a line, surface or a domain in space are continuously deformable into one another, with resp. without selfintersection, if and only if they are homotopic, respectively isotopic.

So, during the deformation the complex can intersect itself. This shows again that their description of the term "homotop" should for example allow for a continuous deformation in the plane from a circle into a lemniscate. Their definition of the term "homotop" does not include this possibility which they apparently also want to consider. Also, homotopy does not necessarily imply continuous deformation with self-intersection ("Selbstdurchdringung"). For curves on surfaces, they come back to the concept "Homotopie". Here, they introduce the product notation for the law of composition of two curves passing through a given point O. Curves are interpreted as sets of points traversed in a given direction. It is implicitly assumed that O is initial and endpoint of the curves considered here,

> Each curve is homotopic with a curve passing through a fixed point O. For two such curves π_1^1 and π_1^2 run through in a given direction, there is a unique curve π_1 (with singular point in O) defined as the two curves π_1^1 and π_1^2 run through in that order and given direction, we write: $\pi_1 = \pi_1^1 \cdot \pi_1^2$. We can now consider arbitrarily many curves $\pi_1^1, \ldots, \pi_1^n$, which may coincide partially and compose them as above, this composition satisfies the law of associativity; also, only cyclic permutations in the sequence of curves are allowed. Jordan now constructs a canonical fundamental system of curves through O, from which all the others can be obtained by composition, ...

They repeat Jordan's results of 1866 [45] and they remark (where Grundlagen Nr. 3 contains the definition of homeomorphism);

> Jordan's results are also important for higher dimensions: in the set of closed curves with given direction passing through a fixed point O one considers only the non-

> homotopic ones as different from each other. As in the above two such curves again define a unique curve passing through O and having a given direction. The set of curves passing through a given point of a M_n thus forms a discontinuous group, which is characteristic for the manifold. Indeed, as follows from Nr. 3 for each other point of M_n, and as one can easily see, for each point of a manifold which is homeomorphic to M_n, this group is the same. This group has been considered for the first time by Poincaré who named it the fundamental group of M_n.

Here, it becomes clear that they implicitly assume that a manifold M_n is arcwise connected (i.e. that every two points on M_n can be joined by a "Streckenzug" in M_n).

3.3. *Occurrences of homotopy in two papers by Tietze*

3.3.1. *Über die topologischen Invarianten mehrdimensionaler Mannigfaltigkeiten* [84]. We have seen that by the identification of the edges of a polygon 2-dimensional manifolds can be constructed. In the first section of his paper, Tietze generalizes this to construct higher-dimensional manifolds. This form of representation of manifolds will be called a "Zellensystem" or a "Schema" of manifolds. We assume that the cells of dimension 0, 1, etc. are closed and finite in number. For dimensions 2 and 3 the construction is described in detail. For higher dimensions the construction is only briefly sketched. To form the schema of an n-dimensional manifold, Tietze links up m-cells ("Zellen m-ter Dimension"), $m \leqslant n$ which are the generalization of vertices ("Ecken"), edges ("Kanten"), faces ("Lamellen") and 3-dimensional cells ("Zellen des dreidimensionalen Schemas"). A short discussion can be found in [4].

Before introducing the fundamental group of a manifold given by a schema, Tietze spends a paragraph on results belonging to group theory, p. 56:

> In this paragraph we will introduce certain concepts belonging to group theory, as we will consider below certain discrete groups associated to the given connected manifolds. We have to note here that the elements of this group are not certain operations which have a specific meaning, but that it is mostly the law of composition which is relevant so that we are dealing with the general group concept.

Tietze defines a group by generators and relations and describes how to go over from one set of generators and relations to another system of generators and relations without changing the group. To a group correspond the "zur Gruppe gehörende charakteristischen Zahlen" obtained after abelianization and which we now define as follows, see for instance [22]. For a finitely generated abelian group G there exists an integer $n \geqslant 0$, primes $p_1, p_2, \ldots, p_m$ and integers $r_1, r_2, \ldots, r_m$ $(m \geqslant 0,\ r_i \geqslant 1)$ such that G is isomorphic to the group

$$n\mathbb{Z} \oplus \mathbb{Z}_{p_1^{r_1}} \oplus \mathbb{Z}_{p_2^{r_2}} \oplus \cdots \oplus \mathbb{Z}_{p_m^{r_m}},$$

where n denotes the direct sum of n copies of $\mathbb{Z}$. If a group H is isomorphic to the group

$$l\mathbb{Z} \oplus \mathbb{Z}_{q_1^{s_1}} \oplus \mathbb{Z}_{q_2^{s_2}} \oplus \cdots \oplus \mathbb{Z}_{q_k^{s_k}}$$

then G and H are isomorphic if and only if $n = l,\ m = k$ and the numbers $p_1^{r_1}, \ldots, p_m^{r_m}$ and $q_1^{s_1}, q_2^{s_2}, \ldots, q_k^{s_k}$ are equal in pairs. Tietze proves that the isomorphism of two groups

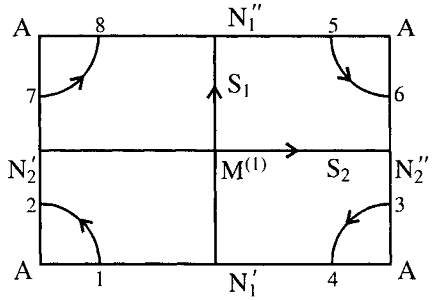


Fig. 7. Tietze's method for the torus.

(not necessarily Abelian) implies that the "zur Gruppen gehörenden charakteristischen Zahlen" must be the same. He realizes that this is only a necessary condition for the isomorphism of two groups. Poincaré had already shown that these numbers cannot determine the manifold's topology and conjectured that the role of the fundamental group might be decisive in classification problems. Poincaré's example of the spherical dodecahedron space also showed that these numbers do not determine the fundamental group. We now know that it is not possible to characterize finitely generated groups by a finite set of numerical invariants [60]. Poincaré was forced to work with the fundamental group as a whole.

Tietze calls the fundamental group a topological invariant belonging to the connected manifold ("einer der zusammenhängenden Mannigfaltigkeit zukommende topologische Invariante"). While Tietze only touches upon the connection between the fundamental group and the behaviour of multi-valued unbranched functions, for Poincaré this behaviour was the starting point for the introduction of the fundamental group. Its elements were interpreted as permutations of the values of multi-valued unbranched functions. With Tietze's paper, the elements of the fundamental group are being stripped of their conception as permutations although they are still referred to as "Operationen". The elements of the fundamental group are the closed paths based in a given point. The product in the group corresponds to our modern definition and Tietze obtains the relations in the group in a way analogous to our modern method by running along the circumference of a simply connected 2-cell (edge path group of a complex, see [84, pp. 77, 78], see also [28]). Referring to Poincaré, Tietze then shows how to calculate the fundamental group of a manifold given by its schema. To illustrate Tietze's method we will apply it on the torus which can be obtained from a rectangle by identification of the opposite sides. The point A in Figure 7 is represented by a closed cycle (Poincaré's terminology). There are two fundamental paths $S_1 = M^{(1)}N_1' + N_1''M^{(1)}$ and $S_2 = m^{(1)}N_2' + N_2''M^{(1)}$. The closed path L_A which surrounds A is represented on the schema by four arcs. When a point runs through L_A on the torus, its movement can be followed on the schema: the path joining position 1 to position 2 is deformable to $N_1'M^{(1)} + M^{(1)}N_2'$, the path joining position 3 to 4 is deformable to $N_2''M^{(1)} + M^{(1)}N_1'$ and so on, such that the sequence of the four arcs is freely homotopic to $S_2^{-1}S_1^{-1}S_2S_1$. Thus it becomes clear that the path L_A can be replaced by a sequence of the generators S_1 and S_2. This replacement corresponds to a continuous deformation of the path L_A such that it can be seen as a nontrivial combination of the fundamental paths, as Tietze says on p. 67 in his note 6):

> We can think of the closed line L_A around A as being deformed and stretched out so that it finally consists out of the segments $M^{(i)}N'_{A,h} + N''_{A,h}M^{(j)}$ of the corresponding partial paths $M^{(i)}N' + N''M^{(j)}$. The thus deformed line represents a closed line going round A, formed out of the fundamental paths, and the relation then says that any function continued along this path will return to its initial value, as has to be the case for an unbranched function y.

The path L_A which corresponds to the "Umgebungsmannigfaltigkeit" of the vertex A, represents the identical operation of the group because of the homogeneity condition required of the manifolds which Tietze considers: L_A is drawn within a simply connected schema. For n-dimensional schemata, the construction is analogous. To calculate the fundamental group of an n-dimensional schema, one of the points $M^{(i)}$ in the cells, $M^{(g)}$ is chosen to be base point ("Grundpunkt"). If the point $M^{(i)}$ is different from the base point then a composition ("Folge") of fundamental paths can be chosen such that there is a path $U_{g,i}$ joining $M^{(g)}$ to $M^{(i)}$ ("Hilfswege"). $U_{g,g}$ represents the identical operation ("identische Operation"). If S_λ is a fundamental path joining $M^{(h)}$ to $M^{(k)}$ then a closed path s_λ is defined by $U^{-1}_{g,h}s_\lambda U_{g,k} = S_\lambda$. The closed paths s_λ are called closed fundamental paths ("geschlossene Fundamentalwege") after Poincaré's "contours fermés fondamentaux". From the relations between the fundamental paths S_λ, he obtains the relations between the closed fundamental paths s_λ. For the first Betti number and the torsion coefficients which he calculates from the obtained fundamental group, he says, p. 80:

> While one can easily determine the equality of two sequences of numbers, one can in general not answer the question whether two groups are isomorphic. Contrary to other topological invariants, the fundamental group thus is an invariant of which the concurrence for two manifolds cannot be decided in every case.

This last assertion is somewhat premature, its proof is given 47 years later by Novikov [60].

Tietze also defines "Gleichartige Transformationen von Mannigfaltigkeiten in sich", which are a special case of homotopic homeomorphisms of a manifold V onto itself. As we know now, a continuous map f from a manifold V (as Tietze defined it) into itself which maps a point $M_0 \in V$ onto itself will induce an endomorphism of the fundamental group F of V. If f is a homeomorphism, the corresponding morphism is an automorphism of F. Two homotopic maps f and g induce the same morphism. An analogous reasoning probably inspired Tietze when he introduced the concept of similarity ("Gleichartigkeit") for homeomorphisms of a manifold V onto itself, pp. 88, 89:

> An invertible and continuous function of a manifold onto itself is called a transformation of the manifold in itself. Let t', t'' be two transformations of V in itself, P a point on V and P', P'' the points in which P is transformed by t', respectively t''. The distance between both transformations is less than ϵ if for each point P on V, the distance between the image points P', P'' is less than ϵ. Two transformations t_1 and t_2 of V in itself are called similar, if there exists a sequence of transformations $t(a)$ such that to each value of the parameter a, where $0 \leqslant a \leqslant 1$ corresponds a transformation $t(a)$, where $t(0) = t_1$ and $t(1) = t_2$ and $t(a)$ is a continuous function of a. This means that to each $0 \leqslant a_0 \leqslant 1$ and to each $\varepsilon > 0$ corresponds a δ such that for $|a - a_0| < \delta$ the distance between the transformations $t(a_0)$ and $t(a)$ is less than ε. Transformations of V similar to the identical transformation are called deformations of V in itself.[2]

[2] ... the introduced definition of distance between points is determining.

In modern terminology, a deformation is homotopic to the identical homeomorphism; Tietze illustrates his definition for the case of the torus. If x is a meridian and y a parallel on the torus, the periods ω_x, ω_y taken on along x and y by an elliptic integral of the first kind do not change if a deformation is applied on the torus. It is clear to him that the "Gleichartigkeit" of f with the identical homeomorphism implies that fx is equivalent to x, fy to y in the sense that for any differential ω of the first kind the integral $\int_{fx} \omega$ equals $\int_x \omega$ and $\int_{fy} \omega$ equals $\int_y \omega$. Tietze introduces the following groups of transformations for a manifold V which provide new topological invariants. T^0 is the group of "Transformationen" leaving a point M_0 in V fixed. D^0 is the group of "Deformationen" leaving the point M_0 in V fixed. (He thus implicitly uses the result that if f and g are gleichartig with the identical transformation then so is fg.) If the manifold V is (arcwise) connected ("zusammenhängend") the choice of M_0 is irrelevant. To each "Transformation" of T^0 corresponds an isomorphism of the fundamental group F of V, p. 90:

> As elements of the fundamental group F of V we can choose a system of paths a_1, a_2 … with M_0 both as initial and end point and which cannot be transformed into one another, the structure of F is then defined by rules for the law of composition of these paths.[3] To each transformation of T^0 corresponds a permutation of the paths a_i, in such a way that relations between paths are preserved after the permutation. The transformations thus determine isomorphisms of F onto itself. Similar transformations correspond to the same permutation, to deformations correspond the identical permutation, …

He thus is aware of the fact that if f is gleichartig with g then for a closed path a_i in V the paths a_i' and a_i'' in which a_i will be transformed by f, g will represent the same element in F, in other words they will be "ineinander überführbar". Apparently, Tietze does not think it is necessary to put this "equivalence" of a_i' and a_i'' in an analogous formal terminology. The concept of "equivalence" as it was defined by Poincaré in 1895 is clear enough for Tietze to work with and hence the idea of using an analogous formal definition does not come up. To each "Operation" (i.e. element) then of the group $G = T^0/D^0$ corresponds an "Operation" in the group of isomorphisms of the fundamental group F of F onto itself. If τ, τ', τ'' are "Transformationen" of V such that $\tau\tau' = \tau''$ then the associated isomorphisms j, j', j'' of F satisfy the relation $jj' = j''$. (We now formulate this in terms of functors.) This way Tietze explicitly introduced a formal definition for the "Gleichartigkeit" of two homeomorphisms f and g. The concept is not our concept of homotopy of maps, as Brouwer introduced it, since the maps $t(a)$ of the family are homeomorphisms. Tietze actually defines isotopy of homeomorphisms.

3.3.2. *"Sur les représentations continues des surfaces sur elles-mêmes"* [85]. In this paper, Tietze digresses on continuous transformations of surfaces as they had been defined by him in 1908. In 1913 the deformations are related to modifications of paths on the surface, an aspect which is less explicit in 1908. Here we find the following lemma:

> Given on a surface S two simple, continuous lines l, l' with the same endpoints, both of them Jordan curves, which only have their endpoints in common with the boundary of S, both lines having the same endpoints, there exists a deformation of S into itself, leaving each boundary point fixed and transforming l' into l.

[3] Two paths run through after each other again define a closed path. A relation between the closed paths $a_i a_k = a_l$ transfers to a corresponding relation between the substitutions of the values of an arbitrary unbranched function in V, which can be associated to the paths a_i, a_k, a_l.

The transformation form l to l' is not the most general form of homotopy, since the deformations are homeomorphisms as required in 1908. For a surface S, which in this paragraph is considered to be multiply connected, he announces the following theorem the proof of which is only roughly sketched, p. 510: G_S is the fundamental group of S.

> Two closed Jordan curves, l, l', drawn on S, having the same corresponding element in G_S, can be transformed into one another by a deformation of S.

Again there is confusion between homotopy and isotopy: two closed simple curves which represent the same element in G_S are homotopic but not necessarily isotopic.

3.4. *M. Dehn's papers on topology*

In [23] M. Dehn discusses how questions in group theory arise in a natural way from the context of topology. For instance, for based closed paths on an orientable surface of genus p (a closed path is called based at a point x_o if its initial and endpoint coincide with x_o) the question whether two such paths are homotopic (As Dehn says: "in einander reduzierbar mit Festhaltung eines Punktes") corresponds to the problem Dehn formulates as follows (p. 140): G is a group defined by generators a_i and relations. For G he uses the fundamental group.

> Our problems are: 1. To find a method by which one can determine in a finite number of steps whether two given operations of G are equal or not given their representation using the a_i and particularly, whether such an operation is the identity.

The question whether two closed paths based in the same point are freely homotopic corresponds to the second problem on p. 140:

> To find a method to decide in a finite number of steps whether for each pair of given substitutions S and T there exists a third one U with $S = UTU^{-1}, \ldots$

Indeed, two such closed paths α and β are freely homotopic if and only if there exists another closed path ν based at the same point such that α is homotopic to $\nu\beta\nu^{-1}$. To discuss these problems Dehn introduces the "Gruppenbild" of the group G. Such a "Gruppenbild" is defined as a (finite or infinite) "Streckenkomplex" C_1 of which the "Strecken" represent the generators of the group or their inverse such that the "Streckenzüge" (these are sequences of "Strecken" where the initial point of a "Strecke" coincides with the endpoint of the previous one in the sequence) of C_1 are in one-to-one correspondence with the elements of the group. A vertex Z is chosen as center of the "Gruppenbild" and represents the identity. If S is a "Streckenzug" of the complex C_1 starting from Z then Dehn associates to S the transformation of C_1 which maps Z to the endpoint of S. This way, to each group element corresponds a transformation of the complex in itself. The "Gruppenbild" is constructed in such a way that to two different elements of the group correspond two different transformations ("Bewegungen des Gruppenbildes") and vice versa. Dehn probably got the idea from the following results for surfaces. For surfaces a "Gruppenbild" of their fundamental group can be obtained [24] as follows: a net ("Netz") can be constructed, which corresponds to our idea of the universal covering surface. The net consists of meshes, these are polygons of which the edges correspond in pairs. Indeed, the one-dimensional complex consisting of the edges of this net can be used as "Gruppenbild". For two closed paths on the surface are homotopic iff the corresponding transformations of the net (now called automorphisms of the universal covering surface) are the same. Dehn adds that the "Gruppenbild", of which he explicitly proves the existence on pp. 141–144, cannot be constructed in a finite number of steps. In modern terminology, this discussion for surfaces relates to the fact that the group of automorphisms of the universal covering surface is isomorphic to the fundamental group of the surface. The same ideas had already been used by Poincaré in 1895. Dehn's student Gieseking used the idea to look for analytic criteria to decide whether two transformations of the "Gruppenbild" are the same or conjugate (whether the associated closed paths are homotopic or freely homotopic), see [34] and [87]. In his paper, it is the first time that homology for closed paths is this explicitly put in a group theoretic context: homologous closed curves represent the same element in the abelian group of the surface. The calculations for the transformations of the net into itself are very elaborate and thus not practical.

3.5. *Homotopy in Brouwer's papers*

In Brouwer's paper [7] we can find implicit occurrences of what we now call homotopic maps. This idea became more explicit in 1912 [8]. Then, Brouwer considers two continuous transformations of a surface into itself which can be continuously transformed into one another. The continuous modification of a continuous univalent transformation is described as follows (p. 527):

> By a continuous modification of a univalent continuous transformation we understand in the following always the construction of a continuous series of univalent continuous transformations, i.e. a series of transformations depending in such a manner on a parameter, that the position of an arbitrary point is a continuous function of its initial position and the parameter.

The explicit statement which says that the position of a point is a continuous function of its initial position and the parameter is the mathematically rigorous definition of the deformation process. It marks the transition of the intuitive understanding of this process to a rigorously defined concept and allows for the extension of the homotopy concept from paths to maps in general. Tietze had already described this dependency in 1908, but in a narrower context: the "Gleichartigkeit" of homeomorphisms of manifolds onto themselves (we now call this isotopy). In [8] transformations which can be obtained from one another by continuous modification are said to belong to the same class. As Freudenthal indicated [8] it is the first time that the term class is used in the sense of homotopy class. It is in these papers that Brouwer also introduced simplicial maps ("simpliziale Abbildung") and simplicial approximation ("modifizierte simpliziale Abbildung", later called simplicial approximation). If a continuous map is given which maps a polyhedron into another one, a simplicial approximation allows the map to be replaced by a map which, perhaps after a subdivision of the polyhedra, is "sufficiently near to it" and is simplicial, i.e. it maps simplexes onto simplexes by "piecewise linear maps". Seen from a methodological point of view the idea of approximating a map by a "piecewise linear map", and not by a map which is piecewise constant as in the case in analysis, was an important new step. These new ideas were used by Brouwer to introduce the "degree" of a map (calculating the algebraic sum of the number of times a point is covered by its image, the sign being determined by the reversal or the preservation of the orientation) and to prove that homotopy classes of maps of the 2-sphere in itself are characterized by their degree. This result sets off the search for characterizing or counting the mapping classes of maps of higher dimensional spheres.

Also in 1912, Brouwer gives a formal description of freely homotopic closed paths in a paper [9] in which he proves the topological invariance of closed plane curves as Schoenflies defined them. According to Schoenflies's [80] definition, a closed curve in the plane is a perfect bounded connected plane set which divides the plane in two regions of which it is the common boundary. As Freudenthal [8] mentions, Brouwer proves a broader result, namely the invariance of the number of domains determined by a bounded connected closed planar point set. Brouwer considers a bounded $(h+1)$-foldly connected plane region ("Gebiet") g and an arbitrary point P in g. For such "Gebiete", he knows that there are h simple closed fundamental curves $c_1, c_2, \ldots, c_h$ through P such that any continuous closed curve σ (i.e. the continuous image of a circle) can be deformed continuously in g into a finite composition of the curves c_υ (p. 523):

> Let g be a bounded $(h+1)$-foldly connected plane region, P any point on g. We can choose h, simple closed curves $c_1, c_2, \ldots, c_h$ through P, which are to be considered as fundamental curves, and which only intersect in P and have the property that each closed continuous curve σ in g can be transformed by continuous deformation in g into a canonic continuous curve ("kanonische stetige Kurve") φ consisting of a finite number of the curves c_υ^{**}.
>
> **This continuous deformation means that a plane region which is bounded by two concentric circles k_1 and k_2 can be continuously mapped in g such that σ corresponds to k_1, and φ corresponds to k_2.

This continuous deformation ("stetiger Abänderung") is an example of free homotopy. We now know that the fundamental group of the region g is the free group generated by h closed paths $c_1, c_2, \ldots, c_h$ each of which goes once round one of the h bounded regions determined by g. Since the fundamental group is a topological invariant the number of regions determined by g (number of generators of the fundamental group) does not change when a homeomorphism is applied to g. But it is exactly the construction of the generators and of the homotopic deformation which is most difficult. One has to define continuous maps α_i from the circle C into g for the paths c_i and for any given path σ one has to determine a continuous map $F : C \times [0,1] \to g : (x,t) \mapsto F(x,t)$. As seen above in Poincaré's and Tietze's work this problem was avoided by using what we now call a cell decomposition of the manifold under consideration. To define the fundamental group Tietze used a generator for each edge and the relations were obtained from running round the circumference of the 2-cells. Poincaré used what we now call the universal covering manifold and its automorphisms to justify an analogous method. If the fundamental group is defined in this way it has to be proved that the choice of a cell decomposition is irrelevant. Brouwer uses the concept of chains ("Ketten"), a concept used by Cantor [86] in the context of infinite plane sets. Brouwer describes them as follows (p. 523):

> By a chain we mean a finite, cyclic ordered set of points, by an ε-chain a chain for which the distance between two subsequent points is smaller than ε.

He also introduces modifications of these chains (p. 523):

> By an ε'-modification of a chain χ we understand firstly a displacement $< \varepsilon'$ of a point of χ, by way of which χ goes over into an ε'-chain, secondly an addition of a new point between two subsequent points of χ, by way of which χ goes over into an ε'-chain. If two subsequent points of a chain coincide, they are considered as the same

point. This way, after an ε'-modification, the number of points in the chain remains constant, increases with one or decreases with one.

Brouwer proves that for each $\varepsilon > 0$ small enough there exists an $\eta > 0$ and h ε-chains a_i such that to every ε-chain χ in π, a finite composition of the a_i can be associated which can be obtained from χ by ε-modifications. This is a property which is invariant under homeomorphisms, a result which follows from what we now call the uniform continuity of such maps on compact subsets of the plane. Brouwer's discussion is in terms of a metric space (the Euclidean plane) and the properties of continuous maps into such spaces. The result is put in the context of set theory. The use of ε-chains as introduced by Cantor [10], although he does not call them ε-chains yet, also reflects this. He thus obtains his result without having to rely on a cell decomposition of the set π; the existence of such a decomposition was proved later in 1925 by Rado [73]. Brouwer's paper also reflects the status of group theoretic thinking at that time. The ε_1-chains (more precisely the closed "Streckenzüge" determined by them) in g_ε through P which can be deformed into each other by ε-modifications can be considered as representing one element of a group, if during the modifications all intermediate "Züge" pass through P. This group is isomorphic to the fundamental group of g_ε. The fact that Brouwer uses free homotopy reflects that he did not think of using group theoretic terminology.

4. Higher homotopy groups

The consideration of based maps of an n-sphere ($n > 1$) into a space can in a natural way be seen as a generalization of based paths in a space and for $n > 1$ also, a group structure can be defined on the homotopy classes of these maps. This was done by Hurewicz [41] in 1934. Before Hurewicz, Čech [18] had already introduced these groups in a short note which did not attract much attention because these groups were Abelian. Since the homology group of dimension one is the fundamental group made Abelian, it contains only part of the information given by the fundamental group. Hence, it was expected that new concepts should be non-Abelian.

Hurewicz introduced the homotopy type: two spaces X and Y have the same homotopy type if there exist mappings $f : X \to Y$ and $g : Y \to X$ such that the composed mappings $f \circ g$ and $g \circ f$ are homotopic to the identity mappings. For instance a simplex or a full sphere have the same homotopy type as a point. Most of the invariants belonging to algebraic topology depend only upon the homotopy type of the topological space under consideration. For simplicial complexes, an equivalence is introduced using combinatorial methods by Whitehead [92]: two simplicial complexes are of the same simple homotopy type if one can pass from one to the other by a finite sequence of elementary simplicial operations.

Hurewicz also treated the classification of homotopy classes of mappings from one space into another. He showed that the description of homotopy groups can be reduced to that of a fundamental group, [39]. One can obtain the sphere S_{n+1} through certain identifications in the topological product (reduced suspension, smash product) of the sphere S_n and the segment $[0, 1]$. Using the equivalence between mappings of the product $Y \times Z$ into X and mappings from Y into the set of mappings of Z into X; by fitting it with the appropriate topology, passing to the homotopy classes and considering the role of the base point, one can demonstrate that the n-th homotopy group of X, $\pi_n(X)$ is isomorphic to the $(n-1)$-th homotopy group of the loop space with base point of X. This determines a recursive formula which finally reduces $\pi_n(X)$ to the fundamental group of $X^{S^{n-1}}$, this recursion can be described using adjoint functors. Since loop spaces are an example of H-spaces, as introduced by Hopf [40] (spaces where a product can be defined "up to homotopy"), their fundamental group is always Abelian, whence the Abelian character of the higher homotopy groups.

Higher homotopy groups are a valuable tool in the theory of obstructions, see [28]. However, the calculation of the higher homotopy groups proved to be difficult, its recursive formula not providing for an efficient calculation because the loop spaces are generally of infinite dimension. [39] mentions results of Hopf, Hurewicz, Freudenthal and Serre (using fibered spaces).

5. Conclusions

The above discussion illustrates how the introduction of homotopy (be it implicit or explicit) depends upon the interests of the mathematicians concerned and how it gradually acquires a more satisfactory definition. The equivalence of paths first meant for certain mathematicians that they led to the same value of the integral of a given, function or that they led to the same value of a multi-valued function. (See, for instance, [11–14, 72, 74, 75].) Later this dependency upon given functions is dropped (by Jordan for surfaces, by Riemann and most explicitly by Poincaré) and this leads to a concept which depends only upon the manifold. It thus becomes a concept which belongs to topology.

As a consequence of this hesitant evolution, there was at first a confusion between concepts. At a later stage relations between them were investigated, as for instance the fact that homotopy equivalence implies homology equivalence. In 1882, Klein gave an example of a closed curve on a surface which is the boundary of a part of the surface but could not be shrunk to a point. In 1904, Poincaré explicitly said that this curve is "homologue à zéro" (null homologous), but not "équivalent à zéro" (null homotopic). Poincaré obtains the homologies from the fundamental group by allowing changes in the order of the terms in the "équivalences". This means also that the equivalences imply homologies but not vice versa.

The success of algebraic methods in topology also had its influence: it explains the preference for theories with "base point" and constrained deformation even though free deformation is a more natural concept. This preference can now adequately be explained using the terminology of categories and functors as introduced by Eilenberg and MacLane in the 1940's. In the category of based spaces and based continuous maps there exists an object which is both initial and final as is also the case for the category of groups or rings and their homomorphisms. This is not the case in the category of spaces without base point, nor in the category of sets. Hence, constrained deformation leads more directly to the description of an algebraic structure (namely the fundamental group) associated with a topological space. (In general this group is not Abelian and hence cannot be characterized by numerical invariants as was the popular inclination around the turn of the century.) This initiated the association of algebraic structures with an object (topological space) to get information about topological properties of the object. In algebraic topology this association is functorial. This means that it is not only possible to associate an algebraic structure $A(S)$ with a topological space S; but the consideration of maps from such a space S into a space T leads in a natural way to a homomorphism between the associated algebraic structures $A(S)$ and $A(T)$. Such an association of algebraic structures in a functorial way with a topological space leads to fruitful results about the topological properties of the space concerned. For instance, if requirements on a map f between spaces X and Y implies incompatible algebraic properties for the morphism $A(f)$, then f cannot exist. Hence also, the richer the algebraic structure $A(X)$ the more abundant the collection of possible results will be.

Bibliography

[1] N.H. Abel, *Recherches sur les fonctions elliptiques*, (Crelle's) J. Reine Angew. Math. **2** (1827), **3** (1828). See also *Oeuvres Complètes*, 268.

[2] N.H. Abel, *Mémoire sur une propriété générale d'une classe très étendue de fonctions transcendantes* (1826), Mémoire des Savants Étrangers **7** (1841), 176–264. See also *Oeuvres Complètes*, 145–211.

[3] E. Betti, *Sopra gli spazi di un numero qualunque di dimensioni*, Ann. Mat. Pura Appl. (2) **4** (1871), 140–158.

[4] M. Bollinger, *Geschichtliche Entwicklung des Homologiebegriffs*, Arch. Hist. Exact Sci. **9** (1972), 94–170.

[5] U. Bottazzini, *The Higher Calculus: A History of Real and Complex Analysis from Euler to Weierstrass*, Springer, Berlin (1986).

[6] U. Bottazzini, *Riemanns Einfluss auf E. Betti und F. Casorati*, Arch. Hist. Exact Sci. **18** (1977), 27–37.

[7] L.E.J. Brouwer, *Beweis der Invarianz der Dimensionenzahl*, Math. Ann. **70** (1911), 161–165. See also *Collected Works*, Vol. 2, 430–434.

[8] L.E.J. Brouwer, *Continuous one-one transformations of surfaces in themselves*, Proceedings of Koninklijke Nederlandse Akademie van Wetenschappen **15** (1912), 352–360. See also *Collected Works*, Vol. 2, p. 527–535.

[9] L.E.J. Brouwer, *Beweis der Invarianz der geschlossenen Kurve*, Math. Ann. **72** (1912), 422–425. See also *Collected Works*, Vol. 2, 523–526.

[10] G. Cantor, *Über unendliche lineare Punktmannigfaltigkeiten*, Math. Ann. **21** (1883), 545–586. See also *Gesammelte Abhandlungen*, 165–209.

[11] F. Casorati, *Teoria delle Funzioni di Variabili Complesse*, Pavia (1868).

[12] A.L. Cauchy, *Mémoire sur les Intégrales Définies Prises entre des Limites Imaginaires*, Publiée chez de Bure, Paris (1825). See also *Oeuvres Complètes*, Série 2, Vol. 15, 41–89.

[13] A.L. Cauchy, *Analyse mathématique*, Rapport sur un Mémoire présenté à l'Académie par M. Puiseux: Recherches sur les fonctions algébriques, Comptes Rendus **32** (1851), 276. See also *Oeuvres Complètes*, Série 1, Vol. 11, 325–335.

[14] A.L. Cauchy, *Calcul intégral: Sur les intégrales qui s'étendent à tous les points d'une courbe fermée*, Comptes Rendus **23** (1846), 251. See also *Oeuvres Complètes*, Série 1, Vol. 10, 70–74.

[15] A. Cayley, *On the theory of groups, as depending on the symbolic equation* $\theta^n = 1$, Philos. Mag. **7** (1854). See also *Collected Papers*, Vol. 2, 123–130.

[16] A. Cayley, *On the theory of groups, as depending on the symbolic equation* $\theta^n = 1$, *2nd part*, Philos. Mag. **7** (1854). See also *Collected Papers*, Vol. 2, 131–132.

[17] A. Cayley, *The theory of groups*, Amer. J. Math. **1** (1878). See also *Collected Papers*, Vol. 10, 401–403.

[18] E. Čech, *Höherdimensionale Homotopiegruppen*, Verhandlungen des Internationalen Mathematiker-Kongresses, Vol. 3, Zurich (1932), 203.

[19] R.F.A. Clebsch, *Zur Theorie der Riemann'schen Fläche*, Math. Ann. **6** (1872), 216–230.

[20] R.F.A. Clebsch and P. Gordan, *Theorie der Abelschen Funktionen*, Teubner, Leipzig (1866).

[21] W.K. Clifford, *On the canonical form and dissection of a Riemann's surface*, Proc. London Math. Soc. **8** (122) (1877), 292–304. See also *Mathematical Papers*, 241–254.

[22] P.M. Cohn, *Algebra*, Vol. 1, Wiley, New York (1974).

[23] M. Dehn, *Über die Topologie des driedimensionalen Raumes*, Math. Ann. **69** (1910), 137–168.

[24] M. Dehn, *Über unendliche diskontinuerliche Gruppen*, Math. Ann. **71** (1912), 116–144.

[25] M. Dehn, *Transformation der Kurven auf Zweiseitigen Flächen*, Math. Ann. **72** (1912), 413–421.

[26] M. Dehn and P. Heegaard, *Analysis situs*, Enzyklopädie der mathematischen Wissenschaften mit Einschluss ihrer Anwendungen, Vol. 3, Teil I, Hälfte I, Teubner, Leipzig (1907), 153–220.

[27] J. Dieudonné, *Abrégé d'Histoire des Mathématiques 1700–1900*, Vol. 1, 2, 1ière ed., Hermann, Paris (1978).

[28] J. Dieudonné, *A History of Algebraic and Differential Topology 1900–1960*, Birkhäuser, Basel (1989).

[29] L. Ford, *Automorphic Functions*, Chelsea, New York (1951) (1st ed. 1929).

[30] M. Fréchet, *Les Espaces Abstraits*, Paris (1928).

[31] M. Fréchet, *Doctoral thesis*, Rendiconti del Circolo Matematico di Palermo **22** (1906), 1–71.

[32] E. Galois, *Sur les conditions de résolubilité des équations par radicaux* (1831), J. Math. Pures Appl. **11** (1846), 381–444.

[33] C.F. Gauss, *Über das Wesen und die Definition der Funktionen*, Gauss and Bessel (1811). See also *Werke*, Vol. 8, 90–92.

[34] H. Gieseking, *Analytische Untersuchungen über topologische Gruppen*, Hildenbach (1912).

[35] J. Gray, *Linear Differential Equations and Group Theory from Riemann to Poincaré*, Birkhäuser, Boston (1986).

[36] F. Hausdorff, *Grundzüge der Mengenlehre*, Leipzig (1914).

[37] K. Hensel and G. Landsberg, *Theorie der Algebraischen Funktionen*, Teubner, Leipzig (1902).

[38] G. Hirsch, *Comment la topologie est-elle devenue algébrique?* Cahiers Fundamenta Scientiae no. 100, Strasbourg (1982).

[39] G. Hirsch, *Topologie*. See also [27, Chapter 10, pp. 211–265].

[40] H. Hopf, *Eine Verallgemeinerung der Euler–Poincaréschen Formel*, Nachr. Gesellschaft Wiss. Göttingen Math.-Phys. Kl. (1928), 127–136.

[41] W. Hurewicz, *Beiträge zur Topologie der Deformationen. I. Höherdimensionale Homotopiegruppen*; *II. Homotopie- und Homologiegruppen*, Proceedings of Koninklijke Academie Wetenschappen Amsterdam **38** (1935), 112–119 and 521–528.

[42] A. Hurwitz, *Über Riemann'sche Flächen mit gegebenen Verzweigungspunkten*, Math. Ann. **39** (1891), 1–61.

[43] C.G. Jacobi, *Fundamenta nova theoriae functionum ellipticarum*, Regiomonti, Sumtibus Fratrum Borntraeger (1829). See also *Gesammelte Werke*, Vol. 1, 49–239.

[44] C.G. Jacobi, *Über die Curve, welche alle von einem Punkte ausgehenden geodätischen Linien eines Rotationsellipsoides beruhrt*, Nachlass. See also *Werke*, Vol. 7, 72–87.

[45] C. Jordan, *Des contours tracés sur les surfaces*, J. Math. Pures Appl. (2) **11** (1866), 110–130. See also *Oeuvres*, Vol. 4, 91–112.

[46] C. Jordan, *Sur la déformation des surfaces*, J. Math. Pures Appl. (2) **11** (1866), 105–109. See also *Oeuvres*, Vol. 4, 85–89.

[47] C. Jordan, *Cours d'Analyse*, Paris, Gauthier-Villars, 3 vols. (1882/1883/1887).

[48] F. Klein, *Über Riemanns Theorie der algebraischen Functionen und ihrer Integrale*, Teubner, Leipzig (1882). See also *Gesammelte mathematische Abhandlungen*, Vol. 3, 501–573.

[49] F. Klein, *Über eindeutige Funktionen mit linearen Transformationen in sich, Erste Mitteilung: Das Rückkehrschnittheorem*, Math. Ann. **19** (1882). See also *Gesammelte mathematische Abhandlungen*, Vol. 3, 622–626.

[50] F. Klein, *Über eindeutige Funktionen mit linearen Transformationen in sich, Zweite Mitteilung: Das Grenzkreistheorem*, Math. Ann. **20** (1882). See also *Gesammelte mathematische Abhandlungen*, Vol. 3, 627–629.

[51] F. Klein, *Neue Beiträge zur Riemannschen Funktionentheorie*, Math. Ann. **21** (1882), 141–218. See also *Gesammelte mathematische Abhandlungen*, Vol. 3, 630–710.

[52] F. Klein, *Über den Begriff des funktionentheoretischen Fundamentalbereichs*, Math. Ann. **40** (1891/1892). See also *Gesammelte mathematische Abhandlungen*, Vol. 3, 711–720.

[53] F. Klein, *Über die Transformation der elliptischen Funktionen und die Auflösung der Gleichungen fünften Grades*, Math. Ann. **14** (1878/1879). See also *Gesammelte mathematische Abhandlungen*, Vol. 3, 13–75.

[54] F. Klein, *Briefwechsel zwischen F. Klein und H. Poincaré in den Jahren 1881/1882*. See also *Gesammelte mathematische Abhandlungen*, Vol. 3, 587–621.

[55] M. Kline, *Mathematical Thought from Ancient to Modern Times*, Oxford Univ. Press, New York (1972).

[56] J.L. Lagrange, *Essai d'une nouvelle méthode pour déterminer les maxima et les minima des formules intégrales indéfinies*, Miscellanea Philosophica – Mathematica Societatis Privatae Taurinensis **2** (1760/1761). See also *Oeuvres*, Vol. 1, 336–362.

[57] J.L. Lagrange, *Leçons sur le Calcul des Fonctions*, Oeuvres, Vol. 10, 2nd ed., Courcier, Paris (1806).

[58] J. Lüroth, *Note über Verzweigungsschnitte und Querschnitte in einer Riemann'schen Fläche*, Math. Ann. **4** (1871), 181–184.

[59] A.F. Möbius, *Theorie der elementaren Verwandtschaft*, Berichte über die Verhandlungen der Königlich Sächsischen Gesellschaft der Wissenschaften zu Leipzig **15** (1863), 18–57. See also *Werke*, Vol. 2, 433–471.

[60] P.S. Novikov, *On the algorithmic unsolvability of the word problem in group theory*, Trudy Mat. Inst. Steklov 44 (1955); Amer. Math. Soc. Trans. Ser. 2 **9** (1958), 1–122.

[61] H. Poincaré, *Théorie des groupes Fuchsiens*, Acta Math. **1** (1882), 1–62. See also *Oeuvres*, Vol. 2, 108–168.

[62] H. Poincaré, *Mémoire sur les groupes Kleinéens*, Acta Math. **3** (1883), 49–92. See also *Oeuvres*, Vol. 2, 258–299.

[63] H. Poincaré, *Sur un théorème de la théorie générale des fonctions*, Bull. Soc. Math. France **11** (1882/1883), 112–125.

[64] H. Poincaré, *Sur l'Analysis situs*, Comptes Rendus Acad. Sci. Paris **115** (1892), 633–636. See also *Oeuvres*, Vol. 6, 189–192.

[65] H. Poincaré, *Analysis situs*, J. Ecole Polytech. **1** (1895), 1–121. See also *Oeuvres*, Vol. 6, 193–288.

[66] H. Poincaré, *Complément à l'Analysis situs*, Rendiconti del Circolo Matematico di Palermo **13** (1899), 285–343. See also *Oeuvres*, Vol. 6, 290–337.

[67] H. Poincaré, *Second complément à l'Analysis situs*, Proc. London Math. Soc. **32** (1900), 277–308. See also *Oeuvres*, Vol. 6, 338–370.

[68] H. Poincaré, *Cinquième complément à l'Analysis situs*, Rendiconti del Circolo Matematico di Palermo **18** (1904), 45–110. See also *Oeuvres*, Vol. 6, 435–498.

[69] H. Poincaré, *Sur un théorème de géometrie*, Rendiconti del Circolo Matematico di Palermo **33** (1912), 375–407. See also *Oeuvres*, Vol. 6, 499–538.

[70] S.D. Poisson, *Suite du mémoire sur les intégrales définies*, J. Ecole Polytech. **11** (1820), 295–341.

[71] J.C. Pont, *La Topologie Algébrique des Origines à Poincaré*, P.U.F (1974).

[72] V. Puiseux, *Recherches sur les fonctions algébriques*, J. Math. Pures Appl. **15** (1850), 365–480.

[73] Rado, *Über den Begriff der Riemannschen Fläche*, Acta Universitatis Szegediensis **2** (1925), 101–121.

[74] B. Riemann, *Grundlagen für eine allgemeine Theorie der Functionen einer veränderlichen complexen Grösse*, Inauguraldissertation, Göttingen (1851). See also *Werke*, 3–43.

[75] B. Riemann, *Theorie der Abel'schen Functionen*, Borchardt's J. Reine Angew. Math. **54** (1857). See also *Werke*, 88–142.

[76] B. Riemann, *Fragment aus der Analysis Situs*, Nachlass. See also *Werke*, 479–482.

[77] E. Scholz, *Geschichte des Mannigfaltigkeitsbegriffs von Riemann bis Poincaré*, Birkhäuser, Basel (1980).

[78] A. Schoenflies, *Die Entwicklung der Lehre von den Punktmannigfaltigkeiten I*, Jahresbericht der Deutschen Mathematiker-Vereinigung, Bd. 8 (1900), 1–250.

[79] A. Schoenflies, *Die Entwicklung der Lehre von den Punktmannigfaltigkeiten II*, Jahresbericht der Deutschen Mathematiker-Vereinigung, 2er Ergänzungsband (1908).

[80] A. Schoenflies, *Bemerkung zu meinem zweiten Beitrag zur Theorie der Punktmengen*, Math. Ann. **65**, 431–432.

[81] O. Schreier, *Die Verwandtschaft stetiger Gruppen im Grossen*, Abhandlungen aus dem Mathematischen Seminar Hamburgischen Universität **5** (1927), 233–244.

[82] H. Seifert and W. Threlfall, *Lehrbuch der Topologie*, Teubner, Leipzig (1834); Chelsea, New York (1947).

[83] C.L. Siegel, *Topics in Complex Function Theory*, Wiley, New York (1971).

[84] H. Tietze, *Über die topologischen Invarianten mehrdimensionaler Mannigfaltigkeiten*, Monatshefte Math. Phys. **19** (1908), 1–118.

[85] H. Tietze, *Sur les représentations continues des surfaces sur elles-mêmes*, Comptes Rendus Acad. Sci. Paris **157** (1913), 509–512.

[86] H. Tietze and L. Vietoris, *Beziehungen zwischen den verschiedenen Zweigen der Topologie*, Enzyklopädie der Mathematischen Wissenschaften (III) AB 13 (1930), 141–237.

[87] R. Vanden Eynde, *Historical evolution of the concept of homotopic paths*, Arch. Hist. Exact Sci. **29** (1992), 127–188.

[88] K. Weierstrass, *Definition analytischer Functionen einer Veränderlichen vermittelst algebraischer Differentialgleichungen*. Mathematische Werke, Vol. 1 (1842), 75–84.

[89] K. Weierstrass, *Vorlesungen über die Theorie der Abelsche Transcendenten*, Mathematische Werke, Vol. 4 (1875/1876).

[90] K. Weierstrass, *Vorlesungen über Variationsrechnung*, Mathematische Werke, Vol. 7 (1875/1879/1882).

[91] H. Weyl, *Die Idee der Riemannschen Fläche*, 1st ed. Teubner, Stuttgart, 1913; 2nd ed. Chelsea, New York, 1947; 3rd ed. Teubner, Stuttgart 1955.

[92] J.H.C. Whitehead, Proc. London Math. Soc. (2) 45 (1939), 243–327. See also *Mathematical Works*, Vol. 2, 1–83.

[93] H. Wussing, *Die Genesis des abstrakten Gruppenbegriffes*, VEB Deutscher Verlag der Wissenschaften, Berlin (1969).

CHAPTER 4

Development of the Concept of a Complex

Gerhard Burde

Fachbereich Mathematik der Johann Wolfgang Goethe-Universität, Robert Mayer Strasse 6-10, 60054 Frankfurt, Germany

Heiner Zieschang

Abteilung für Mathematik Ruhr-Universität, Universitätstrasse 150, 44780 Bochum, Germany

In algebraic and geometric topology and elsewhere in mathematics the concept of a *complex* played and plays till now a basic rôle: in simplicial and cellular homology, in homotopy theory, in obstruction theory, in the study of PL-manifolds, in the homological theory of groups, etc. Spaces which admit a decomposition into cells are often easier to handle than general topological manifolds. Fortunately, the majority of 'interesting' topological spaces fall into this category. These advantages counterbalance the cumbersome definitions and proofs of the main properties of complexes, in particular, invariance properties.

In the following we will describe the thorny way to the concepts used nowadays. It took quite a long time since the problems considered belonged to other disciplines of mathematics, mainly to analysis and geometry and the topological side became clear only later.

For the following treatment of the history of complexes we obtained much information from survey articles and books with historical remarks or flavour such as [7, 9, 10, 36, 37, 2, 31].

1. The origin of Analysis situs

The famous Königsberg bridge problem was to find a walk through the Prussian city passing every of its seven bridges crossing the river Pregel exactly once. To find a path by walking could be cumbersome, but to show that such one does not exist cannot be done by trial. The solution due to Euler [14] is considered as the first topological argument: He noticed that it suffices to consider the relative position of areas and bridges. If there would be a path as desired there could be at most two areas where an odd number of bridges terminate. This has been the first problem in graph theory, an active mathematical discipline with applications in many other areas. We will not deal with it here.

HISTORY OF TOPOLOGY
Edited by I.M. James

Another result of Euler became more important for algebraic topology: the Euler polyhedron formula. A polyhedron was considered as a convex 3-dimensional subspace of $\mathbb{R}^3$; Euler treated the case where the edges of the boundary are straight segments and the faces are planar polygons. He found and proved the formula $v - e + f = 2$ where the boundary consists of v vertices, e edges, and f faces [15, 16]. However, Descartes already found in 1630 the following theorem: The sum of the plane angles of the faces of a polyhedron is $2f + 2v - 4$, and this equals twice the number of space angles, that is, edges; [8]. This implies the Euler formula.

If one does not assume that faces are planar polygons and the edges straight lines one has to postulate that the edges are homeomorphic to segments. For weaker conditions 1813 L'Huilier constructed counterexamples to the Euler polyhedron formula [21]. One of them is obtained from a polyhedron bounded by two spheres. Probably he was lead to this construction by the study of some crystals.

Several mathematicians – among them Legendre, Cauchy, L'Huilier, v. Staudt, Schläfli, Listing, Jordan (a complete bibliography in [7]) – tried to get the Euler polyhedron theorem for surfaces with nonlinear edges and faces. The first satisfactory proof is due to v. Staudt [33]. The authors assumed different geometrical conditions, only in 1861 Cayley [6] and Listing [22] and in 1866 Jordan [18] recognized the topological nature of the Euler polyhedron formula.

The generalization of the Euler polyhedron formula to 3-dimensional complexes embedded in $\mathbb{R}^3$ or S^3 is due to Listing (1862), a student of Gauss, [22]. He also introduced the term *Topologie* and defined *complexes*. In a letter he explains how he got to the new concepts. We repeat part of it:

> ... Die erste Idee, mich mit der Sache (i.e. topology) zu versuchen, ist mir durch mancherlei Vorkommnisse bei den pract. Arbeiten in der Sternwarte zu Göttingen und durch hingeworfene Aeusserungen von Gauss beigekommen ....

Almost at the same time as Möbius (1858) he describes one-sided resp. non-orientable surfaces.

The generalization of the Euler polyhedron theorem to convex polyhedra in arbitrary dimensions is due to Schläfli (1852); in the formulation, however, there are gaps of the same type as there used to be in the proof before v. Staudt's.

Riemann applied and extended the arguments of the proof of the Euler polyhedron theorem in his articles on the behavior of functions of one complex variable. He obtains surfaces as branched coverings of (subsets of) the sphere by cutting and gluing several copies of the domain of the complex variable. He calculates the "Zusammenhangsordnung" of a surface (twice the genus) by determining – using modern terms – the rank of its first mod 2 homology group.

Since all surfaces and complexes were considered as embedded into $\mathbb{R}^3$, non-orientable surfaces were found quite late. Möbius (1858) described the Möbius band, [23]. Klein found the first non-orientable closed surface in addition to the projective plane: the Klein bottle. He also distinguished between the notions "non-orientable" as a topological property of the space and "one-sided" as a property of the embedding of an n-manifold into an $(n + 1)$-manifold. The properties "non-orientable" and "one-sided" have not been distinguished by all authors, for instance, not by Poincaré; in [7] Dehn and Heegaard used only the term one-sided, meaning non-orientable. v. Dyck [13] classified polyhedral closed surfaces by genus and orientability.

2. Complexes and homology

The development of combinatorial and algebraic topology into an important mathematical field is mainly due to the work of H. Poincaré. Most of his ideas can be found in the note [25] and in the article [26]. He obtained that the alternating sum of the Betti numbers equals the alternating sum of the numbers of cells of different dimensions. His definitions and methods were often not precise and he had to correct them in later articles and this led to new rich concepts. For instance, in the study of manifolds and their homology – responding to critical remarks of Heegaard [17] from 1898 – Poincaré restricted himself to spaces decomposed into cells, to *polyhedra*. In [27] Poincaré used complexes, considered as subspaces of $\mathbb{R}^n$. The complex is determined by *incidence matrices* which state whether a k-simplex is part of the boundary of a $(k+1)$-simplex. Simplicial complexes were known at that time, Poincaré added subdivision properties. (The term "triangulation" had been introduced by Weyl [41].) Generalizing intersection properties of linear subspaces in $\mathbb{R}^n$ to submanifolds of n-dimensional manifolds (tacitly assumed to be orientable) Poincaré obtained that the Betti numbers in the dimensions k and $n-k$ coincide (Poincaré duality theorem). According to him this result was known to and applied by several mathematicians but not formulated as a theorem, [28]. The proof depends on the construction of dual bases and this is done by the "barycentric subdivision". These barycentric subdivisions can be applied to arbitrary cell decompositions and the result is an (abstract) simplicial complex; of course, the dual complex will not be simplicial even if the original one is simplicial.

The first abstract definition of a complex was given in 1907 by Dehn and Heegaard [7]. The question to what extent the results are those of the space or only of complexes, the so called Hauptvermutung, was formulated in 1908 by Steinitz [34]. Using "abstract" constructions of complexes Gieseking first described in 1912 a hyperbolic 3-manifold, see Chapter 15 of this book (3-manifolds). Brouwer introduced in 1911 simplicial mappings and showed the existence of a simplicial approximation of a continuous mapping between two polyhedra, [5].

3. Complexes and homotopy

From the applications of complexes we have mentioned only those to homology. Let us now briefly recall another one: the fundamental group or first homotopy group π_1, also introduced by Poincaré [26]. In general, it is difficult to calculate, even to describe the fundamental group. For the fundamental group of a complex, however, v. Dyck (1882) introduced a presentation $\langle S \mid R \rangle$ by a system S of generators corresponding to the edges and a system R of defining relators corresponding to the 2-dimensional cells [12]. Every group admitting a finite presentation can be realized as the fundamental group of a finite 2-dimensional complex (or a closed orientable 4-manifold [31, p. 180]). The presentation of π_1 depends on the complex and on the choice of a maximal tree; hence, the fundamental group of a polyhedron has many presentations and now the "isomorphy problem" of Tietze arises of deciding whether two presentations define isomorphic groups. In 1908 Tietze describes a procedure for connecting two presentations of the same group, [38]. But it turns out that, in general, the isomorphism problem cannot be decided by a finite algorithm. The narrow relationship between the theory of complexes and the combinatorial group theory has been studied in many articles and is still in mathematical research nowadays.

The well-known Seifert–van Kampen theorem is obvious for complexes, see [30, 39].

4. Formal foundation of PL-topology

During the later twenties and the early thirties a group of authors was concerned with the problem of giving a solid foundation to the new branch of topology which was called "combinatorial". Among them were J.W. Alexander, E. Bilz, L.E.J. Brouwer, M. Dehn, S. Lefschetz, M.H.A. Newman, K. Reidemeister, A.W. Tucker, O. Veblen, J.H.C. Whitehead – an incomplete list. The central object – the complex, respectively the polyhedron – was defined in various ways, which differed only slightly. The same holds for the notion of equivalence of complexes and the moves introduced to generate equivalence. In many contributions the special case of manifolds was considered which presented extra difficulties in its combinatorial version. The process of convergence was furthered by a paper of Alexander [1] which clearly had a lasting influence. His definition of a complex, "symbolic" or "rectilinear", was generally accepted as well as his "simple transformations" and his elegant polynomial formalism. Preparatory work was done by Newman [24]. According to Alexander an n-simplex $A = a_0 \cdot a_1 \cdots a_n$ is the product of its vertices a_i, and a k-component the product of any $k+1$ ($k \leqslant n$) of these vertices. A (finite) complex K is a polynomial with coefficients in $\mathbb{Z}/2\mathbb{Z}$ in variables a_i, the vertices; the homogeneous terms of degree $d+1$ define a "symbolic d-complex" which is the fundamental notion in this paper. (In [31] this is called a "pure" (rein) complex.) Each summand of the polynomial represents the simplex and its components defined by the variables that occur in it. The complex $\overline{K}$ (the boundary) consists of the $(d-1)$-components of K counted mod 2. This polynomial algebra can be used for a formal description of geometric operations in K: the product corresponds to the "join", and the boundary to a derivation. The crucial move – a simple transformation – replaces $K = A \cdot P + Q$ by $a \cdot \overline{A} \cdot P + Q$, A is a component of K, a is a new vertex. This transformation and its inverse – $(A, a)^{\pm 1}$ – generate the combinatorial equivalence of complexes – together with isomorphisms. It is the "stellar" subdivision also written as:

$$K \to \big(K - A \cdot \operatorname{link}(A; K)\big) \cup \big(a \cdot \overline{A} \cdot \operatorname{link}(A; K)\big).$$

There is a natural correspondence between symbolic complexes and rectilinear simplicial complexes. The latter is a space in some $\mathbb{R}^N$ composed of a finite set of Euclidean simplices such that any two simplices either are disjoint or have a face in common. The space – a polyhedron – can be partitioned into simplices in many ways. Two rectilinear complexes are equivalent, if they possess isomorphic partitions. Alexander proves that combinatorial equivalence corresponds to the rectilinear one.

Two German textbooks on combinatorial topology appeared in the thirties: One was the famous "Lehrbuch der Topologie" by Seifert and Threlfall, the other one was Reidemeister's "Topologie der Polyeder". The latter is a remarkable example of "purity in method". It presents all the current results of the time on the basis of a strictly abstract combinatorial approach using consequently Alexander's polynomial algebra. Reidemeister's philosophically inclined understanding of mathematics led him to insist on the importance of fixing clearly the logical pattern in which a theory was to be axiomatically developed. Seifert and Threlfall on the other side pragmatically declared themselves to the "méthode mixte" in the preface of their book. Combinatorial arguments and the use of continuity are applied according to the situation in hand, an attitude that on the whole won the day. In the sense of Reidemeister's definition this book is not "on combinatorial topology" – it does not even introduce the notion of equivalence for simplicial complexes, but uses simplicial approximation instead. It is, though, one of the most successful monographs on topology ever written; it was first published in 1934, reprinted as a Chelsea edition after the second world war, and finally translated into English in 1980.

5. CW-complexes

In the following decade, the forties, the theme of combinatorial topology was clearly dominated by J.H.W. Whitehead. He had been in Princeton on a Commonwealth Fund Fellowship (1929–1932), and under the influence of Veblen, Lefschetz and Alexander turned to combinatorial topology. In his fundamental paper [42], he expanded the theory of complexes significantly. Adjoining an n-simplex σ^n to a simplicial complex K, where all $(n-1)$-faces of σ^n but one belong to K, he called an "elementary expansion", the inverse an elementary contraction (collapse). The equivalence generated by these moves is the "simple homotopy type" of K – in Whitehead's paper called "nucleus". He proved that it is a combinatorial invariant. Another fundamental concept developed in this paper was that of a "regular neighbourhood U" of a finite subcomplex K in a combinatorial n-manifold. U is a combinatorial n-manifold collapsing onto K with regard to a suitable subdivision. It exists uniquely up to equivalence. Whitehead proves a number of important theorems relating his nuclei to homotopy type in this and the subsequent paper [43], e.g., the homotopy classification of lens spaces. The complexes used in these papers are still of a piecewise linear structure, though Whitehead abandons the classical simplicial complex in favour of what he calls a "membrane complex". The n-simplices are replaced by n-elements in the sense of Newman, that is, by combinatorial n-cells E_i^n. The definition of a membrane complex K is given by induction on n, attaching E_i^n to the $(n-1)$-skeleton $K^{(n-1)}$ by a simplicial map $\partial E_i^n \to K^{(n-1)}$ not necessarily injective. The inductive procedure was already employed by Veblen [40] in the definition of a simplicial complex, and will reappear in the definition of CW-complexes.

Several other generalizations and variations of complexes are used by different authors. Lefschetz in [20] introduces a "topological simplicial complex", a homeomorphic image of a rectilinear one, or, in a simplicial complex simplices are grouped together to "blocks" or "cells" to form a "cell complex". Such a cell can be characterized in different ways, i.e. by imposing certain homological conditions [31].

It seems that the encumbrances which arose in the homotopy theory of complexes induced Whitehead to seek a more general and flexible notion of complex.

In [43] he developed the adequate concept, his closure-finite cell-complex with the weak topology. Such a complex K is built of topological cells by attaching inductively n-cells e_i^n to the $(n-1)$-skeleton $K^{(n-1)}$ of K: The continuous attaching map maps the boundary ∂e_i^n into a finite subcomplex of $K^{(n-1)}$ but leaves the interiors of the n-cells e_i^n disjoint from $K^{(n-1)}$ and from each other. This concept proved to be immensely useful. For example, homotopy equivalence can be characterized by the classical algebraic invariants, the homotopy groups π_i, for CW-complexes. If two homotopy equivalence classes $[X]$, $[Y]$ can be represented by CW-complexes X, Y, then $[X] = [Y] \Leftrightarrow \pi_i(X) \cong \pi_i(Y)$ for all $i = 0, 1, \ldots$, where the isomorphisms are supposed to be induced by a mapping $f : X \to Y$. Using CW-complexes it is rather easy to construct examples with given homotopy groups. Also the cellular homology which belongs in a natural way to a CW-complex provides a comfortable bridge between simplicial and singular homology. Dowker's metric complexes should be mentioned [11]. They consist of affine cells, and are not required to be countable or star-finite. The topology induced by the metric is coarser than Whitehead's.

6. PL-results and problems

Of course, simplicial complexes remained alive in spite of the introduction of Whitehead's CW-complexes, especially in low dimensions. Here, in dimensions $\leqslant 3$ the Hauptvermutung for closed manifolds was proved by Moise, see Chapter 16 of this book (Hauptvermutung). But also in higher dimensions piecewise linear topology (PL-topology) remained a category of interest in itself. An expository paper by W. Graeub (= Greub) of 1950 is meritorious in which PL-techniques are described from scratch and which contains a detailed proof of Alexander's theorem which states that any simplicially embedded 2-sphere in a 3-sphere separates the latter into two simplicial 3-balls.

One reason for the lasting interest in the concept of complex, simplicial or CW, especially in low dimensions, is the theory of groups. Every group can be realized as the fundamental group of a 2-dimensional complex which is the starting point of combinatorial group theory. Covering theory leads to the concept of the Cayley graph (Dehn's Gruppenbild), a 1-complex with additional structure (coloured).

These combinatorial methods have produced a wealth of group-theoretical results which seem to be out of reach by purely algebraic arguments. Buildings are examples in higher dimensions.

The theory of 2-complexes poses a couple of crucial (and unsolved) problems: The Andrews–Curtis conjecture [3] which states that a group presentation associated with a finite contractible CW-complex should be transformable into the trivial (empty) presentation by certain elementary operations (Q^{**}-transformations) on generators and relators.

Zeeman's conjecture claims [47] that a compact contractible polyhedron K^2 has the property, that the product $K^2 \times I$ with the unit interval I collapses to a point, $K^2 \times I \searrow *$, meaning that a finite sequence of Whitehead's elementary contractions will reduce $K^2 \times I$ to a point. Both conjectures are closely related to the 3-dimensional Poincaré conjecture – Zeeman's conjecture implies it. Finally, there is a conjecture by Whitehead himself [44]: Let K^2 be a connected 2-dimensional CW-complex and $\pi_2 K^2 = 0$ (K^2 is aspherical). Whitehead conjectures that every connected subcomplex $L \subset K$ is also aspherical, $\pi_2 L = 0$.

In the case of higher dimensions the PL-view can offer fundamental or computational simplifications, for instance in the theory of obstructions.

In the sixties Zeeman and Stallings obtained striking results in the field of PL-topology. Zeeman's "Seminar on combinatorial topology" [46] gave a thorough and up-to-date introduction to the theory. He used a modified definition for a polyhedron and its equivalence class: A set X is endowed with a "polystructure" if there is a PL-atlas whose charts are injective maps from finite Euclidean simplicial complexes into X with connecting functions which are PL-embeddings. The concept mirrors the usual procedure of putting a differential or geometric structure on a manifold. A polyhedral space then is a set X furnished with a polystructure. Thus a polyhedral category can be constructed in the obvious way, supplied with a natural notion of equivalence. The polyhedral category does not only contain polyhedra in the usual sense, but they are contained in it.

The seminar notes contain Zeeman's results on unknotting of pairs of spheres or balls for codimension $\geqslant 3$ in the PL-category. They retain their special importance since they differ from corresponding results in the topological or differentiable category. Among other things the topic of "engulfing" is introduced and treated which leads to a proof of the generalized Poincaré conjecture in the PL-case for dimensions $\geqslant 5$, [32, 45].

Bibliography

[1] J.W. Alexander, *The combinatorial theory of complexes*, Ann. Math. **31** (1930), 294–322.

[2] P. Alexandroff and H. Hopf, *Topologie I*, Springer, Berlin (1935).

[3] J.J. Andrews and M.L. Curtis, *Free groups and handlebodies*, Proc. Amer. Math. Soc. **16** (1965), 122–195.

[4] E. Betti, *Sopra gli spazi di un numero qualunque di dimensioni*, Ann. Mat. Pura Appl. (2) **4** (1871), 140–158.

[5] L.E.J. Brouwer, *Beweis der Invarianz der Dimensionenzahl*, Math. Ann. **70** (1911), 161–165.

[6] A. Cayley, *On the theory of groups*, Proc. London Math. Soc. **9** (1878), 126–133.

[7] M. Dehn and P. Heegaard, *Enzyklopädie Math. Wiss.* vol. III AB/3 (1907), 153–220.

[8] R. Descartes, *Oeuvres Inédites de Descartes*, par le comte Foucher de Careil, Bd. 2 (1859).

[9] J. Dieudonné, *Abrégé d'Histoire des Mathématiques 1700–1900*, Tom I et II, Hermann, Paris (1978), Section de Topologie de G. Hirsch. German translation: *Geschichte der Mathematik 1700–1900*, Vieweg, Braunschweig/Wiesbaden (1985).

[10] J. Dieudonné, *A History of Algebraic and Differential Topology 1900–1960*, Birkhäuser, Basel/Boston (1989).

[11] C.H. Dowker, *Topology of metric complexes*, Amer. J. Math. **74** (1952), 555–586.

[12] W. v. Dyck, *Gruppentheoretische Studien*, Math. Ann. **20** (1882), 1–44.

[13] W. v. Dyck, *Beiträge zur Analysis situs I*, Math. Ann. **32** (1888), 457–512.

[14] L. Euler, *Solutio problematis ad geometriam situs pertinentis*, Commentatio 53 indicis Enestroemiani Commentarii academiae scientiarum Petropolitanae **8** (1736), 128–140 and Opera Omnia I, Bd. 7 (1741), 1–10.

[15] L. Euler, *Brief an Goldbach*, Note für Akad. Wiss. Berlin (1750) and Opera Omnia I, Bd. 26, 71–93.

[16] L. Euler, *Elementa doctrinae solidorum*, Novi Comment. Acad. Sci. Petrop. **4** (1752), 109–140 and Opera omnia I, Bd. 26, 71–93; see also [15].

[17] P. Heegaard, *Forstudier til en topologisk Teori for de algebraiske Flader Sammenhäng*, Dissertation Copenhagen (French translation: *Sur l'Analysis situs*, Soc. Math. France Bull. **44**, 161–242 (1916)).

[18] C. Jordan, *Recherches sur les polyédres*, J. Reine Angew. Math. **66** (1866), 22–85.

[19] S. Lefschetz, *Topology*, Amer. Math. Soc. Colloq. Publ. No. 27, Providence, RI (1930).

[20] S. Lefschetz, *Introduction to Topology*, Princeton Univ. Press, Princeton (1949).

[21] S. L'Huilier, Ann. Math. Pures Appl. **3** (1812/1813), 169–192.

[22] J.B. Listing, *Der Census räumlicher Complexe (oder Verallgemeinerung des Eulerschen Satzes von den Polyedern)*, Abh. Wiss. Ges. Göttingen, **10** (1862), 97–180.

[23] A.F. Möbius, *From Nachlaß of A.F. Möbius*, Gesammelte Werke II, Hirzel, Leipzig (1886), 519–521.

[24] M.H.A. Newman, *On the foundations of Combinatorial analysis situs I, II*, Proc. of Akad. Wetenschap. Amsterdam **29** (1926), 611–641; **30** (1927), 670–673.

[25] H. Poincaré, Computes Rendus Acad. Sci. Paris (1882).

[26] H. Poincaré, *Analysis situs*, J. Ecole Polytech. (2) **1** (1895), 1–121.

[27] H. Poincaré, *1r Complément à l'Analysis situs*, Rend. Circ. Mat. Palermo **13** (1899), 285–343.

[28] H. Poincaré, *Oeuvres*, Vol. 6, 228.

[29] L. Schläfli, *Réduction d'une intégrale multiple*, J. Math. Pures Appl. **20** (1855), 359–394.

[30] H. Seifert, *Konstruktion dreidimensionaler geschlossener Räume*, Ber. Sächs. Akad. Wiss. **83** (1931), 26–66.

[31] H. Seifert and W. Threlfall, *Lehrbuch der Topologie*, Teubner, Leipzig (1934).

[32] J. Stallings, *Polyhedral homotopy-spheres*, Bull. Amer. Math. Soc. **66** (1960), 485–488.

[33] C. von Staudt, *Geometrie der Lage*, Nürnberg (1847).

[34] E. Steinitz, *Beiträge zur Analysis situs*, Sitz.-Ber. Berlin Math. Ges. **7** (1908), 29–49 and Arch. Math. Physik **13** (1908).

[35] E. Steinitz, *Vorlesungen über die Theorie des Polyeder*, Springer, Berlin (1934).
[36] J. Stillwell, *Classical Topology and Combinatorial Group Theory*, Grad. Texts Math. vol. 72, Springer, Berlin (1980).
[37] J. Stillwell, *Mathematics and Its History*, Springer, Berlin (1989).
[38] H. Tietze, *Über die topologischen Invarianten mehrdimensionaler Mannigfaltigkeiten*, Monatsh. Math. Phys. **19** (1908), 1–118.
[39] E.R. van Kampen, *On the connection between the fundamental groups of some related spaces*, Amer. J. Math. **55** (1933), 255–260.
[40] O. Veblen, *Analysis Situs*, Amer. Math. Soc., New York (1922).
[41] H. Weyl, *Die Idee der Riemannschen Fläche*, Teubner, Leipzig (1913).
[42] J.H.C. Whitehead, *Simplicial spaces, nuclei and m-groups*, Proc. Lond. Math. Soc. **45** (1939), 243–327.
[43] J.H.C. Whitehead, *On incidence matrices, nuclei and homotopy types*, Ann. Math. **42** (1941), 1197–1239.
[44] J.H.C. Whitehead, *On adding relations to homotopy groups*, Ann. Math. **42** (1941), 409–428.
[45] E.C. Zeeman, *The generalized Poincaré-conjecture*, Bull. Amer. Math. Soc. **76** (1961), 270.
[46] E.C. Zeeman, *Seminar on combinatorial topology*, Inst. Hautes Études Sci. (1963–1966).
[47] E.C. Zeeman, *On the dunce hat*, Topology **2** (1964), 341–358.

CHAPTER 5

Differential Forms

Victor J. Katz
University of the District of Columbia, USA

Differential forms are defined (loosely) as "things which occur under integral signs", that is, expressions of the form

$$\omega = \sum f_{\alpha_1\alpha_2\ldots\alpha_k}\, \mathrm{d}x_{\alpha_1}\, \mathrm{d}x_{\alpha_2} \ldots \mathrm{d}x_{\alpha_k},$$

where the summation is taken over all k-tuples $(\alpha_1, \alpha_2, \ldots, \alpha_k)$, with $1 \leqslant \alpha_1 < \alpha_2 < \cdots < \alpha_k \leqslant n$, and the f_α are differentiable function in some region of n-dimensional space. Such a form will be called a k-form in n-space. In this article, we will discuss the history of these things in relation to differential topology. In fact, it was the study of integration of differential forms in certain regions which was one of the roots of subject of differential topology.

Although the subject of differential forms can be traced back to eighteenth century work on differential equations, the connection with topology was first made by Augustin-Louis Cauchy (1789–1857) in two papers of 1846 dealing with 1-forms in n-space. These papers contained the bare statement of several theorems, without proofs. Cauchy promised to provide the proofs later, but apparently did not do so. The theorems deal with a function k of several variables $x, y, z, \ldots$ which is to be integrated along the boundary curve Γ of a surface S lying in a space of an unspecified number of dimensions. The most important results are collected in the following:

THEOREM. *Suppose*

$$k = X\frac{\mathrm{d}x}{\mathrm{d}s} + Y\frac{\mathrm{d}y}{\mathrm{d}s} + Z\frac{\mathrm{d}z}{\mathrm{d}s} + \cdots,$$

where $X\,\mathrm{d}x + Y\,\mathrm{d}y + Z\,\mathrm{d}z + \cdots$ *is an exact differential. (To say that this differential is exact is to say that* $\partial X/\partial y = \partial Y/\partial x$, $\partial X/\partial z = \partial Z/\partial x$, $\partial Y/\partial z = \partial Z/\partial y, \ldots$.*) Suppose that the function k is finite and continuous everywhere on S except at finitely many points*

HISTORY OF TOPOLOGY
Edited by I.M. James

P, P', P'', ... in its interior. If α, β, γ, ... are closed curves in S surrounding these points, respectively, then

$$\int_\Gamma k\,\mathrm{d}s = \int_\alpha k\,\mathrm{d}s + \int_\beta k\,\mathrm{d}s + \int_\gamma k\,\mathrm{d}s + \cdots.$$

In particular, if there are no such singular points, then

$$\int_\Gamma k\,\mathrm{d}s = 0.$$

In the 2-dimensional case, where S is a region of the plane and k is an arbitrary differential, then

$$\int_\Gamma k\,\mathrm{d}s = \pm \iint_S \left(\frac{\partial X}{\partial y} - \frac{\partial Y}{\partial x}\right) \mathrm{d}x\,\mathrm{d}y.$$

If k is an exact differential, then $\partial X/\partial y = \partial Y/\partial x$, so the right side, and therefore the left, vanish [9, 10].

Although we can derive the Cauchy integral theorem from the last statement, more interesting for our purposes is the appearance here both of the concept of a line integral in n-dimensional space and of the statement (in the next to the last sentence) of the theorem today generally known as Green's theorem. Finally, the expression of the line integral around the boundary of the surface as a sum of line integrals around isolated singular points, whose values are called *periods*, marked the beginning of the study of the relationships of integrals to surfaces over which they are not defined everywhere. Cauchy noticed in fact that if the path of integration goes m times around P, m' times around P', m'' times around P'', and so on, and if the periods for these points are designated by I, I', I'', and so on, then $\int k\,\mathrm{d}s$ can be written as $mI + m'I' + m''I'' + \cdots$. But because Cauchy never published the proof of this theorem, one can only speculate as to how far he carried all of these new concepts.

It was Bernhard Riemann (1826–1866) who restated Cauchy's results a few years later, with full proofs, and extended the result on periods far beyond Cauchy's conception. As part of this process, he introduced the basic ideas of what we now call the topology of a Riemann surface. In other words, instead of concentrating on the points of discontinuity of the coefficient functions of the differential form, he focused his attention on the connectedness of the domains over which they were defined. In 1851, he sketched this idea in his dissertation, but he explained it more fully in a paper of 1857.

Riemann began by observing that the integral of an exact differential $X\,\mathrm{d}x + Y\,\mathrm{d}y$ vanished when taken over the perimeter of a region of the (Riemann) surface R which covers the x-y plane. He then continued:

> Hence, the integral $\int(X\,\mathrm{d}x + Y\,\mathrm{d}y)$ has the same value when taken between two fixed points along two different paths, provided the two paths together form the entire boundary of a region of R. Thus, if every closed curve in the interior of R bounds a region of R, then the integral always has the same value when taken from a fixed initial point to one and the same endpoint, and is a continuous function of the position of the endpoint which is independent of the path of integration. This gives rise to a distinction among

surfaces: simply connected ones, in which every closed curve bounds a region of the surface – as, for example, a disk – and multiply connected ones, for which this does not happen – as, for example, an annulus bounded by two concentric circles [3, pp. 52, 53], [23].

Riemann proceeded to refine the notion of multiple connectedness: "A surface F is said to be $(n+1)$-ply connected when n closed curves $A_1, \ldots, A_n$ can be drawn on it which neither individually nor in combination bound a region of F, while if augmented by any other closed curve A_{n+1}, the set bounds some region of F" [23], [3, p. 53]. Riemann noted further that an $(n+1)$-ply connected surface can be changed into an n-ply connected one by means of a cut, a curve going from one boundary point through the interior to another boundary point. For example, an annulus, which is doubly connected, can be reduced to a simply connected region by any cut q which does not disconnect it. A double annulus needs two cuts to be reduced to a simply connected region.

Using the idea of cuts, Riemann was able to describe exactly what happened when one integrated an exact differential on an $(n+1)$-ply connected surface R. If one removes n cuts from this surface, there remains a simply connected surface R'. Integration of the exact differential $X\,\mathrm{d}x + Y\,\mathrm{d}y$ from a fixed starting point over any curve in R' then determines, as before, a single-valued continuous function Z of position on this surface. However, whenever the path of integration crosses a cut, the value jumps by a fixed number dependent on the cut. There are n such numbers, one for each cut. This notion of multiple connectedness turned out to be important in physics, particularly in fluid dynamics and electromagnetism, and so it was extended to regions of 3-dimensional space by such physicists as Hermann von Helmholtz (1821–1894), William Thomson (1824–1907), and Clerk Maxwell (1831–1879).

Helmholtz extended Riemann's definition to 3-dimensional regions in a paper of 1858: "An n-ply connected space (in 3-dimensional space) is one which can be cut through by $n-1$, but no more, surfaces without being separated into two detached portions" [15, p. 27]. Thus Helmholtz's surfaces replaced Riemann's cuts. Helmholtz noted that certain important theorems in potential theory failed to hold in a multiply connected region precisely because integrals of exact differentials could not then be considered as single-valued functions.

The British physicists picked up on Helmholtz's ideas shortly after his paper was translated into English. Thomson in 1869 explained what happened when line integrals were taken in an n-ply connected 3-dimensional space. He illustrated his discussion with pictures of pretzel-like regions and interconnected rings. Using Helmholtz's definition of n-ply connected spaces, he defined numbers analogous to Cauchy's periods. Namely, if $F\,\mathrm{d}s = u\,\mathrm{d}x + v\,\mathrm{d}y + w\,\mathrm{d}z$ is an exact differential and β_i is one of Helmholtz's barrier surfaces, then, if points P and Q are "each infinitely near a point B of β_i, but on the two sides of this surface," he defined κ_i to be $\int F\,\mathrm{d}s$ taken along any curve in the space joining P and Q without cutting any other barrier β. Thomson showed that this value is the same for any such curve and for any point B on β_i [24, pp. 43, 44].

A few years later, Maxwell further generalized this idea [18]. Namely, if a closed curve Γ passes m_i times through β_i, then the corresponding line integral $\int_\Gamma F\,\mathrm{d}s$ will be given by

$$\int_\Gamma F\,\mathrm{d}s = m_1\kappa_1 + m_2\kappa_2 + \cdots + m_n\kappa_n.$$

Around the same time, Enrico Betti (1823–1892) generalized this notion even further, by defining connectivity in n-dimensional spaces.

Betti's definition was as follows: For each dimension $m < n$, a region R is said to have m-dimensional order of connectivity $p_m + 1$, if there are p_m closed m-dimensional spaces $A_1, A_2, \ldots, A_{p_m}$ in R, which together do not form the boundary of a connected $(m+1)$-dimensional region of R, while any additional closed m-dimensional space together with some subset of the $\{A_j\}$ forms such a boundary [2]. (A closed space, for Betti, was one without a boundary.) So, for example, in a region whose m-dimensional order of connectivity is 1, any closed m-dimensional space is the boundary of an $(m+1)$-dimensional space. For $n = 2$ and $m = 1$, Betti's definition is the same as Riemann's original definition. We should note, of course, that Betti's definition requires a theorem to show that it is consistent. Betti provided one, but his proof was deficient. The situation was not entirely clarified until the work of Poincaré.

Generalizing the work of Riemann, Betti showed that to make a space simply connected (1-ply connected) in the m-th dimension, one had to remove from it p_m $(n-m)$-dimensional cross-sections. For example, if $m = 1$, one must remove p_1 $(n-1)$-dimensional cross-sections from R to make the remainder R' simply connected in the first dimension.

Betti then went on to compare n-fold integrals with $(n-1)$-fold integrals. For example, he considered an n-dimensional region R bounded by t closed $(n-1)$-dimensional spaces $S_1, S_2, \ldots, S_t$, given respectively by equations $F_1 = 0$, $F_2 = 0, \ldots, F_t = 0$. We will just consider the case $t = 1$, where the hypersurface S is given by the equation $F = 0$. If $f_1, f_2, \ldots, f_n$ are functions on R, Betti's aim was to express the n-fold integral

$$\Omega_n = \int_R \left(\frac{\partial f_1}{\partial z_1} + \frac{\partial f_2}{\partial z_2} + \cdots + \frac{\partial f_n}{\partial z_n} \right) \mathrm{d}z_1 \, \mathrm{d}z_2 \ldots \mathrm{d}z_n$$

in terms of an $(n-1)$-fold integral. To do this, he parametrized the hypersurface S through n parameters $z_i = z_i(u_1, u_2, \ldots, u_{n-1})$ $(i = 1, 2, \ldots, n)$ and then, via a proof generalizing the fundamental theorem of calculus, showed that

$$\Omega_n = \int \sum_{i=1}^{n} (-1)^i f_i \frac{\partial(z_1, \ldots, \hat{z}_i, \ldots, z_n)}{\partial(u_1, u_2, \ldots, u_{n-1})} \mathrm{d}u_1 \mathrm{d}u_2 \ldots \mathrm{d}u_{n-1},$$

where the integral is taken over the appropriate parametric region in $(n-1)$-space. Ω_n can be expressed more simply as

$$\Omega_n = \int_S \sum_{i=1}^{n} (-1)^i f_i \, \mathrm{d}z_1 \ldots \mathrm{d}\hat{z}_i \ldots \mathrm{d}z_n.$$

This result of Betti's generalizes the classical divergence theorem. He also noted that if the space R is simply connected in the $(n-1)$-st dimension, then any two $(n-1)$-dimensional regions having the same boundary Γ together form a closed space. Thus, we note that if $\sum \partial f_i / \partial z_i = 0$, then the integral $\int \sum (-1)^i f_i \, \mathrm{d}z_1 \ldots \mathrm{d}\hat{z}_i \ldots \mathrm{d}z_n$ will always have the same value over any spaces with boundary Γ.

Betti also generalized the classical Stokes' theorem, by looking at one-forms $\sum X_i \, dz_i$ in n-dimensional space. Assuming that the curve γ represented by $z_i = z_i(u)$ bounds the region C given by $z_i = z_i(v_1, v_2)$, he defined the integral

$$\Omega_1 = \int_\gamma \sum_{i=1}^{n} X_i \, dz_i = \int \sum_{i=1}^{n} X_i \frac{dz_i}{du} \, du$$

and calculated that this expression is equal to

$$\iint_C \sum_{i,j=1; i>j}^{n} \left(\frac{\partial X_i}{\partial z_j} - \frac{\partial X_j}{\partial z_i} \right) dz_i \, dz_j .$$

Betti then assumed that the region R had connectivity $p + 1$ in the first dimension. This means that there are p $(n-1)$-dimensional cross-sections $s_1, s_2, \ldots, s_p$, such that on removing these sections from R, the remainder R' will be simply connected. Furthermore, there are p closed curves $L_1, L_2, \ldots, L_p$, which, respectively, meet the sections s_i, and such that any other closed curve γ forms with the L's the boundary of a 2-dimensional space C. Betti could then conclude that if $\partial X_i/\partial z_j - \partial X_j/\partial z_i = 0$ for all i, j, then $\int \sum X_i \, dz_i = 0$, where the integral is taken over the entire boundary system of C, namely, $\gamma, L_1, L_2, \ldots, L_p$. It follows that if $M_t = \int_{L_t} \sum X_i \, dz_i$, then "the integral $\int \sum X_i \, dz_i$, taken from z_0 to z_1 along any curve which meets certain sections s_j, differs from that taken along any curve from z_0 to z_1 which does not meet any of the sections s_j by the quantities M_j relative to the intersected sections s_j; these quantities are taken to be positive or negative depending on whether they [the curves] intersect the section [while] progressing in one or the other direction" [2, p. 158]. Hence, if R is simply connected in the first dimension, "the integral taken along any curve in R from z_0 to z_1 always has the same value." It is easy to see that this result is the same as Maxwell's result, except, of course, that it is valid for an arbitrary number of dimensions.

The conditions above insuring that the line integral $\int \sum X_i \, dz_i$ is independent of the path of integration and depends only on the endpoints in a simply connected space were called the *integrability conditions* by Henri Poincaré (1854–1912) in 1887. He used this name because these conditions imply the existence of an "integral" for $\sum X_i \, dz$, that is, a function f such that $df = \sum X_i \, dz_i$. Poincaré went on to consider similar conditions for surface integrals in n-dimensional space, motivated by the aim of generalizing the work of Cauchy on functions of one complex variable to functions of two complex variables [20].

Poincaré succeeded in this aim, although in the process he had to consider surface integrals of real functions in n-space, namely integrals of the form

$$J = \iint \sum (X_i, X_k) \, dx_i \, dx_k ,$$

where each symbol (X_i, X_k) denotes a function of the n variables x_i, where $(X_k, X_k) = 0$ and $(X_i, X_k) = -(X_k, X_i)$ for all values of i and k; and where the summation is taken over all pairs of indices. Poincaré defined this integral by parametrizing the surface, thereby converting J into an ordinary double integral in the plane. He was careful to remark that the order of integration of the parametric variables was crucial; this was his reason for insisting on the skew symmetry of the functions (X_i, X_k).

Poincaré went on to derive the integrability conditions he was seeking, namely, the conditions under which the integral does not depend on the surface of integration, but only on the boundary curve. These conditions turned out to be the conditions

$$\frac{\partial(X_i, X_k)}{\partial x_h} + \frac{\partial(X_k, X_h)}{\partial x_i} + \frac{\partial(X_h, X_i)}{\partial x_k} = 0.$$

In essence, the case $n = 3$ of this result had been known to Betti, but Poincaré's proof was different from that of Betti. In addition, the case $n = 4$ gave Poincaré the result he wanted for functions of two complex variables.

Having obtained the result for 2-dimensional integrals, Poincaré then generalized it to integrals of higher order. Thus, for a triple integral of the form

$$\iiint \sum (X_\alpha, X_\beta, X_\gamma)\, \mathrm{d}x_\alpha\, \mathrm{d}x_\beta\, \mathrm{d}x_\gamma,$$

where the symbols $(X_\alpha, X_\beta, X_\gamma)$ satisfy analogous properties to those for functions of two variables, the conditions under which the integral only depends on the 2-dimensional boundary of the 3-dimensional space over which the integral is taken are

$$\frac{\partial(X_\alpha, X_\beta, X_\gamma)}{\partial x_\delta} - \frac{\partial(X_\beta, X_\gamma, X_\delta)}{\partial x_\alpha} + \frac{\partial(X_\gamma, X_\delta, X_\alpha)}{\partial x_\beta} + \frac{\partial(X_\delta, X_\alpha, X_\beta)}{\partial x_\gamma} = 0.$$

Poincaré noted that similar results would hold in any dimension, with the signs between the individual terms alternating in the odd-dimensional cases and always being positive in the even-dimensional cases. Vito Volterra (1860–1940), in fact, wrote out and proved these conditions in full in 1889, although in terms of solutions of partial differential equations rather than in terms of integrals depending on boundaries [26].

Now both Poincaré and Volterra in these results had assumed connectivity of the domain only in the sense that the functions involved could not have any singularities in the regions of integration. But in 1895, Poincaré discussed what happened to integrals over regions of multiple connectivity. In his fundamental paper "Analysis Situs", Poincaré defined the notions of homology and Betti number, further clarifying these four years later: A *homology relation* exists among p-dimensional subvarieties $v_1, v_2, \ldots, v_r$ of an n-dimensional variety V, written

$$v_1 + v_2 + \cdots + v_r \sim 0,$$

if for some integer k, the set consisting of k copies of each of the v_i constitutes the complete boundary of a $(p+1)$-dimensional subvariety W [21, p. 203], [22, p. 291]. (A "variety" for Poincaré (now generally called a manifold in English) was the generalization to higher dimensions of a curve in one dimension or a surface in two dimensions and was generally thought of as being defined, at least locally, either as the set of zeroes of an appropriate system of functions or parametrically as the image of a certain set of such functions.) Poincaré introduced "negatives" of varieties by considering orientation. That is, $-v$ is the same variety as v but with the opposite orientation. As an example of a homology relation, let v_1 and v_2 be the outer and inner boundaries respectively of a ring, with opposite orientations. Then v_1 and v_2 together form the complete boundary of the ring, and, since $-v_2$ has the opposite orientation as v_2, the relation $v_1 - v_2 \sim 0$ holds. Poincaré further observed that the varieties in homology relations can be added, subtracted, and multiplied by integers and therefore was able to call a set of varieties *linearly independent* if there were no homology relation among them with integer coefficients.

To clarify the notion of multiple-connectedness, Poincaré went on to define the p-dimensional *Betti number* B_p of a variety V to be one more than the maximum number of linearly independent, closed, p-dimensional subvarieties, where a closed variety is one without a boundary. Thus according to Poincaré the one-dimensional Betti number of the ring is 2, while that of the double ring is 3. On the other hand, the one-dimensional Betti number of the disc is 1. (Poincaré's definition is nearly the same as that of Betti; the difference is that Betti had failed to consider the possibility that a multiple of a variety was a boundary, while the variety itself was not. On the other hand, the definitions of both were modified in the 1920s, when the p-dimensional Betti number was defined to be exactly the number of independent, closed p-dimensional subvarieties. Today, however, the Betti number is defined to be the rank or dimension of a particular homology or cohomology group.)

Having defined the Betti numbers, Poincaré was ready to consider integrals of the form

$$\int \sum Z_{\alpha_1 \dots \alpha_p} \, \mathrm{d}x_{\alpha_1} \dots \mathrm{d}x_{\alpha_p},$$

where the integrals are now taken over multiply-connected varieties. He then generalized his earlier result on the integrability conditions: If the complete boundary of an $(m+1)$-dimensional variety W is composed of k m-dimensional varieties $V_1, V_2, \dots, V_k$, then, assuming the integrability conditions are satisfied, the algebraic sum of the integrals over the V_i will also be zero.

Hence, since there are $B_m - 1$ independent closed m-dimensional varieties $V_1, V_2, \dots,$ $V_{B_m - 1}$ such that any closed variety U is (up to homology) a linear combination of these and since, therefore, a multiple of U together with the same multiple of this linear combination forms the boundary of an $(m+1)$-dimensional variety, Poincaré could conclude that the integral taken over U is simply a linear combination of the values that the integral takes over the V_i. Poincaré called these values, which are the generalizations of similar values appearing in the works of other authors, the *periods* of the integral. In fact, he noted that Betti had essentially come to the same conclusion, but only for dimensions 1 and $n-1$. Poincaré knew, naturally, that the maximum number of linearly independent periods is equal to $B_m - 1$, but he also claimed, without proof, that there always exist integrals for which the maximum number of periods is attained [21, p. 95]. A full explanation and proof of this remark took another thirty years.

Even though Poincaré had extensively used differential forms and showed how to integrate them, they still lacked even a formal definition. This was provided in 1899 by Elie Cartan (1869–1951) [5]. His definition was a "purely symbolic" one; namely, he defined "differential expressions" as homogeneous expressions formed by a finite number of additions and multiplications of the differentials $\mathrm{d}x$, $\mathrm{d}y$, $\mathrm{d}z$, ... and certain differentiable coefficient functions. Thus, for instance, $A\,\mathrm{d}x + B\,\mathrm{d}y$ was a first degree expression and $A\,\mathrm{d}x\,\mathrm{d}y + B\,\mathrm{d}z\,\mathrm{d}y$ was a second degree expression. These forms were to have as their algebra the exterior algebra of Hermann Grassmann (1809–1877). His work, though originally neglected when it appeared in 1844, had by the late 19-th century been discussed by many mathematicians. In fact, Cartan in 1908 wrote an encyclopedia article explaining Grassmann's work [7].

The major idea for Cartan was the exterior multiplication starting with an n-dimensional real vector space V; namely, there are to be n independent "units" of the first class ε_1, ε_2, $\ldots$, ε_n in V with an associative multiplication characterized by the relations $\varepsilon_j\varepsilon_j = 0$ and $\varepsilon_i\varepsilon_j = -\varepsilon_j\varepsilon_i$. Of course these elements $\varepsilon_i\varepsilon_j$ are no longer in V but in a new vector space of dimension $n(n-1)/2$ in which the independent units (of the second class) are the elements $\varepsilon_{ij} = \varepsilon_i\varepsilon_j$, where $1 \leqslant i < j \leqslant n$. One can similarly define units of the k-th class for any k with $0 \leqslant k \leqslant n$; there are then $n!/k!(n-k)!$ such units.

For Cartan, the units of the first class were the differentials $\mathrm{d}x_i$, $i = 1, 2, \ldots, n$, while the units of higher classes were simply formal products of these differentials. Interestingly, Cartan had already used these Grassmannian rules with very little comment in a paper of 1896 [4], in which he even showed that formal multiplication of the differential forms $\mathrm{d}u = (\partial u/\partial x)\,\mathrm{d}x + (\partial u/\partial y)\,\mathrm{d}y$ and $\mathrm{d}v = (\partial v/\partial x)\,\mathrm{d}x + (\partial v/\partial y)\,\mathrm{d}y$ gave the change-of-variable formula for double integrals:

$$\mathrm{d}u\,\mathrm{d}v = \frac{\partial(u, v)}{\partial(x, y)}\,\mathrm{d}x\,\mathrm{d}y.$$

Besides developing the algebra of differential forms, Cartan also developed their calculus. Namely, in [5] he defined the derived expression (now called the *exterior derivative*) of a one-form $\omega = \sum A_i\,\mathrm{d}x_i$ to be the two-form $\mathrm{d}\omega = \sum \mathrm{d}A_i\,\mathrm{d}x_i$. For example, the derived expression of the form $\omega = A\,\mathrm{d}x + B\,\mathrm{d}y$ is the form

$$\mathrm{d}\omega = \left(\frac{\partial A}{\partial x}\,\mathrm{d}x + \frac{\partial A}{\partial y}\,\mathrm{d}y\right)\mathrm{d}x + \left(\frac{\partial B}{\partial x}\,\mathrm{d}x + \frac{\partial B}{\partial y}\,\mathrm{d}y\right)\mathrm{d}y = \left(\frac{\partial B}{\partial x} - \frac{\partial A}{\partial y}\right)\mathrm{d}x\,\mathrm{d}y.$$

Note that this derived expression appears in the statement of Green's theorem, while the exterior derivative of the one-form $A\,\mathrm{d}x + B\,\mathrm{d}y + C\,\mathrm{d}z$ appears in the statement of Stokes' theorem. In 1901, Cartan generalized his definition of the exterior derivative to forms of any degree [6]. Namely, if $\omega = \sum a_{ij\ldots k}\,\mathrm{d}x_i\,\mathrm{d}x_j \ldots \mathrm{d}x_k$, then the exterior derivative $\mathrm{d}\omega$ is defined to be $\sum \mathrm{d}a_{ij\ldots k}\,\mathrm{d}x_i\,\mathrm{d}x_j \ldots \mathrm{d}x_k$. It is straightforward to show then that the exterior derivative of the two-form $A\,\mathrm{d}y\,\mathrm{d}z + B\,\mathrm{d}z\,\mathrm{d}x + C\,\mathrm{d}x\,\mathrm{d}y$ is the 3-form

$$\left(\frac{\partial A}{\partial x} + \frac{\partial B}{\partial y} + \frac{\partial C}{\partial z}\right)\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z,$$

the expression which shows up in the divergence theorem.

Although Cartan realized that these three theorems of vector calculus could be easily stated using differential forms, it was Edouard Goursat (1858–1936) in 1917 who first noted that these theorems were all special cases of a generalized Stokes' theorem, which could be written in the simple form

$$\int_S \omega = \int_T \mathrm{d}\omega,$$

where ω is a p-form in n-space, and S is the p-dimensional boundary of the $(p+1)$-dimensional region T [14]. (This general theorem had already been stated, in coordinate form, by Volterra in 1889 in [25].) Goursat also used differential forms to state and prove the Poincaré lemma and its converse, namely, that if ω is a p-form, then $d\omega = 0$ if and only if there is a $(p-1)$-form η with $\omega = d\eta$. Goursat did not notice, however, that the "only if" part of the result depends on the domain of ω and is not true in general.

Cartan in 1922 gave a counterexample [8]. Namely, he noted that the two-form

$$\omega = \frac{x\,dy\,dz + y\,dz\,dx + z\,dx\,dy}{(x^2+y^2+z^2)^{3/2}},$$

defined in $D = R^3 - 0$ had the property that $d\omega = 0$, but that there could not be a one-form η defined in D with $d\eta = \omega$. For if such an η existed, then by Stokes' theorem, with Σ being the unit sphere and $\partial\Sigma$ its boundary, $\int_\Sigma \omega = \int_\Sigma d\eta = \int_{\partial\Sigma} \eta = 0$, since the sphere has no boundary. On the other hand, a straightforward calculation shows that $\int_\Sigma \omega = 4\pi$. Thus, the Poincaré lemma implies that R^3 is cohomologically trivial, while Cartan's example shows that with manifolds whose higher order Betti numbers are positive, one needs to refine this result considerably. This work was accomplished by Cartan and Georges DeRham in the following decade.

To understand the work of DeRham, we need a quick review of the progress in algebraic topology after Poincaré's original papers of the late 1890s. The basic change was that Poincaré's varieties were replaced by simplexes and complexes. That is, the p-dimensional submanifolds were considered not as solutions to systems of equations but as being formed from certain simple p-dimensional manifolds, each of which was the continuous image of a p-dimensional "triangle". The appropriate definitions were completely worked out by James W. Alexander (1888–1971) by 1926 when he defined a *p-simplex* to be the p-dimensional analogue of a triangle and a *complex* to be a finite set of simplexes such that no two had an interior point in common and such that every face of a simplex of the set was also a simplex of the set [1]. An elementary i-chain of a complex was defined to be an expression of the form $\pm V_0V_1 \ldots V_i$, where the Vs are vertices of an i-simplex. The expression changes sign upon any transposition of the Vs, thus giving each chain an orientation. An elementary i-chain was then an i-dimensional "face" of a p-simplex, while an arbitrary i-chain was a linear combination of elementary i-chains with integer coefficients. As an example, the tetrahedron with vertices V_0, V_1, V_2, V_3 is a 3-simplex while it together with its four faces (each a 2-simplex), its four edges (each a 1-simplex), and its four vertices (each a 0-simplex), form a complex. The face $V_0V_1V_2$ is then an elementary 2-chain of the 3-simplex. Alexander next defined the boundary of the elementary i-chain $K = V_0V_i \ldots V_i$ to be the $(i-1)$-chain $K' = \sum(-1)^s V_0 \ldots \widehat{V_s} \ldots V_i$ and extended this to arbitrary i-chains by linearity. Thus the boundary of $V_0V_1V_2$ is $V_0V_1 - V_0V_2 + V_0V_1$. An easy calculation with this example shows that the boundary of the boundary is zero, and one can show that this result is true in general.

Alexander gave his definition of homology applied to *closed chains* (cycles), chains whose boundary is zero. Namely, a closed chain K is homologous to zero, $K \sim 0$, if it is the boundary of a chain L. Two chains K and K^* are homologous, $K \sim K^*$, if $K - K^*$ is homologous to zero. The p-th Betti number of a complex is then the maximum number of closed p-chains which are linearly independent with respect to boundary, that is, such that no linear combination is homologous to zero. (Note that this number is one less than the number according to Poincaré's original definition.)

With a commutative operation ("addition") having an inverse being considered on the set of closed chains, it should be clear to modern readers that there is a group hiding among Alexander's definitions. It was Emmy Noether (1882–1935) who suggested to the Göttingen mathematicians in the late 1920's that they apply group-theoretic ideas to combinatorial topology. As Pavel Aleksandrov (1896–1982) put it in his address in her memory,

> When in the course of our lectures she first became acquainted with a systematic construction of combinatorial topology, she immediately observed that it would be worthwhile to study directly the groups of algebraic complexes and cycles of a given polyhedron and the subgroup of the cycle group consisting of cycles homologous to zero; instead of the usual definition of Betti numbers and torsion coefficients, she suggested immediately defining the Betti group as the (quotient) group of the group of all cycles by the subgroup of cycles homologous to zero [13, p. 130].

With Noether's remarks and the subsequent publications of Leopold Vietoris (1891–) and Heinz Hopf (1894–1971), the subject of algebraic topology began in earnest. Vietoris in 1927 defined the *homology group* $H(A)$ of a complex A to be the quotient group of cycles modulo boundaries, as Noether recommended. About the same time, Hopf defined several other Abelian groups, namely, the groups L^p, Z^p, R^p, and $\overline{R}^p$ generated by the p-simplexes, the p-cycles, the p-boundaries (those chains which were the boundary of some chain), and the p-boundary-divisors (those chains for which a multiple was a boundary), respectively. Then for Hopf, the factor group $B_p = Z^p/\overline{R}^p$ was a free group (a group none of whose elements had a multiple equal to 0) whose rank (the number of basis elements) turned out to be the p-th Betti number of the complex [16].

Matters progressed so quickly in this new field that just a year later Walther Mayer (1887–1948) published an axiom system for defining homology groups [19]. Namely, Mayer was no longer concerned with the topological complexes themselves, but solely with the algebraic operations which were defined on them. Thus a complex ring Σ was a collection of elements (complexes) $K^{(p)}$, to each of which was attached a dimension p. The p-dimensional elements formed a finitely generated free Abelian group K^p. For each p, a homomorphism $R_p : K^p \to K^{p-1}$ is defined such that $R_{p-1}(R_p(K^p)) = 0$. (R_p is called the p-th boundary operator. Often, one just uses R, without subscripts, and then writes the last equation in the form $R^2 = 0$.) Given these axioms, Mayer defined the group of p-cycles C^p to be those elements K of K^p for which $R(K) = 0$ and the group of p-boundaries to be $R(K^{p+1})$. Modifying Hopf's definition slightly, he defined the p-th homology group of Σ to be the factor group $H_p(\Sigma) = C^p/R(K^{p+1})$.

Using these definitions, DeRham in 1931 was able to complete the generalization of Poincaré's lemma in the way which Poincaré had indicated earlier. In his doctoral dissertation of that year, published in [11], DeRham first defined the q-chains, namely, the objects over which differential q-forms are to be integrated in a variety V. An elementary q-chain c^q is the differentiable homeomorphic image in V of a convex q-dimensional polyhedron T (a q-simplex) while an arbitrary q-chain is a linear combination of elementary ones with integral coefficients. The boundary of an elementary q-chain is the chain formed by restricting c^q to all of the faces of T in turn with appropriate signs. Then if ω is a q-form and c is a q-chain, $\int_c \omega$ is defined by "pulling back" everything to the appropriate polyhedra and performing an ordinary integration there.

Next DeRham made explicit the analogies between forms and chains by introducing definitions for the forms to match the corresponding definitions for chains: A form ω is closed if its exterior derivative $d\omega$ is 0 while a chain c is closed if its boundary is 0. A q-form ω is homologous to zero if there is a $(q-1)$-form η with $d\eta = \omega$, while a q-chain is homologous to zero if there is a $(q+1)$-chain C whose boundary is c. Both the derivative and the boundary operators give zero when repeated. By the generalized Stokes' theorem, therefore, the integral of a form homologous to zero over a closed chain is zero, as is the integral of a closed form over a chain homologous to zero. Finally, DeRham assumed that V has a "polyhedral subdivision", that is, there is a finite system of elementary chains $\{a_i^q\}$ which "form" V. In other words, V has a differentiable triangulation. Under this hypothesis, he demonstrated that for each q there are a finite number p_q of "elementary" closed q-forms such that every closed q-form is homologous to a linear combination of these elementary ones and also a finite number p'_q of elementary closed q-chains with the same property. It is perhaps surprising that DeRham did not mention one further analogy. He did define the homology groups for a chain complex as Mayer had. But although the forms satisfy essentially the same axioms as the chains, he made no mention of defining similar groups for them.

In a private communication, DeRham said that he saw no reason to introduce these "cohomology" groups at the time since they are nothing but the homology groups of the reciprocal complex. In particular, in an n-dimensional manifold, a q-chain and an $(n-q)$-form are simply two aspects of the same more general notion which DeRham later developed extensively under the name "current".

In any case, after defining the period of a closed form ω over a closed chain c to be $\int_c \omega$, DeRham completed the analogies with three important theorems. First, a closed form of which all the periods are zero is homologous to zero. Second, given p closed q-chains among which there is no homology (therefore $p \leqslant p'_q$), there exists a closed q-form which, integrated over these q-chains, gives p given values. (This is the result stated earlier by Poincaré.) Third, if B_q is the q-th Betti number of the complex of the $\{a_i^q\}$, then $B_q = p_q = p'_q$. In current terminology, these theorems say that for each q, the singular homology group $H_q(V, R)$ and the differential cohomology group $H^q(V)$ are dual real vector spaces via the operation $\int_c \omega$.

Within the next few years, cohomology groups were in fact introduced. First, Eduard Kähler in 1932–1934 defined the ring of differential forms essentially as a module over the ring of functions in n-space generated by "symbols" dx, $d(x_i, x_k), \ldots, d(x_{i_1}, x_{i_2}, \ldots, x_{i_p})$ with appropriate operations [17]. And DeRham himself, in lectures delivered in Hamburg in 1938, called this type of ring an "alternating ring", noting of course that the "integrands in n-dimensional space, otherwise called differential forms", are a prime example [12]. He further defined the exterior derivative, then noted that the subring of "total forms" – those which are homologous to zero – is an ideal in the subring of closed forms. Hence, one can consider the residue class ring, that is the "cohomology ring" as a graded ring, each dimension of which is a "cohomology group".

Over the next fifteen years, the notion of differential forms was involved in the development of several new concepts, including the ideas of fibre bundles, sheaves, and differential graded rings. It was out of these concepts, in fact, that a set-theoretic definition of differential forms could be given. Namely, a differential form was defined to be a section of a vector bundle over a manifold, where at each point the vector space was ΛT^*, the exterior algebra of the dual space of the tangent space to the manifold at that point.

But even with this very abstract definition, it is as "things under integral signs" that most mathematicians think of differential forms. And they continue to prove useful in numerous areas.

Bibliography

[1] J.W. Alexander, *Combinatorial Analysis Situs*, Trans. Amer. Math. Soc. **28** (1926), 301–329.

[2] E. Betti, *Sopra gli spazi di un numero qualunque di dimensioni*, Annali di Mathematica (2) **4** (1871), 140–158; *Opere*, Vol. 2, 273–290.

[3] G. Birkhoff (ed.), *A Source Book in Classical Analysis*, Harvard Univ. Press, Cambridge (1973).

[4] E. Cartan, *Le principle de dualité de certaines intégrales multiples de l'espace tangentiel de l'espace reglé*, Bull. Soc. Math. France **24** (1896), 140–177; *Oeuvres* (2), Vol. 1, 265–302.

[5] E. Cartan, *Sur certaines expressions différentielles et sur le problème de Pfaff*, Ann. École Normale **16** (1899), 239–332; *Oeuvres* (2), Vol. 1, 303–397.

[6] E. Cartan, *Sur l'intégration de certaines systèmes de Pfaff de caractère deux*, Bull. Soc. Math. France **29** (1901), 233–302; *Oeuvres* (2), Vol. 1, 483–554.

[7] E. Cartan, *Nombres complexes*, Encyclopédie Sci. Math. (edition Française) **15** (1908); *Oeuvres* (2), Vol. 1, 107–246.

[8] E. Cartan, *Leçons sur les Invariants Intégraux*, Hermann, Paris (1922).

[9] A.-L. Cauchy, *Sur les intégrales qui s'étendent a tous les points d'une courbe fermée*, Comptes Rendus **23** (1846), 251–255; *Oeuvres* (1), Vol. 10, 70–74.

[10] A.-L. Cauchy, *Considérations nouvelles sur les intégrales définies qui s'étendent a tous les points d'une courbe fermée, et sur celles qui sont prises entre des limites imaginaires*, Comptes Rendus **23** (1846), 689–704; *Oeuvres* (1), Vol. 10, 153–168.

[11] G. DeRham, *Sur l'Analysis situs des variétés a n-dimension*, J. Math. Pure Appl. (9) **10** (1931), 115–200.

[12] G. DeRham, *Über mehrfache Integrale*, Abhandlungen aus dem Mathematischen Seminar des Hansischen Universität (Universität Hamburg) **12** (1938), 313–339.

[13] A. Dick, *Emmy Noether, 1882–1835*, Birkhäuser, Boston (1981).

[14] E. Goursat, *Sur certains systèmes d'équations aux différentielles totales et sur une généralisation du problème de Pfaff*, Ann Fac. Sci. Toulouse (3) **7** (1917), 1–58.

[15] H. von Helmholtz, *Über Integrale des hydrodynamischen Gleichungen welche den Wirbelbewegungen entsprechen*, Crelle **55** (1858), 25–55; English translation: Philos. Mag. (4) **33** (1868), 485–511.

[16] H. Hopf, *Eine Verallgemeinerung der Euler–Poincaré'schen Formel*, Nachr. Gesellschaft Wiss. Göttingen Math. Phys. Kl. (1928), 127–136. See also *Selecta Heinz Hopf*, Springer, Berlin (1964), 5–13.

[17] E. Kähler, *Forme differenziali e funzioni algebriche*, Memorie della Reale Accademie D'Italia, Classe di Scienze Fisiche Mathematiche e Naturali **3** (1932).

[18] C. Maxwell, *A Treatise on Electricity and Magnetism*, Oxford Univ. Press, London (1873).

[19] W. Mayer, *Über Abstrakte Topologie I*, Monatshefte für Math. Phys. **36** (1929), 1–42.

[20] H. Poincaré, *Sur les résidus des intégrales doubles*, Acta Math. **9** (1887), 321–380; *Oeuvres*, Vol. 3, 440–489.

[21] H. Poincaré, Analysis situs, J. Ecole Polytech. (2) **1** (1895), 1–121; *Oeuvres*, Vol. 6, 193–288.

[22] H. Poincaré, *Complement à l'Analysis situs*, Rendiconti del Circolo Matematico di Palermo **13** (1899), 285–343; *Oeuvres*, Vol. 6, 290–337.

[23] B. Riemann, *Lehrsätze aus des analysis situs für die Theorie der Integrale von zweigliedrigen vollständigen Differentialen*, Crelle **54** (1857), 105–110; *Werke*, 91–96. For English translation see [3].

[24] W. Thomson, *On vortex motion*, Trans. Roy. Soc. Edinburgh **25** (1869), 217–260; Mathematics and Physics Papers, Vol. 4, 13–66.

[25] V. Volterra, *Della variabili complesse negli iperspazi*, Rendiconti Accademia dei Lincei (4) **5**, 158–165; *Opere* **1**, 403–410.

[26] V. Volterra, *Sulle funzioni coniugate*, Rendiconti Accademia dei Lincei (4) **5** (1889), 599–611; *Opere*, Vol. 1, 420–432.

CHAPTER 6

The Topological Work of Henri Poincaré

K.S. Sarkaria

Department of Mathematics, Panjab University, Chandigarh 160014, India

Introduction

Topology, as we know it today, started with Poincaré's "*Analysis Situs*" [59] and its five *Compléments* [61, 62, 66, 67, 69]. My objective is to describe the contents of these classics, together with some remarks relating them to further developments.

So we will not go into the details of Poincaré's life (1854–1912) and the various honours bestowed on him. However, a mathematician will enjoy Appell's anecdotes [3] about their year together in high school in Nancy, e.g., the way in which Poincaré would draw figures on a wall with his finger to explain his reasoning to his fellow students, or how he instantly gave very creative solutions to geometry problems that were posed to him (Appell gives a number of such examples). Also very informative is Darboux's eulogy [15] of 1913.

As we shall see, Poincaré covered a lot of ground in the papers mentioned above. In the light of this, it seems almost incredible that this was really *only a small part of the huge canvas on which he was working during this time*. In a series of long papers starting from 1880 he had created the *qualitative theory of ordinary differential equations*. Then, impelled with the desire to solve linear differential equations having algebraic coefficients, he had created and developed yet another theory, that of *Fuchsian and Kleinian groups*. Hard on the heels of this had come his prize-winning 1890 paper [55] on the 3-body problem, which was now being elaborated further in his three volume *treatise on celestial mechanics* [57]. Add to this dozens of courses delivered in almost every imaginable area of *theoretical physics*, and we are left gasping at the very idea that he had any time left to create and develop yet another very original mathematical theory!

Darboux tells us that Poincaré's "answers came with the rapidity of an arrow" and that "when he wrote a memoir, he drafted it at one go, limiting himself to just some crossings out, without coming back to what he had written". Despite this, Poincaré's writings are characterized by great lucidity of thought, an intuitive ability of getting at once to the heart of the matter, and clarity of exposition.

At Mittag-Leffler's request Poincaré wrote in 1901 an analysis of his own work [63]. Of these hundred odd pages only four, pp. 100–103, deal with "Analysis Situs" and its first two Complements. Here he recalls (this line occurs in the Introduction of "A.S." as well) that "*geometry is the art of reasoning well with badly made figures. Yes, without*

HISTORY OF TOPOLOGY
Edited by I.M. James

Henri Poincaré (1854–1912)

doubt, but with one condition. The proportions of the figures might be grossly altered, but their elements must not be interchanged and must conserve their relative situation. In other terms, one does not worry about quantitative properties, but one must respect the qualitative properties, that is to say precisely those which are the concern of Analysis Situs." Thus his hope was that topology would render in higher dimensions much the same service which these "badly made" figures give to ordinary geometry. After mentioning the previous work of Riemann and Betti in this direction he continues as follows.

"*As for me, all the diverse paths on which I was successively engaged have led me to Analysis Situs. I had need of the ideas of this science to pursue my studies on curves defined by differential equations and for extending these to higher order differential equations and in particular to those of the three body problem. I had need of it for the study of multi-valued functions of 2 variables. I had need of it for the study of periods of multiple integrals and for the application of this study to the development of the perturbation function. Finally I glimpsed in Analysis Situs a means of attacking an important problem in the theory of groups, the search for discrete or finite groups contained in a given continuous group. It is for all these reasons that I devoted to this science a fairly long work.*"

Indeed Poincaré's other works probably contain just as much interesting "topology" – in the wide sense of the word – as "Analysis Situs" and its five Compléments! For example, his memoirs on the qualitative theory of differential equations contain the **Poincaré index formula** giving the Euler characteristic of a surface as the sum of the local degrees of a generic vector field at its isolated singularities: this was generalized later to higher dimensions by Hopf [27]. And, of course, the study of periods of multiple integrals is "de Rham–Hodge theory"; and of invariant integrals, which he introduced while doing celestial mechanics, that of "symplectic transformations"; and the work on perturbation functions of astronomy the "small divisors problem". (A seminar run by A. Chenciner has recently been analysing Poincaré's treatise on celestial mechanics.) The **last geometric theorem** [70] which Birkhoff [5] resolved shortly after Poincaré's premature death, is also equally "topology". It says that if a volume preserving diffeomorphism of the annulus moves its two bounding circles in opposite directions than it must have two fixed points. (A recent paper of Golé and Hall [24] shows that the existence of a fixed point does follow by slightly modifying Poincaré's original attempt.)

However, we shall confine ourselves in the following to "Analysis Situs" and its five "Compléments". Section 1 is a summary of "Analysis Situs". Section 2 contains notes on this summary, intended mostly to connect Poincaré's contributions with future developments. For the "Compléments" (these contain more material than "A.S." itself) we have summarized and annotated in tandem in Section 3. We shall pause for just a few remarks before we embark on this task. We have not hesitated to use modern notations, and even ideas, whenever this seemed to help in understanding Poincaré's mathematics. For example, Riemann's *connectivity* of a surface was 1 more than b_1, so Poincaré defined his Betti numbers to be 1 more than the modern ones: we have lowered them by 1. Again, we have discarded Poincaré's *congruences*, and just used $\partial w = c$ to denote a boundary. On the other hand, for homotopy between loops, we liked Poincaré's *equivalences* $A \equiv B$, and have, like him, combined these using *additive*, rather than the modern multiplicative notation. Lastly, Poincaré's grade in art class notwithstanding – see Darboux [15, p. XIX] for the surprising answer! – it is clear to anybody who reads him that he thought via *figures*: so we have added some, but we remark that, of those given below, five are his own.

1. A summary of "Analysis Situs"

Introduction. The branch of Geometry called *Analysis Situs* describes the relative situation between some points, lines, and surfaces, without bothering about their sizes. There is a similar Analysis Situs in more than three dimensions as has been demonstrated by Riemann and Betti (and which we shall develop further in this paper). We expect it will have many applications, e.g., the following three.

"The classification of algebraic curves by means of their genus is based, following Riemann, on the classification of closed real surfaces, made from the viewpoint of *Analysis Situs*. An immediate induction now tells us that the classification of algebraic surfaces and the theory of their birational transformations is intimately tied to the classification of closed real (hyper)surfaces in 5-space from the viewpoint of *Analysis Situs*. M. Picard, in a work which has been hailed by the Académie des Sciences, has already stressed this point."

"Besides, in a series of memoirs published in the *Journal de Liouville* and entitled "*Sur les courbes définis par les équations différentielles*", I have used ordinary 3-dimensional Analysis Situs to study (second order) differential equations. The same researches have also been pursued by M. Walther Dyck. One sees easily that a generalized *Analysis Situs* would permit us to similarly treat higher order equations, and in particular those of Celestial Mechanics."

"M. Jordan has analytically determined the groups of finite order which are contained in the linear group of n variables. M. Klein had previously, by a geometrical method of rare elegance, solved the same problem for the linear group of two variables. Could not one extend the method of M. Klein to a group of n variables, or even *an arbitrary continuous group*? I have not been able to do this so far, but I have thought long on this question, and it appears to me that the solution should depend on a problem of *Analysis Situs* and that the generalization of the celebrated theorem of Euler should play a role in this."

§ 1. *Première définition des variétés.* A nonempty subset V of n-space defined by p equations $F_\alpha(x_1, \ldots, x_n) = 0$ and q inequalities $\phi_\beta(x_1, \ldots, x_n) > 0$, where the functions F and ϕ are continuously differentiable, will be called a *variety of dimension $n - p$* if the rank of the matrix $[\partial F_\alpha / \partial x_i]$ is equal to p at all points V.

When a variety is defined only by inequalities, i.e. when $p = 0$, then it is called a *domain*. Furthermore, varieties which are one-dimensional, respectively, not one-dimensional but having codimension one, are called *curves*, respectively, (hyper) *surfaces*. A variety will be called *bounded* (finie) if the distance of all its points from the origin is less than some constant.

We will only consider *connected* (continue) varieties, regarding others we only remark that they can be decomposed into a finite or infinite number of connected varieties. For example, the plane curve $x_2^2 + x_1^4 - 4x_1^2 + 1 = 0$ is the disjoint union of the two connected curves obtained by adjoining to its defining equation either the inequality $x_1 < 0$ or else $x_1 > 0$. (See Fig. 1.)

By the *complete boundary* (frontière complète) of a variety V we will mean the set of all points of n-space satisfying $\{F_\alpha = 0,\ 1 \leqslant \alpha \leqslant p,\ \phi_\beta = 0;\ \phi_\gamma > 0,\ 1 \leqslant \gamma \neq \beta \leqslant q\}$ for some $1 \leqslant \beta \leqslant q$. However, sometimes we shall think of the largest $(n - p - 1)$-dimensional variety contained in this set as the true *boundary* (we shall denote this by ∂V) of V. A *boundaryless* (illimité) variety will be one which has empty true boundary; if furthermore it is connected and bounded we shall call it *closed* (fermée).

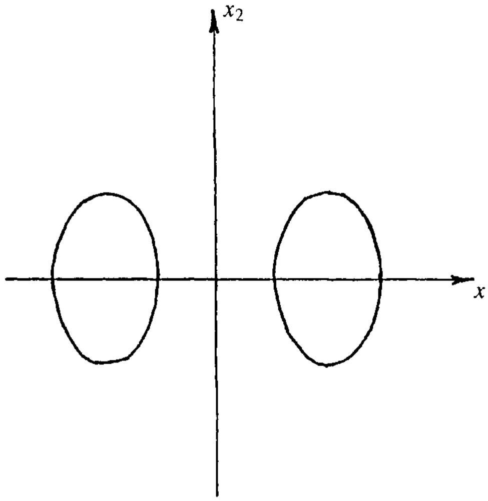


Fig. 1. $x_2^2 + x_1^4 - 4x_1^2 + 1 = 0$.

§ 2. *Homéomorphisme.* Consider the "group" formed by all maps between open subsets of n-space for which the functional determinant is always nonzero: the science whose object is the study of this and some other analogous "groups" is called *Analysis Situs*.

By a *diffeomorphism* (homéomorphisme) between two varieties of n-space we shall mean a bijection between them which extends to a differentiable bijection between open Euclidean sets obtained by replacing their defining equalities $F_\alpha = 0$ by some inequalities $-\varepsilon < F_\alpha < +\varepsilon$. A similar definition can be given for more complicated figures, made up of many varieties, of n-space.

§ 3. *Deuxième définition des variétés.* Consider first m-dimensional varieties v of n-space satisfying a system of n equations $x_i = \theta_i(y_1, \ldots, y_m)$ with $\operatorname{rank}[\partial\theta_i/\partial y_j] \equiv m$, and some inequalities $\psi(y_1, \ldots, y_m) > 0$.

For example, the system of three equations $x_1 = (R + r\cos y_1)\cos y_2$, $x_2 = (R + r\cos y_1)\sin y_2$ and $x_3 = r\sin y_1$ defines a *torus*. (See Fig. 2).

Indeed in the following definition we may only use those v's which, unlike that of the above example, have a one–one θ. Furthermore, we can assume these functions to be (real) analytic: this follows because we can always replace θ by an arbitrarily close real analytic θ'.

Given two such varieties v and v' we shall say that they are *analytic continuations* of each other iff their intersection $v \cap v'$ is also an m-dimensional variety of the above type. As per our new definition a "variety" – or sometimes, to use a different word, a *manifold* – will mean any *connected network* (réseau continu) M of varieties v related to each other by analytic continuation (i.e. a graph whose vertices are varieties of the above type, with two vertices contiguous in the graph iff they are analytic continuations of each other).

We shall see later that such an M need not be definable by equations of the type given in § 1; however, as shown below *any variety V of § 1 is also a variety as per this second definition.*

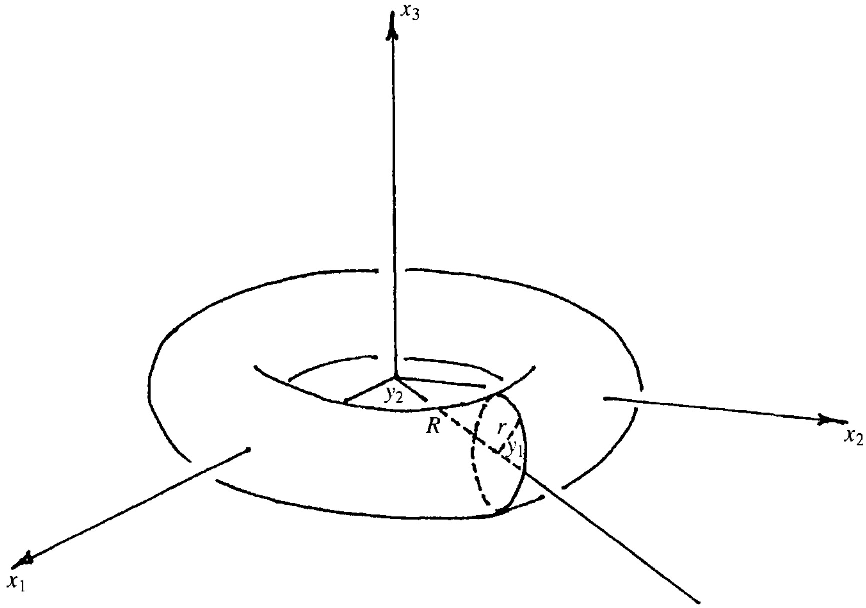


Fig. 2. $(x_1^2+x_2^2+x_3^2+R^2-r^2)^2-4R^2(x_1^2+x_2^2)=0.$

To see this we shall use the well-known result that if the n real analytic equations $y_i = F_i(x_1, \dots, x_n)$ are such that their functional determinant is nonzero at x, then they have real analytic solutions $x_i = \theta_i(y_1, \dots, y_n)$ valid in some neighbourhood of $F(x)$.

Now let P be any point of V, defined as in § 1 by p equations $F_\alpha(x_1, \dots, x_n) = 0$ and some inequalities $\phi(x_1, \dots, x_n) > 0$. It clearly suffices to find an $(n-p)$-dimensional variety ν_P of the above type such that $P \in \nu_P \subseteq V$. To see this choose any $n-p$ additional analytic functions $F_{p+1}, \dots, F_n$ of n variables, which vanish at P, and are such that the functional determinant of all the n functions F_i is nonzero at P. So we can solve the n equations $u_i = F_i(x_1, \dots, x_n)$ to get real analytic solutions $x_i = \theta_i(u_1, \dots, u_n)$ in some neighbourhood of $F(P) = 0$ specified by some inequalities $\lambda(u_1, \dots, u_n) > 0$. By making this neighbourhood smaller, if need be, we will assume also that these inequalities imply the defining inequalities $\phi(x_1, \dots, x_n) > 0$ of V. Thus the n equations $x_i = \theta_i(0, \dots, 0, \ y_1, \dots, y_{n-p})$ and the inequalities $\lambda(0, \dots, 0, \ y_1, \dots, y_{n-p}) > 0$ are satisfied by P and imply the defining p equations $F_\alpha(x_1, \dots, x_n) = 0$ and inequalities $\phi(x_1, \dots, x_n) > 0$ of V, and so give a ν_P such that $P \in \nu_P \subseteq V$.

§ 4. *Variétés opposées.* We will assume that if we interchange two of the defining equations of a V as in § 1 then we no longer get V, but the *opposite variety* $-V$. More generally, given any nonsingular matrix $A_{\alpha\beta}$ of functions, the ordered set of equations $\sum_\alpha A_{\beta\alpha} F_\alpha = 0$ gives V, respectively. $-V$, iff $\det(A_{\alpha\beta})$ is positive, respectively, negative.

Likewise, for a ν as in § 3, we shall assume that interchanging any two of the m parameters y_i no longer gives ν, but the opposite variety $-\nu$, and more generally, if the parameters undergo a transformation $y_1, \dots, \ y_m \mapsto z_1, \dots, z_m$, we shall assume that the resulting variety is ν or $-\nu$, depending on whether the transformation's functional determinant is positive or negative.

The two concepts will be tied to each other by stipulating that if $\nu_P \subseteq V$ as in § 3, then ν_P has the correct orientation iff the $n \times n$ functional determinant mentioned there is positive.

Also the order of the defining equations of an $(n-p-1)$-dimensional nonsingular variety occurring in the true boundary of V will be deemed to be that in which one first writes the equations of V and then puts the new equation $\phi = 0$ in the very end.

§ 5. *Homologies.* Suppose V is a subvariety of a manifold M whose oriented boundary consists of k_i copies of the variety ν_i for $1 \leqslant i \leqslant a$, and s_j copies of the variety $-\mu_j$ for $1 \leqslant j \leqslant b$. Then we shall write

$$k_1\nu_1 + \cdots + k_a\nu_a \simeq s_1\mu_1 + \cdots + s_b\mu_b,$$

and refer to this relation as a *homology* of M. These "homologies can be combined with each other just like ordinary equations" (i.e. the sum of any two homologies will also be deemed to be a homology, and we can take any term to the other side provided we change its sign, and so on).

In case M has a boundary, the notation $k_1\nu_1 + \cdots + k_a\nu_a \simeq \varepsilon$ will indicate that the sum of the varieties on the left is homologous to a sum of varieties contained in this boundary.

§ 6. *Nombres de Betti.* The cardinality of a maximal *linearly independent* set – i.e. one for which there is no nontrivial homology between its members – of closed r-dimensional subvarieties of M will be called the *r-th Betti number* $b_r(M)$ of M. (In the paper it is $b_r(M) + 1$ which is called the r-th Betti number and is denoted by P_r.)

Let us make these definitions clearer by an example. Let D be a domain of 3-space bounded by n disjoint surfaces S_i. Then its Betti numbers are $b_1(V) = (1/2)\sum_i b_1(S_i)$ and $b_2(V) = n - 1$, where each $b_1(S) + 1$ is necessarily odd, being the connectivity of S as defined by Riemann.

§ 7. *Emploi des intégrales.* The integral

$$\int_V \sum \omega_{\alpha_1 \ldots \alpha_r}(x_1, \ldots, x_n)\, \mathrm{d}x_{\alpha_1} \cdots \mathrm{d}x_{\alpha_r}, \qquad 1 \leqslant \alpha_i \leqslant n,$$

or briefly $\int_V \omega$, over any r-dimensional variety V (which is equipped with an orientation as in § 4) of n-space, will be defined to be

$$\sum_\nu \int \sum \omega_{\alpha_1 \ldots \alpha_r}(x_1, \ldots, x_n) \det\left(\partial x_{\alpha_i}/\partial y_j\right) \mathrm{d}y_1 \cdots \mathrm{d}y_r,$$

where $V = \sum \nu$, and for each ν, the multiple integral is evaluated, using the equations $x_i = \theta_i(y_1, \ldots, y_r)$ of ν, between the limits of y_i prescribed by the inequalities of ν.

About the functions $\omega_{\alpha_1 \ldots \alpha_r}$ – or $\omega(\alpha_1, \ldots, \alpha_r)$ – being integrated it will be assumed that they merely change sign when any two of the indices α_i are interchanged. The result below is from paragraph 2, entitled "*Conditions d'intégrabilité*", of Poincaré [54], 1887.

The integrals $\int_V \omega$ are zero for all closed varieties V of n-space if and only if the $\binom{n}{r+1}$ cyclic sums

$$\sum (-1)^{ri} \partial/\partial x_{\alpha_i} \left[\omega(\alpha_{i+1}, \ldots, \alpha_{r+1}, \alpha_1, \ldots, \alpha_{i-1})\right],$$

are identically zero.

The proof given there shows also that if these $\binom{n}{r+1}$ conditions hold in the vicinity of a given m-dimensional submanifold M of n-space, then the above integrals are still zero over closed subvarieties V of M; in fact, for this, just $\binom{m}{r+1}$ analogous conditions suffice.

For any functions ω satisfying these conditions, one can find at most $b_r(M)$ numbers such that the integral $\int_V \omega$ of ω over any closed r-variety V of M is a linear integral combination of these numbers (we omit the proof given). In other words, *the indefinite integral $\int \omega$, of any functions ω satisfying the conditions of integrability near M, has at most $b_r(M)$ periods.* Further, it can be shown that this bound is the best possible, i.e. there exist such functions ω having exactly $b_r(M)$ periods. For $r = 1,\ m-1$, this interpretation of the numbers $b_r(M)$ was given by Betti himself.

§ 8. *Variétés unilatères et bilatères.* A manifold M (as defined in § 3) will be called *two-sided* (bilatère) iff we can assign an orientation (as in § 4) to each of the varieties ν of its connected network, in such a way that the $m \times m$ determinant $\det(\partial y_i/\partial y'_j)$ is positive whenever ν is contiguous to ν'.

Otherwise M will be called *one-sided* (unilatère) and deemed equal to its own opposite $-M$. This happens iff either, its network contains a contiguous pair $\{\nu, \nu'\}$ with the determinant not of the same sign in all the components of $\nu \cap \nu'$, or else, has a *one-sided circuit* $(\nu_1, \ldots, \nu_q)$, i.e. one for which making the determinant between ν_i and ν_{i+1} positive for $1 \leqslant i \leqslant q-1$, makes the determinant between ν_q and ν_1 negative.

However, to justify these definitions (i.e. to see that one- or two-sidedness is a property of the space M) one also must check (we omit the proof given) that the same alternative continues to hold if a new local parametrization ν^* is added to the connected network.

Everyone knows of the one-sided surface which one obtains by folding a paper rectangle $ABCD$ and then gluing the edges AB and CD in such a way that A is glued to C and B to D. (See Fig. 3.)

Examples of two-sided manifolds are easier to give: for example, in n-space, any domain, or any curve, or any closed $(n-1)$-dimensional surface, are all two-sided. Indeed much more is true: *the varieties V of § 1 are all necessarily two-sided* (we omit the proof given).

This shows that "variétés" as defined in § 3 (i.e. manifolds) do not all satisfy equations of the type given in § 1.

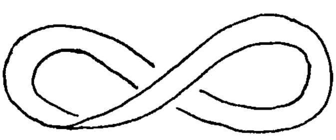


Fig. 3.

§ 9. *Intersection des deux variétés.* Given two points x and x' of n-space lying in oriented varieties ν and ν' of dimensions p and $n-p$ we denote by $S(x, x') \in \{-1, 0, +1\}$ the sign of the $n \times n$ determinant

$$\begin{vmatrix} \partial x_i/\partial y_j \\ \partial x_i'/\partial y_k' \end{vmatrix}, \quad 1 \leqslant i \leqslant n,\ 1 \leqslant j \leqslant p,\ 1 \leqslant k \leqslant n-p,$$

and using it deem the *algebraic number of intersections* of ν and ν' to be $N(\nu, \nu') = \sum\{S(x, x')\colon x = x'\}$. More generally one likewise (we omit the details given) counts the algebraic number of intersections of two oriented complementary dimensional submanifolds of any oriented n-manifold M (intuitively one counts an intersection x as $+1$ iff the orientation of ν at x followed by that of ν' agrees with that of M).

We note that $N(\nu, \nu')$ changes sign if the orientation of any one of the three manifolds $\{M, \nu, \nu'\}$ is reversed, and that

$$N(\nu', \nu) = (-1)^{\dim \nu \, \dim \nu'} N(\nu, \nu').$$

If closed $(n-p)$-dimensional varieties $V_i \subset M$ are such that there exists a p-dimensional cut C of M having intersection number $\sum_i k_i N(C, V_i)$ nonzero then we cannot have $\sum_i k_i V_i \simeq 0$; and conversely, if this homology does not hold, then such a cut C can be found. Here, by a *cut* (coupure) of M we mean either any closed subvariety, or else one whose boundary is contained in the boundary of M.

For case $p = 1$ and M closed (we omit the proof given for the case $\partial M \neq \emptyset$) the direct part follows because if $\partial W = V_1 + \cdots + V_t$, then the oriented closed curve C must go as many times from the complement of W into W, as it goes from W into its complement. Conversely the given conditions ensure (we omit details given) that there is no nontrivial homology amongst these V_i's. So the complement W of $V_1 \cup \cdots \cup V_t$ in M must be connected, for otherwise the boundary of any component of this complement will furnish a nontrivial homology between some of these V_i's. Now we can obtain the required closed curve C by joining the extremities y and z, of a small arc yxz cutting V_1 at x, to each other in W. (Regarding the sketched generalization see the First Complement.)

If follows that, for a closed M, the Betti numbers equidistant from the two ends are equal, i.e. that $b_p(M) = b_{n-p}(M)$ for $0 \leqslant p \leqslant n$. "This theorem has not been, I believe, ever been stated; it is, however, known to many, who have even found some applications of it."

To see this choose in M maximal sets of linearly independent p- and $(n-p)$-dimensional closed oriented varieties $\{C_1, \ldots, C_\lambda\}$ and $\{V_1, \ldots, V_\mu\}$, where $\lambda = b_p(M)$ and $\mu = b_{n-p}(M)$. If the number λ of linear equations $\sum_i x_i N(C_j, V_i) = 0$ was less than the number μ of unknowns x_i, they would have a nontrivial solution $x_i = k_i$. Then (by the direct part of the previous result) we would have $\sum_i k_i N(C, V_i) = 0$ for *all* closed r-dimensional C's. So (by the converse part of that result) we would have $\sum_i k_i V_i \simeq 0$ in M. Since this is not so we must have $\lambda \geqslant \mu$. Likewise $\mu \geqslant \lambda$.

Let us now consider the middle Betti number $b_{n/2}(M)$ for the case n even: *if $n \equiv 2 \bmod 4$ then $b_{n/2}(M)$ is even.*

To see this choose $b = b_{n/2}(M)$ linearly independent closed $(n/2)$-dimensional subvarieties $V_1, V_2, \ldots$ of M, and consider the $b \times b$ determinant $N = [N(V_i, V_j)]$, where by $N(V, V)$ we mean $N(V, V')$ for a suitable $V' \simeq V$. Since $n/2$ is odd this determinant is skewsymmetric, and so, if b were odd, it would be zero. So we would be able to find k_i's not all zero such that $\sum_j k_j N(V_i, V_j) = 0$. So as in last argument we would have $\sum_j k_j N(C, V_j) = 0$ for all $(n/2)$-varieties C, which implies $\sum_j k_j V_j \simeq 0$ in M, a contradiction.

This is no longer true if 4 divides n, nor if M is one-sided, as we shall see later by means of examples.

§ 10. *Représentation géométrique.* "There is a way of describing three-dimensional varieties situated in four space which facilitates their study remarkably", viz., as some polytope(s) P having an even number of facets, with the facets identified in pairs.

For a two-sided variety these *conjugate* facets $F \equiv F'$ are such that if we walk on P along ∂F keeping F to our left, then the corresponding walk on P along $\partial F'$ should keep F' to our right.

Let me recall something similar from ordinary space, viz. cutting a torus along a meridian and a parallel, we can describe it as a square $ABDC$ with identifications $AB \equiv CD$, $AC \equiv BD$, of its sides. Likewise we can identify pairs of facets of a cube in, for example, the following five ways, which all satisfy the above criterion for two-sidedness. (In the paper, the fifth example, i.e. $\mathbb{R}P^3$, is defined by the antipodal conjugation of the facets of an octahedron instead of the cube.) (See Fig. 4.)

Nevertheless, not all of the above facet conjugations of the cube can occur: we shall see below that Examples 1, 3, 4 and 5 are admissible but Example 2 is not.

First note that – in complete analogy with the formation of cycles in the theory of Fuchsian groups – the prescribed facet conjugations partition off the sets of edges and vertices into *cycles* consisting of edges or vertices which get identified to each other, e.g., for Example 2 these are $AB \equiv B'D \equiv C'C \equiv B'A' \equiv AC \equiv DD'$, $AA' \equiv DC \equiv C'A' \equiv B'B \equiv C'D' \equiv DB$, $A \equiv B' \equiv C' \equiv D$, and $B \equiv D' \equiv C \equiv A'$.

For each cycle α of vertices let f_α = its cardinality, e_α = half the sum of the number of facets incident to each member of α, and v_α = number of cycles of edges incident to

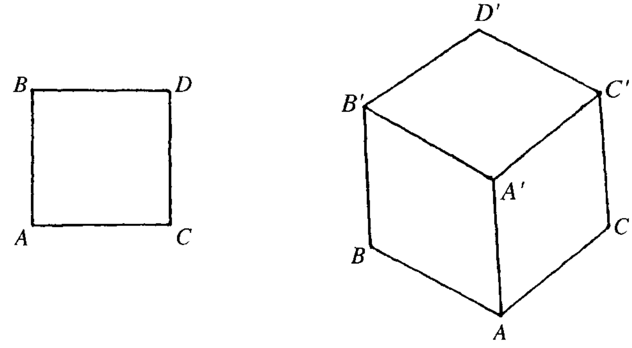


Fig. 4. Square and cube.

	Example 1	*Example* 2	*Example* 3	*Example* 4	*Example* 5
$ABDC \equiv$	$A'B'D'C'$	$B'D'C'A'$	$B'D'C'A'$	$B'D'C'A'$	$D'C'A'B'$
$ACC'A' \equiv$	$BDD'B'$	$DD'B'B$	$DD'B'B$	$BDD'B'$	$D'B'BD$
$ABB'A' \equiv$	$CDD'C'$	$DD'C'C$	$C'CDD'$	$CDD'C'$	$D'C'CD$

vertices of α, taking care to count each such edge cycle twice if both vertices of an edge are in α. We assert that *for a conjugation of facets to be admissible, it is necessary and sufficient, that one has* $v_\alpha - e_\alpha + f_\alpha = 2$ *for all* α.

To see this note that the subdivision, of the portion of the variety consisting of all points at a distance $< \varepsilon$ from any vertex α, is diffeomorphic to a *star* (aster), i.e. a figure formed by some solid angles arranged around a single vertex in such a way that each point of space belongs to one and only one of them. Since v_α, e_α, and f_α are the number of rays, faces and solid angles of this star, the required condition follows by using Euler's formula.

The above condition holds (we omit the computations given) for all our examples excepting the second, for which $v_\alpha - e_\alpha + f_\alpha = 0 \forall\alpha$.

§ 11. *Réprésentation par un groupe discontinu.* In analogy with the theory of Fuchsian groups one may sometimes describe a three-dimensional variety via a properly discontinuous group of substitutions S of ordinary space.

Indeed, consider any fundamental domain D of this group. Subdivide its boundary into surfaces F which it shares with neighbouring translates $S(D)$, with F' denoting the surface shared by P and $S^{-1}(D)$. Then the variety can be obtained, just as in the last article, from D, by gluing all its conjugate pairs of facets F, F' to each other.

EXAMPLE 6. Consider the group G_T of transformations of 3-space generated by

$$(x, y, z) \mapsto (x+1, y, z), \quad (x, y, z) \mapsto (x, y+1, z), \quad \text{and}$$
$$(x, y, z) \mapsto (\alpha x + \beta y, \gamma x + \delta y, z+1),$$

where α, β, γ, and δ are four chosen integers with $\alpha\delta - \beta\gamma = 1$, i.e. such that $T = \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} \in \mathrm{SL}(2, \mathbb{Z})$. One can check (we omit the proof given) that this group is discontinuous, with the unit cube P as a fundamental domain. So we obtain a variety M_T by conjugating pairs of facets of a subdivided cube P_T.

The simplest case is when $T = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, now $P_T = P$, and one recovers Example 1. For $T = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$, once again $P_T = P$, but now the conjugations are that of Example 4. When $T = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, then P has a nontrivial subdivision P_T, which with its facet conjugations is shown in Fig. 5. More generally, any P_T has the same (unsubdivided) vertical facets as P, but the number of its top and bottom cells will increase with the size of the entries of the matrix T.

§ 12. *Groupe fondamental.* Suppose given a system $\mathcal{F}$ of multiple-valued locally defined continuous functions F_α on the variety V, which return to their initial values if we trace small loops on the variety. We will denote by $g_{\mathcal{F}}$ the group of all permutations of the branches which ensue if we follow them over all closed loops starting and ending at a given base point b of the variety.

To be specific we may consider solutions F_α of an equation

$$\mathrm{d}F_\alpha = \sum_{1 \leqslant i \leqslant n} X_{\alpha,i}(x_1, \ldots, x_n; F_1, \ldots, F_\lambda)\, \mathrm{d}x_i,$$

$$
\begin{aligned}
AA'C'C &\equiv BB'D'D\\
ABB'A' &\equiv CDD'C'\\
ABC &\equiv A'D'C'\\
BCD &\equiv B'A'D'
\end{aligned}
$$

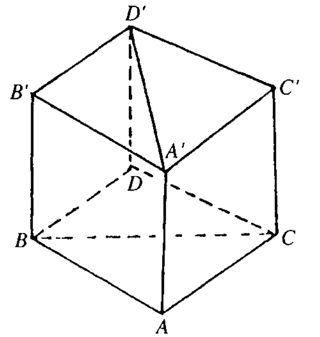


Fig. 5. P_T for $T = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$.

where the $n\lambda$ coefficients $X_{\alpha,i}$ are given functions of x_k and F_i, single valued and continuous, together with their derivatives, in a neighbourhood of V, and satisfying there the conditions of integrability

$$\frac{\partial X_{\alpha,i}}{\partial x_j} + \sum_\beta \frac{\partial X_{\alpha,i}}{\partial F_\beta} X_{\beta,j} = \frac{\partial X_{\alpha,j}}{\partial x_i} + \sum_\beta \frac{\partial X_{\alpha,j}}{\partial F_\beta} X_{\beta,i}.$$

We note that if we trace a *lacet* C, i.e. go along any path from b to c, followed by a small loop at c, and then return to b along the original path, then we only get the identity substitution $S_C \in g_{\mathcal{F}}$. Also for the loop C_1C_2, i.e. C_1 followed by C_2, one has $S_{C_1C_2} = S_{C_1}S_{C_2}$.

Motivated by this we shall set $C \equiv 0$ for all lacets, and $C_1 + C_2 \equiv C_1C_2$. A general *equivalence*

$$k_1C_1 + k_2C_2 + \cdots \equiv k_\alpha C_\alpha + k_\beta C_\beta + \cdots$$

will be between integral combinations of loops based at b. One adds them just like homologies but the order of the terms cannot be interchanged. So, e.g., $A \equiv B$ and $C \equiv D$ implies $A + C \equiv B + D$ but not $C + A \equiv B + D$. Also note that from $2A \equiv 0$ one does not have the right to conclude $A \equiv 0$. Another difference from homologies is that a base point b is involved in their definition.

(The above careful distinction between equivalences and homologies notwithstanding at one point it is erroneously written that the boundary of a two dimensional variety of V is equivalent to zero; also see [66, p. 390]; analogously on p. 293 of [61] $C_1 \equiv 0$ should be $C_1 \simeq 0$. This surprising error is rectified in [69], see pp. 451–452.)

For any $g_{\mathcal{F}}$, we obviously have (1) $C \equiv C_1 + C_2 \Rightarrow S_C = S_{C_1}S_{C_2}$ and (2) $C \equiv 0 \Rightarrow S_C = \mathrm{Id}$. We shall denote by G the *fundamental group* G of substitutions S_C satisfying (1) and the stronger (2′) $C \equiv 0 \Leftrightarrow S_C = \mathrm{Id}$. There is thus an epimorphism from G onto any $g_{\mathcal{F}}$. This can be one–one, but is in general not so, because some loop C, which is not decomposable into lacets, may still give the identity substitution in $g_{\mathcal{F}}$.

§ 13. *Équivalences fondamentales.* One can always find some *fundamental loops* $C_1, \ldots, C_p$ such that any loop is equivalent to a combination of these. The relations subsisting between them, which determine the form of the group G, will be called *fundamental equivalences*.

For a variety described as in § 10, using just one polytope P, each pair of conjugate facets $\{F, F'\}$ gives a fundamental loop C: proceed along a straight line from the base point $b \in \mathrm{int}P$ to a point $x \in F$, and then along another straight line from the conjugate point $x' \in F'$ back to b. Denoting each edge of P as the product of its two incident facets any cycle of edges is of the type $F_1 F'_\mu \equiv F_2 F'_1 \equiv \cdots \equiv F_\mu F'_{\mu-1}$. Each of these gives us a fundamental equivalence $C_1 + C_2 + \cdots + C_\mu \equiv 0$.

Ignoring the order of the terms in these fundamental equivalences, one obtains the *fundamental homologies* between these loops. These give $b_1(M)$, which for these three-dimensional varieties M, also equals $b_2(M)$.

For Example 3 of § 10 (computations are also given for Examples 1, 4 and 5) this method gives fundamental equivalences $2C_1 \equiv -2C_2 \equiv 2C_3$, $4C_1 \equiv 0$, which show that G is isomorphic to the *hypercubic* order eight group $\langle i, j, k\rangle$, and that $b_1(M) = b_2(M) = 0$.

For Example 6 of § 11 the fundamental group is evidently isomorphic to G_T. Denoting (the loops inducing) the three defining substitutions of this group, respectively, by C_1, C_2, and C_3, we see that

$$C_1 + C_2 \equiv C_2 + C_1, \quad C_1 + C_3 \equiv C_3 + \alpha C_1 + \gamma C_2, \quad \text{and}$$
$$C_2 + C_3 \equiv C_3 + \beta C_1 + \delta C_2.$$

These are fundamental equivalences, because using them, any member of G_T can be written $m_3C_3 + m_1C_1 + m_2C_2$, which can be checked to be the identity substitution iff $m_1 = m_2 = m_3 = 0$. The fundamental homologies are thus $(\alpha - 1)C_1 + \gamma C_2 \simeq 0$ and $\beta C_1 + (\delta - 1)C_2 \simeq 0$. These homologies are trivial iff $T = I$. In this, and only this, case does one have $b_1 = b_2 = 3$. For $T \neq I$, the above homologies are proportional iff the determinant $\begin{vmatrix} \alpha - 1 & \gamma \\ \beta & \delta - 1 \end{vmatrix}$ vanishes, i.e. iff $\mathrm{tr}(T) = \alpha + \delta = 2$. So in this case, and only in this case, the Betti numbers are $b_1 = b_2 = 2$. In all other cases the homologies are nontrivial and nonproportional, and so we have $b_1 = b_2 = 1$.

§ 14. *Conditions de l'homeomorphisme.* One knows that *closed 2-manifolds are diffeomorphic iff their Betti numbers are same*. This follows, for example, from the study of the periods of Abelian functions. In any Riemann surface R with z as variable, one can introduce a new complex variable t, such that z is a Fuchsian function of t and that t, considered as function of z, has no singular point on the surface R. The Fuchsian group is obviously nothing else but the fundamental group G of R. This rules out the possibility that some cycle of vertices of the Fuchsian polygon, R_0 or $R_0 + R'_0$, has angle sum $2\pi/n$ with $n > 1$, for then we would get a nonidentity substitution as we describe a lacet around this point of the variety. The possibility of a non simply connected Fuchsian polygon $R_0 + R'_0$ is ruled out because then a nontrivial loop C of this polygon would yield the identity substitution of the Fuchsian group. Thus only Fuchsian groups of the first kind with angle sums 2π at all cycles of vertices can occur. All of these groups which are of the same genus are isomorphic, and it is for this reason that all closed two-dimensional manifolds having the same Betti number are diffeomorphic.

However, in dimensions > 2 the questions of *Analysis Situs* become much more complicated and, as we shall see, it is no longer the case that closed varieties having the same Betti numbers must be diffeomorphic.

Let us return to our sixth example (§§ 11 and 13). We shall say that our $T \in \mathrm{SL}(2; \mathbb{Z})$ is *hyperbolic, parabolic*, or *elliptic*, according as its two eigenvalues are real distinct, equal, or imaginary. If $G_{T'} \cong G_T$ then a pair of elements of G_T can correspond to the elements C_1' and C_2' of $G_{T'}$ only if they are commuting but linearly independent. So let $a_3C_3 + a_1C_1 + a_2C_2$ and $b_3C_3 + b_1C_1 + b_2C_2$ be any two elements of G_T such that no nonzero multiple of either equals a multiple of the other. *For T hyperbolic these elements commute iff $a_3 = 0$ and $b_3 = 0$, and for T elliptic or $T = -I$ this happens iff $a_3 \equiv 0 \bmod \nu$ and $b_3 \equiv 0 \bmod \nu$, where $\nu > 1$ is such that $T^\nu = I$* (we omit the proof given).

One can say more: *the groups G_T and $G_{T'}$ are isomorphic iff T is in the same conjugacy class as T'*. To check this (we omit the very long details) we choose generators $C_1,\ C_2,\ C_3$ of G', such that C_1 and C_2 are commuting but linearly independent and such that $C + C_3 \equiv C_3 + T(C)$ for all $C \in \langle C_1, C_2\rangle$. The idea of the proof is that, in these relations, T gets replaced by the *similar* (transformée) matrix UTU^{-1} if we replace C by $U(C)$. We give, for various cases, sequences of such elementary moves by means of which we finally replace T by T' in these relations.

The number of these conjugacy classes is infinite because similar matrices must have the same *trace*; however conversely, just like noncongruent quadratic forms can have the same determinant, two linear substitutions can be nonsimilar even if they have the same trace.

Thus there are infinitely many nondiffeomorphic M_T's. Since, on the other hand, their Betti number $b_1 = b_2$ can be only 1, 2, or 3, it follows that, *for two closed varieties to be diffeomorphic, it does not suffice that the Betti numbers be the same.* This follows equally because, for our third example G was of order 8, for the fifth ($\cong \mathbb{R}P^3$) of order 2, and for the unit sphere of 4-space it is of order 1, yet for all of these $b_1 = b_2 = 0$. So it seems natural that only those varieties should be called *simply connected* for which G is null.

It would be interesting to know which fundamental equivalences can actually arise, and how one can construct these varieties, and whether two varieties having the same G must be diffeomorphic? Such questions need a difficult and long study, so I will not pursue these here.

However, I do want to draw attention to one point. Riemann had studied algebraic curves as two-dimensional varieties, likewise algebraic surfaces are four-dimensional varieties. M. Picard has shown that for all but some very special algebraic surfaces one always has $b_1 = 0$. This paradoxical looking result appears less so now: the group G can be quite complex and yet the Betti numbers can be very small.

§ 15. *Autres modes de génération.* One may give other definitions of varieties which are, so to speak, intermediate between the two given before, e.g., if the equations of § 1 depended on q parameters, then the dimension of our variety would increase by q, or, if the parameters of § 3 were subject to λ equations, the dimension would decrease by λ.

Also, given a variety W, and a group G which preserves it, one may construct a variety V, to each point of which corresponds one and only one *orbit* (système de points) of W. This variety will be two-sided iff the functional determinants of the substitutions of G are positive with respect to compatible parametrizations of the two-sided variety W.

EXAMPLE 7. Let V be the sphere $y_1^2 + y_2^2 + y_3^2 = 1$ of ordinary space and let G be $\langle (y_1, y_2, y_3); (-y_1, -y_2, -y_3) \rangle$. If then, e.g.,

$$x_1 = y_1^2, \quad x_2 = y_2^2, \quad x_3 = y_3^2, \quad x_4 = y_2 y_3, \quad x_5 = y_3 y_1, \quad x_6 = y_1 y_2,$$

the x's will not change if the y's change signs. This two-dimensional variety V of six-dimensional space is one-sided because the spherical substitution $(\phi, \theta); (\phi + \pi, \pi - \theta) \in G$ has functional determinant -1.

EXAMPLE 8. Let W be the $(2q-2)$-dimensional variety W of $2q$-dimensional space given by the equations $y_1^2 + \cdots + y_q^2 = 1$ and $z_1^2 + \cdots + z_q^2 = 1$; its points correspond to ordered pairs (Q, Q') of points of the hypersphere $y_1^2 + \cdots + y_q^2 = 1$. The $q(q+3)/2$ combinations

$$y_i + z_i, \qquad y_i z_i, \qquad y_i z_k + z_k y_i,$$

give us $n = q(q+3)/2$ new variables $x_1, x_2, \ldots, x_n$ which do not change if we interchange the y's and the z's, i.e. a variety V whose points correspond to unordered pairs $\{Q, Q'\}$ of points of the hypersphere S^{q-1}.

For $q = 2$ this V is not closed, however, for $q \geqslant 3$ it does have an empty boundary ∂V; furthermore V is one-sided for q even and two-sided for q odd (we omit the proofs given).

The *nonzero Betti numbers of W are* $b_0(W) = b_{2q-2}(W) = 1$ *and* $b_{q-1} = 2$. By duality it suffices to compute b_i, $i \leqslant q-1$. Any subvariety of dimension less than $q-1$ can be deformed into the ball $W \setminus (U_1 \cup U_2)$, where U_1, respectively, U_2 denotes all $(Q, Q') \in W$ such that $Q = Q_0$, respectively, $Q' = Q_0$. In dimension $q-1$, U_1 and U_2 are linearly independent: for, if J is the usual *volume element* of S^{q-1} in spherical coordinates, then $\int (J_1 + \lambda J_2)$ satisfies the conditions of integrability of § 7, and, for λ irrational, its periods over U_1 and U_2 are integrally independent. Lastly (we omit the proof given) any closed $(q-1)$-dimensional variety v of W is homologous to some $mU_1 + nU_2$.

On the other hand, *for* $q \geqslant 3$, *the nonzero Betti numbers of V are* $b_0(V) = b_{q-1} = b_{2q-2}(V) = 1$ (we omit the argument). This shows, as announced in § 9, that there exist one-sided, respectively, two-sided, varieties of dimensions $4k+2$, respectively, $4k$, having middle Betti number odd.

§ 16. *Théorème d'Euler.* This tells us that if S, A and F are, respectively, the number of vertices (sommets), edges (arêtes) and faces of a convex polyhedron, then one must have $S - A + F = 2$. This theorem has been generalized by M. l'amiral de Jonquières to nonconvex polyhedra. One now has $S - A + F = 2 - b_1$, where b_1 denotes the Betti number of the bounding surface. The fact that the faces are planar is of no importance, and the same result is true for any subdivision of a closed surface into *cellular* (simplement connexe) regions.

We shall generalize this result to an arbitrary closed variety V of dimension p. This will be subdivided into some varieties v_p of dimension p which are not closed, and the boundaries of these v_p's will be made up of some varieties v_{p-1} which are not closed,

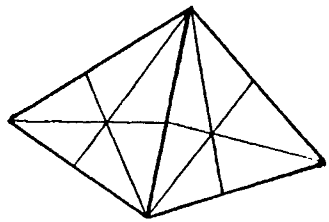


Fig. 6. Derived tetrahedron.

and so on till some points ν_0. If the regions ν_i's are all cellular, then we shall call such a subdivided V a *polyhedron*. We propose to calculate the number

$$N = \alpha_p - \alpha_{p-1} + \cdots \pm \alpha_0,$$

where α_q denotes the number of the ν_q's in the polyhedron.

Two polyhedra arising from the same V will be called *congruent*. Further, if the regions of the first are contained in those of the second, then the first will be said to be a *derived* of the second, e.g., we can derive a tetrahedron into 24 triangles as shown in Fig. 6.

We shall show that *the number N is the same for any two congruent polyhedra*. Since two congruent polyhedra have a common derived it suffices to show $N(P') = N(P)$ for any derived P' of P. If a ν_i, $i \leqslant q-2$, is incident to exactly 2 ν_{i+1}'s we shall say that ν_q is a *singular region* of the polyhedron. We allow these because then we can go (we omit the argument given) from P' to P, in a number of steps, done in order of increasing dimension, each involving erasing a ν_i having exactly two incident ν_{i+1}'s, followed by an annexation of these three regions. Clearly N is unchanged after each of these steps.

However, this argument is open to objections, e.g., during the above operations the regions may not remain cellular? Before modifying our proof so as to overcome these objections, let us compute some N's.

For the boundary V of any $(p+1)$-cell one has $N = 2$ if p is even and $N = 0$ if p is odd. By the above we can use, e.g., the boundary of a generalized tetrahedron (we omit this calculation), or a generalized cube $-1 \leqslant x_i \leqslant +1$, $+1 \leqslant i \leqslant p+1$: for the latter $\alpha_q = 2^{p+1-q}\binom{p+1}{q}$, so $(1-2)^{p+1} = 1 - \alpha_p + \cdots \pm \alpha_0 = 1 - N$, i.e. $N = 1 - (-1)^{p+1}$.

Our rigorous proof of the invariance of N will be by induction on p. The regions ν_i, $i > q$, incident to a given $\nu_q \in P$ will be said to constitute the *star* (aster) of ν_q. The induction hypothesis, and the above computation, imply that, *for the star of any $\nu_q \in P$ one has*

$$\gamma_p - \gamma_{p-1} + \cdots \pm \gamma_{q+1} = 1 + (-1)^{p-q-1} \tag{A}$$

where γ_t denotes the number of ν_t's in the star.

Next, let us take a *quadrillage*, i.e. a cubical subdivision of n-space by n pencils of nonaccumulating hyperplanes parallel to the coordinate planes, $x_i = a_{i,k}$, $1 \leqslant i \leqslant n$. Then, *if the mesh of this quadrillage is small, the intersection of each of its $(n-t)$-cubes D_{n-t} with V is a $(p-t)$-cell ν_t, and these cells give us a polyhedron Q covering V*. Let P' be a polyhedron which is a derived of P and of Q.

To see $N(P') = N(P)$ we go from P' to P by erasing the hyperplanes $x_i = a$, one by one. Let δ_q denote the number of q-cells of P' on this plane, δ'_q the number contiguous to it on the $(x_i < a)$-side of the plane, and δ''_q the number contiguous to it on the other side. Since $\delta'_q = \delta_{q-1} = \delta''_q$ (with also $\delta'_0 = 0 = \delta''_0$ and $\delta_p = 0$), the suppression of this hyperplane decreases each α_q by $\delta_q + \delta_{q+1}$, and since the alternating sum over q of these numbers is zero, N remains same.

To see $N(P') = N(Q)$ we can assume (by making the mesh small) that the interior c of each cell of Q intersects only *one* least dimensional cell v_q of P: thus the cells of P intersecting c are precisely those that have v_q on their boundary. We now go from P' to Q by erasing all cells of P' which are in p-cells c of Q but which have lesser dimension than p. So in each c we are erasing the least dimensional v_q and, for each $p > t > q$, γ_t incident cells of dimension t. Moreover, the number of p-cells within c was γ_p before and 1 after. Thus the total decrease in N is $-1 + \gamma_p - \gamma_{p-1} + \cdots \pm \gamma_{q+1} \mp 1$, which by the result above is zero. Next we erase all cells of P' which are in $(p-1)$-cells of Q but which have lesser dimension than $p - 1$, and so on. The same verification shows that N remains same at each step.

§ 17. *Cas où p est impair.* For any polyhedron P (subdividing V as in § 16) we shall denote by $\beta_{\lambda\mu}$ the sum, over all v_λ, of the number of v_μ's which are incident to v_λ. Note that $\beta_{\lambda\lambda} = \alpha_\lambda$ and $\beta_{\lambda\mu} = \beta_{\mu\lambda}$.

If the dimension p of V is even, the number N depends on the Betti numbers of V (see § 18) but *if p is odd, then a closed variety V always has* $N = 0$. To see this consider the following tableau

$$\begin{array}{llllll}
\beta_{p,p-1} - \beta_{p,p-2} & + \beta_{p,p-3} & - \beta_{p,p-4} & + \cdots \\
\quad + \beta_{p-1,p-2} & - \beta_{p-1,p-3} & + \beta_{p-1,p-4} & - \cdots \\
 & \quad + \beta_{p-2,p-3} & - \beta_{p-2,p-4} & + \cdots \\
 & & \cdots\cdots\cdots & \cdots\cdots \\
 & & & \cdots\cdots
\end{array}$$

The sum of the first row is the sum of the N's of the bounding $(p-1)$-spheres of the α_p p-cells of P, so it equals $2\alpha_p$. Likewise that of second row is zero and that of third is $2\alpha_{p-2}$, etc. Thus the sum of the tableau is twice $\alpha_p + \alpha_{p+2} + \cdots$. On the other hand the sums of the columns are $2\alpha_{p-1}$, 0, $2\alpha_{p-3}$, 0, ... by Eq. (A) of § 16. Thus the sum of the tableau is also twice $\alpha_{p-1} + \alpha_{p-3} + \cdots$. Equating the two values one gets $N = 0$.

§ 18. *Deuxième démonstration.* This proof will tell us how N depends on Betti numbers. I will first give an exposition of it for the *Case* $p = 2$, i.e. an ordinary polyhedron P with α_0 vertices, α_1 edges and α_2 faces, and show that $N = 2 - b_1$.

Assign to each of the α_0 vertices any number, and to each of its oriented α_1 edges the difference δ of the numbers of its two vertices. These α_1 numbers δ depend on the α_0 numbers, and conversely determine them up to an additive constant, so *there are in all* $\alpha_1 - \alpha_0 + 1$ *linear relations between the* δ*'s*. These linear relations are given by setting equal to zero, the algebraic sum of the δ's, of some cycle K of edges. Firstly, each of the oriented α_2 faces furnishes a cycle, viz. its perimeter Π. Secondly, from any chosen b_1 homologously independent cycles C of V, we construct cycles C'' of edges as follows – see Fig. 7.

We assert that any relation between the δ's is a linear combination of the $\alpha_2 + b_1$ relations given by the Π's and the C''s. To see this, let K be any cycle of edges. Adding a suitable linear combination of the C''s to it we get L, which is homologous to zero. Being a cycle

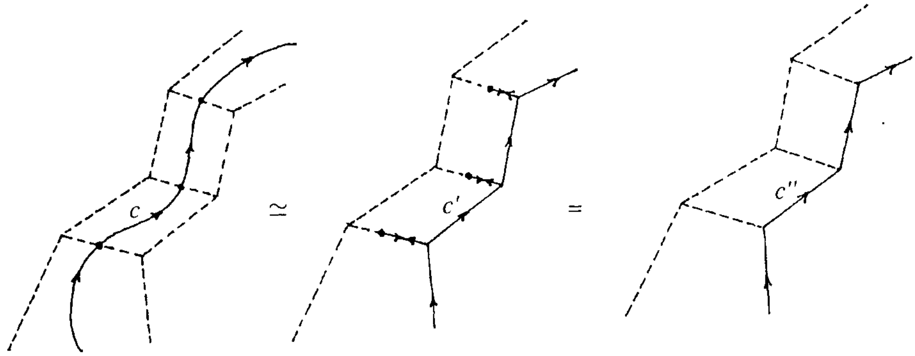


Fig. 7.

of edges of P, this L must be the boundary of a sum of faces of P, and so a sum of their perimeters Π. Also, the sum of *all* the oriented perimeters is zero, but no *partial sum* of the Π' is zero. Thus our new count shows $\alpha_2 + b_1 - 1$ linearly independent relations between the δ's. Equating this with $\alpha_1 - \alpha_0 + 1$ gives the required formula.

For any subdivision P of a closed p-dimensional variety V one has

$$\alpha_p - \alpha_{p-1} + \alpha_{p-2} - \cdots = b_p - b_{p-1} + b_{p-2} - \cdots .$$

This follows by a generalization (we omit the proof which is written out for the *Case* $p = 3$ only) of the above argument. Since the Betti numbers equidistant from the extremes are equal the above formula again shows that $N = 0$ when the dimension p is odd.

2. Notes

With reference to the Introduction of A.S.

NOTE 1. Starting with his dissertation [73], 1851, Riemann had visualized the graph of a multi-valued analytic function – e.g., the function y of x defined by a *polynomial equation* $f(x, y) = 0$ *in two variables* – as a surface obtained by gluing some complex "sheets" to each other along some "cuts", and in [74] he showed that the connectivity of this surface characterizes a nonsingular algebraic curve up to birational equivalence. This connectivity is defined on p. 11 of [73], and in [75] he left some ideas about a similar notion in higher dimensions; these higher connectivities were defined by Betti [4]. In 1880, Picard launched an analogous programme for *polynomial equations* $f(x, y, z) = 0$ *in three variables*, leading eventually to the famous treatise [49] on algebraic surfaces which he wrote with Simart.

NOTE 2. Poincaré developed his *qualitative theory of differential equations* in the three part memoir [51]. The index formula for generic vector fields is proved for all surfaces on pp. 121–125 of the second part – the 2-sphere case is there even in his C.R. note [50] of 1880 – while pp. 192–197 of the third part deal with the case of all n-spheres, $n \geqslant 3$. For Dyck's work on Analysis Situs see [19]. We note also that Poincaré's "A.S." was preceded by two Comptes Rendus notes [56, 58].

NOTE 3. The assertion that the (next to impossible) problem of *classifying the finite subgroups of* $GL(n, \mathbb{C})$ had been solved is of course wrong, however, Jordan [33] had given (modulo two groups of order 168 and 169 which he missed) the classification of all finite subgroups of $GL(3, \mathbb{C})$. The case $n = 2$ had been done previously by Klein [35]. The classification is now known for all $n \leqslant 7$: for references, and other information on this subject, see Dixon [17].

With reference to § 1 of A.S.

NOTE 4. *Poincaré's uses of the word "variety" have at least four modern connotations.* For instance a "closed variety" V, or rather its closure, can be thought of as a closed **pseudomanifold**, e.g., for $q \geqslant 4$, Poincaré's eighth example (§ 15) only gives pseudomanifolds. The "variety" M defined in § 3 as a "reseaux connexe" is more or less today's abstract closed **manifold**, while the special kind of "varieties" v used in its definition are the **local parametrizations** of M. In § 4 "varieties" are oriented, and then in §§ 5 and 6 Poincaré considers integral or rational linear combinations of the oriented varieties of an M to define its homologies and Betti numbers: in this context it is best to think of his "varieties" as **smooth chains** of M.

With reference to § 2 of A.S.

NOTE 5. This definition of Analysis Situs is in harmony with Klein's famous *Erlangen Program* [34] of 1872, even though now the "group" in question is really only a pseudogroup or a groupoid.

With reference to § 3 of A.S.

NOTE 6. The reader will note that Poincaré's "analytic continuation" works equally well with C^1 or C^∞ charts, and is just the way one would nowadays define an **abstract manifold** M, together with an immersion in n-space. Poincaré's focus will always be on the abstract M, he never enters into questions related to the immersion, and only exploits the convenience of n-space to present without fuss some important ideas whose simplicity is obscured if one insists – the book of Milnor [45] being a beautiful exception – on a totally intrinsic treatment.

The idea of an abstract manifold goes back to Riemann, but became popular only much later after Weyl [88], 1918.

With reference to § 5 of A.S.

NOTE 7. Interpreting ∂ as the oriented boundary of smooth oriented chains Poincaré's homologies are the same as those of Eilenberg [22]. It was probably because of this that Eilenberg remarks, on p. 408 of his 1942 paper [21] on **singular homology**, that the singular method of defining homology "*is as old as topology itself*".

NOTE 8. At this stage Poincaré's "just like equations" is confusing, for it is not clear that he allows division by nonzero integers, i.e. whether he wants to use integral or rational coefficients? However, this point gets clarified in the first "Complément".

With reference to § 6 of A.S.

NOTE 9. As Poincaré pointed out in [60], the numbers defined by Betti [4] himself were not the same as these! In modern terms Betti had considered *the number of elements required to generate* $H_r(M; \mathbb{Z})$, rather than the rank of the free part of this group: see the First Complement.

With reference to § 7 of A.S.

NOTE 10. Poincaré's indefinite integrals are uniquely determined by their skewsymmetric integrands or *differential forms* ω. This section of "Analysis Situs" inspired É. Cartan [12, 13]. Following him, the **exterior derivative** $d\omega$, of an r-form ω of a tubular neighbourhood U of a manifold $M \subset \mathbb{R}^n$, is the $(r+1)$-form defined by

$$(d\omega)(\alpha_1, \dots, \alpha_{r+1}) = \sum_i (-1)^i \partial/\partial x_{\alpha_i} \big[\omega(\alpha_1, \dots, \widehat{\alpha_i}, \dots, \alpha_4)\big].$$

One has $d \circ d = 0$, i.e. these differential forms constitute a *cochain complex* $(\Omega(U), d)$. Now Poincaré's "conditions d'intégrabilité" read $d\omega = 0$, and the result, nowadays called **Poincaré's Lemma**, which he quotes from his earlier paper, says that $H^*(\Omega(\mathbb{R}^n), d)$ vanishes in all positive dimensions. Given any r-form satisfying $d\omega = 0$, integration over cycles gives an additive group homomorphism $H_r(U; \mathbb{R}) \to (\mathbb{R}, +)$ (likewise $H_r(M; \mathbb{C}) \to (\mathbb{C}, +)$ if one uses complex valued forms) whose image is called the *period group* of ω. Poincaré checks that this free Abelian group has rank $\leqslant b_r(M)$, and asserts without proof that this bound is the best possible. For simple cases – like, e.g., $U = \mathbb{C} \setminus \{\text{some points}\}$ when one can use *Cauchy's integral formula*, also see Poincaré's use of the volume form J in § 15 – one can check this by giving explicit closed r-forms having $b_r(M)$ periods. The assertion is true in general, and equivalent to a generalization of the Poincaré lemma proposed by Cartan [13], which soon became **de Rham's theorem** [72], viz., the cohomology $H^*(\Omega(U), d)$ defined via differential forms is isomorphic to $H^*(M; \mathbb{R})$.

NOTE 11. As observed in Sarkaria [76], dropping the requirement that the components of ω be skewsymmetric with respect to the indices gives a bigger cochain complex $(\Omega_{\text{assoc}}(U), d)$ with d defined exactly as above, and furthermore, intermediate between $\Omega(U)$ and $\Omega_{\text{assoc}}(U)$ one has yet another, $(\Omega_{\text{cycl}}(U), d)$, consisting of all ω's skewsymmetric with respect to rotations of their indices. The cohomology of $(\Omega_{\text{assoc}}(U), d)$ is also $H^*(M; \mathbb{R})$, but the **cyclic cohomology** is somewhat different, being

$$H^*\big(\Omega_{\text{cyc}}(U), d\big) \cong \bigoplus_{j \geqslant 0} H^{*-2j}(M; \mathbb{R}).$$

Cyclic subcomplexes were first observed by Connes [14], however, the cyclic manner in which Poincaré displayed his "conditions d'intégrabilité" *could have* suggested $(\Omega_{\text{cyc}}(U), d)$ even to Cartan?

With reference to § 8 of A.S.

NOTE 12. *Orientability is not sufficient to ensure that a manifold can be defined as in* § 1. Such a $V^{n-p} \subset \mathbb{R}^n$ has a trivial normal bundle, so all its *characteristic classes* must vanish.

For example, since $p_1(\mathbb{C}P^2) \neq 0$, the complex projective plane, i.e. for $q = 3$, Poincaré's eighth example, cannot be so defined.

With reference to § 9 *of A.S.*

NOTE 13. Here *rational coefficients are necessary*, e.g., the double of a nonbounding 1-cycle V of Poincaré's fifth example $\mathbb{R}P^3$ (see § 10) bounds, so $N(C, V) = 0$ for all two cycles C. With rational coefficients it is true that a $(n-p)$-cycle is nonbounding iff its intersection number with some p-dimensional cut is nonzero. The proof which Poincaré gives of this assertion for the case $p = 1$ is okay, however, for $p \geqslant 2$ the sketched generalization is flawed: see the First Complement.

NOTE 14. Picard had used $b_1 = b_3$ for complex surfaces. Besides **Poincaré duality**, i.e. $b_p = b_{n-p}$ for closed oriented manifolds, the above assertion about cuts also implies *Lefschetz duality*, i.e. $b_{n-p}(\operatorname{int} M) = b_p(M, \operatorname{bd} M)$ for oriented manifolds with boundary. More generally, one has $b_{n-p}(M \setminus A) = b_p(M, A)$ if $M \setminus A$ is an orientable n-manifold, e.g., one has *Alexander duality* between the Betti numbers of a closed $A \subset S^n$ and its complement. This generalized *Jordan curve theorem* shows, e.g., that the Betti numbers of the example of § 6, i.e. of $S^3 \setminus$ {some bouquets of circles}, are indeed those given by Poincaré.

NOTE 15. Recall that two integral $n \times n$ matrices A and B are called *congruent* iff $A = PBP'$ for some $P \in \mathrm{GL}(n, \mathbb{Z})$. The unimodular **intersection matrix** $N(V_i, V_j)$ of size $b_{n/2}(M)$ which Poincaré considers is well defined up to congruence, and is especially important for $n = 4k$ when it is symmetric. For example, Whitehead [91] showed that the homotopy type of a closed simply connected 4-manifold is determined by the congruence class of this matrix, and a theorem of Donaldson [18] says that, if definite, such a matrix must be congruent to $\pm I$.

Combining this with Freedman [23] it follows, e.g., that there are about a 100 million distinct simply connected closed four-dimensional **topological manifolds** with $b_2 = 32$, and having intersection matrix – now of course defined via *cup products* – positive or negative definite, yet only two of these manifolds can carry a smooth structure! *Topological manifolds, or for that matter all of point set topology, came long after "Analysis Situs"*: unless explicitly stated otherwise, these notes also are about smooth manifolds and polyhedra.

With reference to § 10 *of A.S.*

NOTE 16. The assumption that manifolds are obtainable from polytope(s) by facet conjugations is equivalent to their **triangulability**. In § 16 Poincaré suggests a cell subdivision via "quadrillages", and in the first "Complément" (§ XI) he gives yet another with more details. For proofs of triangulability of smooth manifolds see Cairns [10, 11], and Whitehead [89]. For topological manifolds, triangulability is a *much* more delicate question, e.g., Casson has shown that some such closed 4-manifolds (related to Poincaré's homology 3-sphere) are not homeomorphic to any simplicial complex: see Akbulut and McCarthy [1, p. xvi].

NOTE 17. *There are in all seven orientable closed 3-manifolds obtainable by conjugating opposite facets of a cube*, these are listed in Sarkaria [77]. We note also that instead of Poincaré's star criterion one may simply check that the 3-complex resulting from the facet conjugations has Euler characteristic zero, then it will automatically – see Seifert and Threlfall [78, p. 216] – be a manifold. For *topological triangulations* stars can be funny, e.g., Edwards [20] gives a simplicial subdivision of S^5, in which one of the edges has as link the 3-manifold of Poincaré's "Cinquième Complément".

With reference to § 11 *of A.S.*

NOTE 18. Poincaré's sixth example amounts to identifying the 2 ends of $\mathbb{T}^2 \times [0, 1]$ using the diffeomorphism of the torus $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$ defined by $T \in \mathrm{SL}(2, \mathbb{Z})$. Starting instead with two copies of a solid torus, and identifying their bounding tori using T, one obtains the **lens spaces** L_T of Tietze [87]. For $T = \begin{bmatrix} 1 & 0 \\ h & 1 \end{bmatrix}$ this toral diffeomorphism commutes with the projections of the two solid tori onto the 2-disk, so yielding all the *circle bundles* $L_T \to S^2$: cf. Steenrod [84]. As against this, Poincaré's M_T's are $\mathbb{T}^2$-bundles over S^1, and in the "Troisième Complément" he will also consider other surface bundles over S^1.

NOTE 19. One can check that *the top and the bottom squares of* P_T *are each made up of exactly* $|\alpha| + |\beta| + |\gamma| + |\delta| - 1$ *facets.*

With reference to § 12 *of A.S.*

NOTE 20. Poincaré gives four approaches to his groups g and G. Firstly, as all *deck transformations* of a **covering space** over M, viz. that whose projection map is the inverse of the multiple valued function F_α (one should allow the number of values to be infinite also). Secondly, his differential equations definition – which plays a major rôle in Sullivan [85] – gives g as the **holonomy group** of a "curvature zero" or *integrable connection* on a vector bundle over M (for nonintegrable connections holonomy groups need not be quotients of $\pi_1(M)$). Thirdly, his definition using "loops", "equivalences" and "lacets" amounts to that which one usually finds in most text books. Lastly, in § 13, for any M obtained from a polytope by facet conjugations, Poincaré defines $\pi_1(M)$ via some simple and elegant (yet intriguing) cyclic relations.

Much later Hurewicz [31], 1935, defined his *higher homotopy groups* as fundamental groups of iterated *loop spaces*: $\pi_i(M) = \pi_1(\Omega^{i-1}M)$. That $\pi_3(S^2)$ is nontrivial was seen by using the *Hopf map* [29], i.e. the projection $L_T \to S^2$ (see Note 18) for case $T = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}$, when $L_T = S^3$.

With reference to § 13 *of A.S.*

NOTE 21. *It looks curious to a modern reader that Poincaré speaks of the "fundamental group" but never of the (first) "homology group"*, and this even though he speaks of "fundamental equivalences" in tandem with "fundamental homologies"! This is because – cf. [69, p. 450] – at that time, the word "group" was used in a more restricted sense: one spoke of groups of *transformations* (= substitutions = permutations etc.) but not of groups of *points*. For example, $\mathbb{R}$ equipped with addition was seldom called a group, but one spoke of the group of translations of $\mathbb{R}$. For equivalences, Poincaré had given such an interpretation via substitutions induced by monodromy, for homologies he had not. This undefinable distinction between transformations and points was discarded later, likewise function and path spaces entered topology.

With reference to § 14 *of A.S.*

NOTE 22. The theory of Fuchsian groups was created and developed by Poincaré: see, e.g., [52]. Examples (of all three kinds) of these discontinuous groups of motions of the *non-Euclidean plane* will occur later in course of the arguments of the third, fourth and fifth Complements. (This last also contains a more topological argument – via *Morse theory*! – for the classification of surfaces.) Poincaré also started work on the harder theory of discontinuous groups of motion of the *non-Euclidean space* – see, e.g., [53], also see [68, pp. 64–68] for his popular account of a non-Euclidean world – and probably got interested first in 3-manifolds while examining fundamental domains of these groups. The recent work of Thurston [86] shows that going back to these geometric "roots" may lead to a classification of 3-manifolds.

NOTE 23. As observed in Sarkaria [77] the main result of this section needs to be corrected slightly: *the groups G_T and $G_{T'}$ are isomorphic iff T or T^{-1} is in the same conjugacy class, in* GL(2, $\mathbb{Z}$), *as T'*. This is also then easily seen to be necessary and sufficient for the manifolds M_T and $M_{T'}$ to be diffeomorphic to each other.

Since $\mathrm{SL}(2, \mathbb{Z})/\{\pm I\}$ is isomorphic to a free product of $\mathbb{Z}/2$ and $\mathbb{Z}/3$, it follows that the *finite orders* ν which can occur are 1, 2, 3, 4 or 6. There is just one conjugacy class of SL(2, $\mathbb{Z}$) corresponding to each of these ν's, viz. those of

$$I,\ -I,\ \begin{bmatrix} -1 & -1 \\ 1 & 0 \end{bmatrix},\ \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix},\ \text{and}\ \begin{bmatrix} 0 & -1 \\ 1 & 1 \end{bmatrix},$$

respectively. The conjugacy classes of *parabolic elements* are also easy and are given by Poincaré: representatives are

$$\begin{bmatrix} 1 & h \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} -1 & h \\ 0 & -1 \end{bmatrix}, \quad h \in \mathbb{Z}.$$

However, it is not easy to make Poincaré's classification of the M_T's more explicit, because a *complete enumeration of the conjugacy classes of* GL(2, $\mathbb{Z}$) *is unknown*, but one does know that the number of hyperbolic conjugacy classes having a fixed trace t equals – see [77] – the **class number** of the real quadratic field $\mathbb{Q}[(t^2 - 4)^{1/2}]$. For a different connection between the topology of Poincaré's M_T's and number theory, read Hirzebruch–Zagier [26, pp. ix–xii].

NOTE 24. *One of the questions asked by Poincaré in* § 14 *can be answered by using the analogous manifolds L_T* (see Note 18) for which an explicit classification was found by Reidemeister [71], 1935: L_T is homeomorphic to L_T, iff $\gamma = \gamma'$ and either $\delta \equiv \pm\delta' \bmod \gamma$ or $\delta\delta' \equiv \pm 1 \bmod \gamma$. Here $\pi_1(L_T) \cong \mathbb{Z}/\gamma$. So one obtains nonhomeomorphic closed 3-manifolds having the same fundamental group. Indeed, by Whitehead [90], one also has nonhomeomorphic L_T's having the same homotopy type.

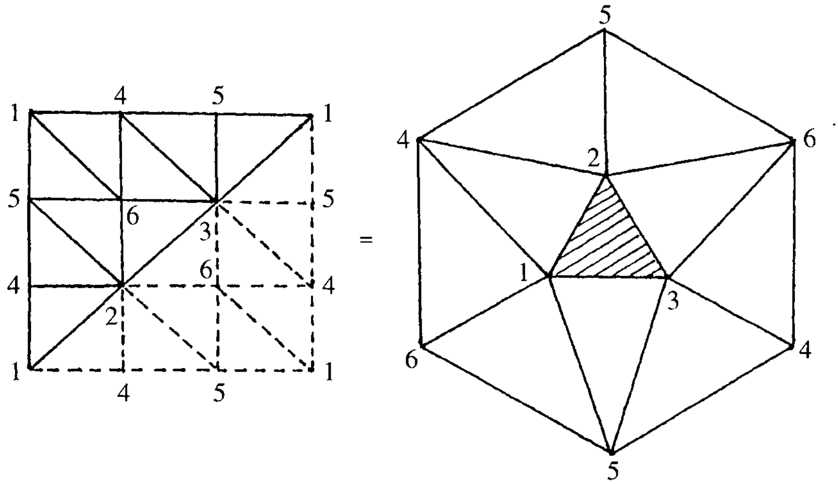


Fig. 8. Symmetric square of S^1.

With reference to § 15 *of A.S.*

NOTE 25. The image of the map of Example 7 is actually *contained in a* 4*-sphere* – this follows because $x_1 + x_2 + x_3 = 1$ and $x_1^2 + x_2^2 + x_3^2 + 4x_4^2 + 4x_5^2 + 4x_6^2 = 1$ – thus showing that $\mathbb{R}P^2$ embeds in the 4-sphere; more generally by Whitney [92] *any closed* M^n *embeds in* S^{2n}. We note also that Kuiper [38] has checked that the image of the analogous map $(y_1, y_2, y_3) \mapsto (y_1\bar{y}_1, y_2\bar{y}_2, y_3\bar{y}_3, y_2\bar{y}_3 + \bar{y}_2 y_3, y_3\bar{y}_1 + \bar{y}_3 y_1, \bar{y}_1 y_2 + \bar{y}_1 y_2)$, from the unit sphere of $\mathbb{C}^3$ to $\mathbb{R}^6$, *is equal to a* 4*-sphere*, thus showing that $\mathbb{C}P^2$ mod complex conjugation is S^4. See also Massey [43].

NOTE 26. For $q \geqslant 3$ the link of the diagonal points of the V of Example 8 is $S^{q-2} * \mathbb{R}P^{q-2}$, so for $q \geqslant 4$, V is only a pseudomanifold. For $q = 3$ one gets a manifold, viz. $\mathbb{C}P^2$. More generally $\mathbb{C}P^n$ *is diffeomorphic to the space of all unordered* n*-tuples of points of the* 2*-sphere* – see, e.g., Shafarevich [79, p. 402] – and likewise the symmetric n-th power of any 2-manifold is a $2n$-dimensional manifold. For $q = 2$, V is the Möbius strip, as is shown by the simplicial identifications made below. (See Fig. 8.)

Adding the shaded triangle to the Möbius strip gives the minimal triangulation $\mathbb{R}\mathrm{P}_6^2$ of $\mathbb{R}P^2$. Analogously, the minimal triangulation $\mathbb{C}P_9^2$ of $\mathbb{C}P^2$ – see Kühnel and Banchoff [37] – is close to the result of Kuiper and Massey mentioned in the last note.

With reference to § 16 *of A.S.*

NOTE 27. The reference for the cited work of Admiral de Jonquières is [32]. Incidentally I do not know of any higher-ranking topologist!

NOTE 28. Though in § 14 Poincaré gave the now current definition of **simply connected,** mostly he used it – see p. 275 of this section, or p. 297 of the first "Complément" – to mean a *cell* or, sometimes, its bounding *sphere*: e.g., while asking, on p. 498 of [69], the famous

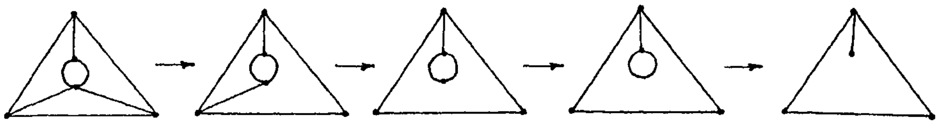


Fig. 9.

question which is now mistakenly called the *Poincaré conjecture*. The Euler–Poincaré formula is of course false if one only demands that the regions be simply connected in the modern sense of the phrase.

NOTE 29. The figure above (Fig. 9: here P is an unsubdivided triangle) shows that Poincaré's "erasing/annexing algorithm" may stop before reaching P.

Indeed *any* algorithm of this kind would give a positive answer to a problem which, to the best of my knowledge, is still open, viz. *is there a combinatorial characterization of the set of all simplicial complexes realizable as geometric subdivisions of a given simplicial complex* (cf. Hudson [30, p. 14])? However, a fundamental **theorem of M.H.A. Newman** [48] does identify the *equivalence relation* generated by "have a common geometric subdivision" with that generated by "have a common stellar subdivision". The invariance of N follows because clearly elementary stellar moves preserve it. This argument is close in spirit to the one being tried by Poincaré in his second attempt in this section.

With reference to § 17 *of A.S.*

NOTE 30. For **simplicial complexes** (these made their appearance in the first Complement) $\beta_{\mu\lambda} = \alpha_\lambda\binom{\lambda+1}{\mu+1}$ for all $\mu \leqslant \lambda$, so then column summation of Poincaré's "tableau" gives the **Dehn–Sommerville equations** [16, 82],

$$\alpha_p\binom{p+1}{\mu+1} - \alpha_{p-1}\binom{p}{\mu+1} + \cdots \pm \alpha_{\mu+1}\binom{\mu+2}{\mu+1} = \left(1 + (-1)^{p-\mu+1}\right)\alpha_\mu,$$

which, for a simplicial sphere, are equivalent to saying that the polynomial $\zeta(z) = \alpha_p z^{p+1} - \alpha_{p-1}z^p + \cdots \pm \alpha_0 z \mp 1$ must obey the *functional equation* $\zeta(z) = \zeta(1-z)$. A complete characterization of these polynomials is now known: see Stanley [83].

With reference to § 18 *of A.S.*

NOTE 31. This attempt – the two first "Compléments" will push it further – at the invariance of N is the one which affected future developments the most. It gives (implicitly) *a new definition of Betti numbers which uses a cell subdivision P of M* and Poincaré is trying to show – with ideas which clearly foreshadow **simplicial approximation** – that these coincide with those of § 6. This programme, in which Brouwer – see, e.g., [9] – played a big rôle, culminated in Alexander [2], 1915, which contains an elegant proof of $H_*(P) \cong H_*(M)$. After this it remains only to check, as Poincaré does, that the alternating sum of the face numbers α_i equals the alternating sum of the Betti numbers b_i of P. This lemma came to fruition with Hopf [28] and Lefschetz [41].

3. Complements

3.1. *The First Complement*

The First Complement opens with a reference to the "très remarquable" work of Heegaard [25], 1898, who had deemed the duality $b_p = b_{n-p}$ of "Analysis Situs" inexact and the proof of it given there without any value. Before examining Heegaard's specific objections Poincaré points out that **Betti's numbers** [4] were quite different from those of "A.S.". For Betti homologies were between *distinct* varieties together forming the boundary of some V, while Poincaré had considered arbitrary ∂V's. At this point Poincaré states that it is convenient to even allow division by nonzero integers (see Note 8). What an example given by Heegaard, or for that matter Example 3 of "A.S." itself, showed (see Note 13), is only that the *duality is false for Betti's numbers*, on the other hand the *duality is very much true for* **Betti numbers** (as defined in "A.S."), and the main object of this paper is to give a new proof of this using the polyhedra P of § 16 of "Analysis Situs".

As for the previous proof (§ 9 of "Analysis Situs"), after showing for $c = 1$ that a homologously nontrivial codimension c cycle V of an oriented closed manifold M admits a c-dimensional transversal C which intersects it nontrivially, Poincaré had hurriedly sketched that the general case could be done by finding an M' of one dimension more which contains V, then using case $c = 1$ to get a one-dimensional transversal cut C' in M', and finally enlarging C' to a complementary dimensional cycle C which intersects M' in C'. **Heegaard's objections** to this were two: how can one find M', or even if one can, how can one enlarge C' to a cycle C of the required kind? Poincaré admits the validity of at least the second of these objections.

Given a polyhedron P its **schema**, i.e. how it is built up from the v_i's, is determined by its *incidence numbers*: one has $\varepsilon_{ij}^q = 0$ if the j-th (oriented) $(q-1)$-cell is not incident to the i-th q-cell, and $= \pm 1$ otherwise, sign depending on whether or not the orientation of the $(q-1)$-cell agrees with the boundary orientation (§ 4 of "A.S.") of the q-cell. Poincaré observes the all-important necessary condition

$$\varepsilon^q \varepsilon^{q-1} = 0 \quad (\text{i.e. } \partial \circ \partial = 0),$$

but points out that this is not all, one has, e.g., the star condition of "A.S.", § 10. Poincaré poses the problem of characterizing schemas of manifolds (Newman's Theorem, Note 29, answers this partially).

Next, given a cell subdivision P of our manifold, we can consider the **reduced Betti numbers** $b_q(P) \leqslant b_q$ defined as the maximum number of linearly independent cellular cycles. Note that here "linearly independent" is still in the sense of § 6 of "A.S.", i.e. the homologies are not required to be cellular. However, Poincaré asserts that *all homologies are generated by the cellular ones*, and an intricate proof of this – only for the case of 3-manifolds – is given in Section VI. (As remarked in Note 31 a full proof, using Brouwer's simplicial approximation, was given much later by Alexander.)

However, before this he shows in Section III how the above assertion implies (this is along lines already sketched in § 18 of "A.S.") the **Euler–Poincaré formula** for the reduced Betti numbers,

$$\alpha_m(P) - \alpha_{m-1}(P) + \cdots = b_m(P) - b_{m-1}(P) + \cdots$$

(i.e. he checks that $\sum_i(-1)^i \dim C_i = \sum_i(-1)^i \dim H_i$ for any chain complex $\cdots \to C_i \to C_{i-1} \to \cdots$ over $\mathbb{Q}$).

In modern parlance, Section IV checks that, by imaging each q-cell of P to the sum of all the smaller compatibly oriented q-cells of a subdivision P', one gets the **chain subdivision map** $C(P) \to C(P')$.

In Section V he proves (again using the assertion that cellular homologies suffice) that the **reduced Betti numbers are subdivision invariants**. For this he deforms each q-cycle of P' over h-cells of P, in order of decreasing h, till finally it is contained in the q-skeleton of P. Then the coefficients of all smaller cells belonging to the same q-cell of P being the same, chain subdivision identifies it with a q-cycle of P. So $b_q(P)$ does not depend on P.

At the end of Section V, Poincaré asserts that given any closed (smooth) cycle one can always subdivide P so that the cycle becomes cellular in this subdivision. Using this **triangulability** assertion it follows that the **reduced Betti numbers coincide with those of "Analysis Situs"**, § 6. He enters into these intricacies in Section X (no use is now made of the "quadrillages" of § 16, "A.S.", instead there is an interesting idea involving *joins of simplicial complexes*) and declares at the end that "on est ainsi débarassé des dernier doutes" about triangulability. (We note that the "simple triangulation method" of Cairns [11] is almost the same as Poincaré's previous method of § 16, "A.S.", viz. intersecting M with a sufficiently fine "quadrillage".)

In Section VII Poincaré puts a vertex $\widehat{\sigma}$ in each $\sigma \in P$, and subdivides inductively by coning the already subdivided boundary of σ over $\widehat{\sigma}$. This gives a **simplicial complex**, viz. the *barycentric derived* P' of P. If one transfers the incidence relation amongst the cells of P to their barycentres, one sees that the simplices of P' have as vertices all totally orderable sets of barycentres, and that a cell σ of P consists of all simplices of P' having *highest* vertex $\widehat{\sigma}$.

Poincaré now defines his **dual cells** σ^* by inductively coning over $\widehat{\sigma}$ the already defined dual cells of the higher dimensional cells incident to σ (that σ^* is indeed a cell follows by the star criterion of "A.S.", § 10). So σ^* consists of all simplices of P' whose *lowest* vertex is $\widehat{\sigma}$.

The dual cells σ^* constitute the **polyèdre réciproque** P^* (Poincaré dual cell complex): note that P and P^* have common subdivision P', so just the subdivision invariance of Betti numbers gives $b_q(P) = b_q(P^*)$. We shall orient the dual cells so that, under the incidence reversing correspondence $\sigma \leftrightarrow \sigma^*$ between the schema of P and P^*, one has

$$\varepsilon_{ij}^q(P) = \varepsilon_{ji}^{n-q}(P^*)$$

(i.e. the boundary ∂ of P becomes the *coboundary* δ of P^*, thus giving at once the modern $H_q(P) \cong H^{n-q}(P^*)$).

Using this duality of incidences Poincaré obtains his **duality** $b_q(P) = b_{n-q}(P^*)$ by showing, in the course of Section VIII, that the reduced Betti numbers can be computed from the schema by using

$$b_q(P) = \alpha_q(P) - r\big(\varepsilon^q(P)\big) - r\big(\varepsilon^{q+1}(P)\big),$$

where $r(A)$ denotes the *rank* of the matrix A. (For the sake of simplicity Poincaré prefers to write all details, starting with the definition of P^*, only for 3-manifolds; however, the general versions can be found in §§ 1 and 3, of the Second Complement.)

Section VIII actually gives an **algorithm** which computes the Betti numbers of a schema. For this he sets up his **tableaux**

$$\begin{bmatrix} I_{\alpha_q} & \varepsilon^q \\ \varepsilon^{q+1} & 0 \end{bmatrix},$$

and then using *elementary row and column operations*, triangularizes the top right and bottom left corners. Working over $\mathbb{Z}$, instead of $\mathbb{Q}$ which suffices for *his* Betti numbers, enables him to show in Section IX that b_q *coincides with Betti's q-th number iff the greatest common divisor of the largest sized nonzero minors of the* $(q+1)$*-th incidence matrix is* 1.

Section IX also contains a cellular version of a result of § 9 of "Analysis Situs", viz. that *it is possible to find a p-cycle V_1 in P such that $N(V_1, V_2)$ is nonzero if and only if the $(n-p)$-cycle V_2 of P^* is not homologous to zero over* $\mathbb{Q}$. This is easy algebra because σ and σ^* have intersection number $N(\sigma, \sigma^*) = \pm 1$. Then, using triangulability, Poincaré again claims the previous results of § 9, "A.S.", in full.

3.2. *The Second Complement*

The Second Complement is only, says Poincaré, to simplify and clarify results already in hand. He begins by precising that, with dual cells oriented as above, one has

$$N(\sigma, \sigma^*) = (-1)^{q(q+2)/2}, \quad \text{where } q = \dim(\sigma).$$

In the previous paper Poincaré had only triangularized the corners of his "tableaux" because he was unaware of Smith [81], 1861, where it had been shown that *a rank r integer matrix can be reduced, by elementary operations over $\mathbb{Z}$, to a unique matrix of the type,* $\operatorname{diag}(d_1, d_2, \ldots, d_r, 0, \ldots, 0)$, $d_1 | d_2 | \ldots | d_r$. Still unaware of Smith's work, he now re-discovers, and gives a nice proof of this result in § 2.

In terms of these important new **torsion invariants** d_i of the schema P he then works out in § 3 that **Betti's q-th number exceeds b_q by the number of invariants of ε^{q+1} bigger than** 1, and that the product of these invariants gives the number of "distinct" cycles whose multiples bound (i.e. the order of the torsion part of $H_q(P)$).

As mentioned in Note 21 Poincaré did not *speak* of homology groups, but of course knowledge of Betti numbers and torsion invariants is equivalent to knowing homology or cohomology groups:

$$H_q(P) \cong \bigoplus \{\mathbb{Z}/d_i(\varepsilon^{q+1})\mathbb{Z} \colon d_i(\varepsilon^{q+1}) > 1\} \oplus \mathbb{Z}^{b_q},$$
$$H^q(P) \cong \bigoplus \{\mathbb{Z}/d_i(\varepsilon^{q})\mathbb{Z} \colon d_i(\varepsilon^{q}) > 1\} \oplus \mathbb{Z}^{b_q}.$$

Computations (§ 4). Poincaré gives the Betti numbers and torsion invariants of Examples 1, 3, 4, and 5 of § 10 of "Analysis Situs" by diagonalizing over $\mathbb{Z}$ the incidence matrices of the *cell complexes* (so ε^q_{ij} can be integers other than 0, ± 1) given by the facet conjugations (a more sophisticated cell complex H was used later for the homological computations of the Fourth Complement).

He also computes the same for Heegaard's [25] example, viz. **the singular link of the complex surface** $z^2 = xy$. Curiously, though it is all but apparent from the cell subdivision which he uses, he fails to notice that *Heegaard's example is actually the same as his own Example* 5 *of "Analysis Situs"*, i.e. diffeomorphic to $\mathbb{R}P^3$. (See Milnor [46] for more on singular links of complex hypersurfaces.)

Poincaré also computes his new invariants for the manifolds M_T of Example 6 (§ 11 of "A.S."). In modern terms he shows that $H_1 \cong \mathbb{Z} \oplus \mathbb{Z}/d_1(T-I)\mathbb{Z} \oplus \mathbb{Z}/d_2(T-I)\mathbb{Z}$ and $H_2 \cong \mathbb{Z}^{b_1}$. This time his method is to note (using § 13 of "A.S.") that H_1 is $\langle C_1, C_2, C_3 \rangle$ mod the relations

$$(\alpha - 1)C_2 + \gamma C_3 \simeq 0,$$
$$\beta C_2 + (\delta - 1)C_3 \simeq 0,$$

and the result follows by reducing the coefficient matrix.

In § 5 (which perfects Section X, and end of Section VII, of the First Complement) there is a direct combinatorial proof of the particular case $b_q(P) = b_q(P^*)$ of the invariance theorem.

In § 6 it is shown that *if the* q*-skeleton of* P *has no "one-sided circuits" in the sense of* § 8 *of "Analysis Situs", then its* $(q - 1)$*-th homology is torsion free.* This condition amounts to saying that if we consider any circuit, with some entries of the q-th incidence matrix as its vertices, and with edges alternatively horizontal and vertical, then the product of its vertices, if nonzero, is $+1$ or -1 depending on whether the length of the circuit is, respectively, 0 or 2 mod 4. (See Fig. 10.)

From this Poincaré deduces that all minors of ε^q must be 0 or ± 1 which of course implies that H_{q-1} is free. (We note that the vanishing of a *Stiefel–Whitney class* w_{n-q} can likewise be interpreted as a milder "orientability condition" on the q-skeleton.)

Poincaré ends by conjecturing that if Betti numbers and torsion invariants are all trivial then the manifold is a sphere. As is well known, he later disproved this via the famous example of the Fifth Complement, but it is to be noted that he already has at least examples of orientable 3-manifolds having the same homology groups but different fundamental groups, e.g., take M_T with $T = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ and $\begin{bmatrix} 2 & 1 \\ 3 & 2 \end{bmatrix}$.

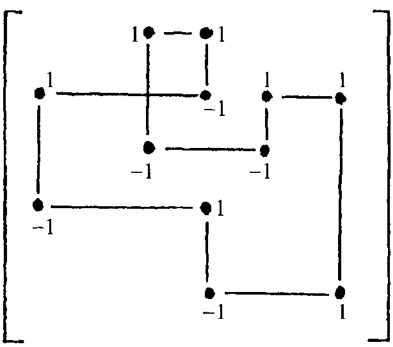


Fig. 10. A one-sided circuit.

3.3. *The Third Complement*

The Third Complement computes $\pi_1(V)$ of some V's satisfying the polynomial equation $z^2 = F(x, y)$. (In [64] Poincaré mentions that he got interested in these surfaces because of a problem regarding perturbation functions of celestial mechanics.) It is assumed that, but for finitely many *singular points* $\{A_1, \dots, A_q\}$ of the complex y sphere, $F(x, y) = 0$ has $2p + 2$ distinct roots $x_0(y), x_1(y), \dots, x_{2p+1}(y)$, $p \geqslant 1$.

First consider y as a constant $\neq A_i$, $1 \leqslant i \leqslant q$. Then $z^2 = F(x, y)$ gives a **complex curve** V of genus p, and the coordinates x and z of its points are *Fuchsian functions* of an auxiliary variable $u = \xi + \mathrm{i}\eta$, having a Fuchsian polygon R of the first type with angle sum 2π, such that opposite pairs of its $4p$ edges get identified under transformations $S_k(u) = \phi_k(\xi, \eta) + \mathrm{i}\psi_k(\xi, \eta)$, $1 \leqslant k \leqslant 2p$, generating the Fuchsian group G' admitted by these functions. The curve being *hyperelliptic* (it has the involution $(x, y, z) \leftrightarrow (x, y, -z)$) one can choose an R which admits a central symmetry (non-Euclidean, if $p \geqslant 2$) and is made up of two symmetric halves R' and R'', with R' being such that its $2p + 1$ vertices lie above x_0 and the mid-points of its edges correspond to the remaining roots $x_1, \dots, x_{2p+1}$ of $F(x, y) = 0$. Each of these tiles R' covers the complex x sphere with the two halves of each of its $2p+1$ edges imaging onto the two "lips" of a cut going from x_0 to $x_1, \dots, x_{2p+1}$ (the genus p surface is obtained by identifying two copies of this cut sphere). (See Fig. 11.) Each member of $\pi_1(V) = G'$ is a product of an even number of central symmetries s_k through points above x_k (e.g., $S_1 = s_1 s_{2k+1}$). These symmetries obey, besides $s_k^2 = 1$, the *Fuchsian relation* $s_0 s_1 \cdots s_{2p+1} = 1$.

What happens if we now let y vary and describe a simple closed curve in $S^2 \setminus \{A_1, \dots, A_q\}$? Our Fuchsian group will vary in a continuous way, likewise the roots $x_0, x_1, \dots, x_{2p+1}$, and the Fuchsian polygon R. After y has described the closed curve, the group will return to the original G', but the points x_i will in general get permuted amongst themselves, so that R might become a different, but still *equivalent* polygon R_1, i.e. still generating G'.

The three-dimensional variety V defined by $z^2 = F(x, y)$, *with y constrained on such a closed curve* is then analyzed via **monodromy**, i.e. as y varies we shall make the Fuchsian variable u vary continuously in such a way that vertices of the original tiling go to vertices of the new (but homeomorphic) tiling, edges to edges, and congruent points go to congruent points. We introduce three real variables ξ, η, and ζ: the first two being the real and imaginary parts of the initial u, and the last a function of y alone which augments by 1 as we describe the closed curve. We can then represent V (see "A.S.", § 11) by the discontinuous group G generated by the $2p + 1$ substitutions

$$(\xi, \eta, \zeta) \mapsto \big(\phi_k(\xi, \eta), \psi_k(\xi, \eta), \zeta\big),$$
$$(\xi, \eta, \zeta) \mapsto \big(\theta(\xi, \eta), \theta_1(\xi, \eta), \zeta + 1\big),$$

where $u_1 = \theta(\xi, \eta) + \mathrm{i}\theta_1(\xi, \eta)$ denotes final position of $u = \xi + \mathrm{i}\eta$ (so the 3-manifold V is the **mapping torus** of the diffeomorphism of the surface of genus p induced by the monodromy $u \mapsto u_1$). Note that $\pi_1(V) = G$ contains a normal subgroup isomorphic to G' which, together with the last substitution Σ, generates it. For the Euclidean case $p = 1$ we have

$$\theta(\xi, \eta) = \alpha\xi + \beta\eta, \quad \theta_1(\xi, \eta) = \gamma\xi + \delta\eta, \quad \text{where } T = \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} \in \mathrm{SL}(2, \mathbb{Z})$$

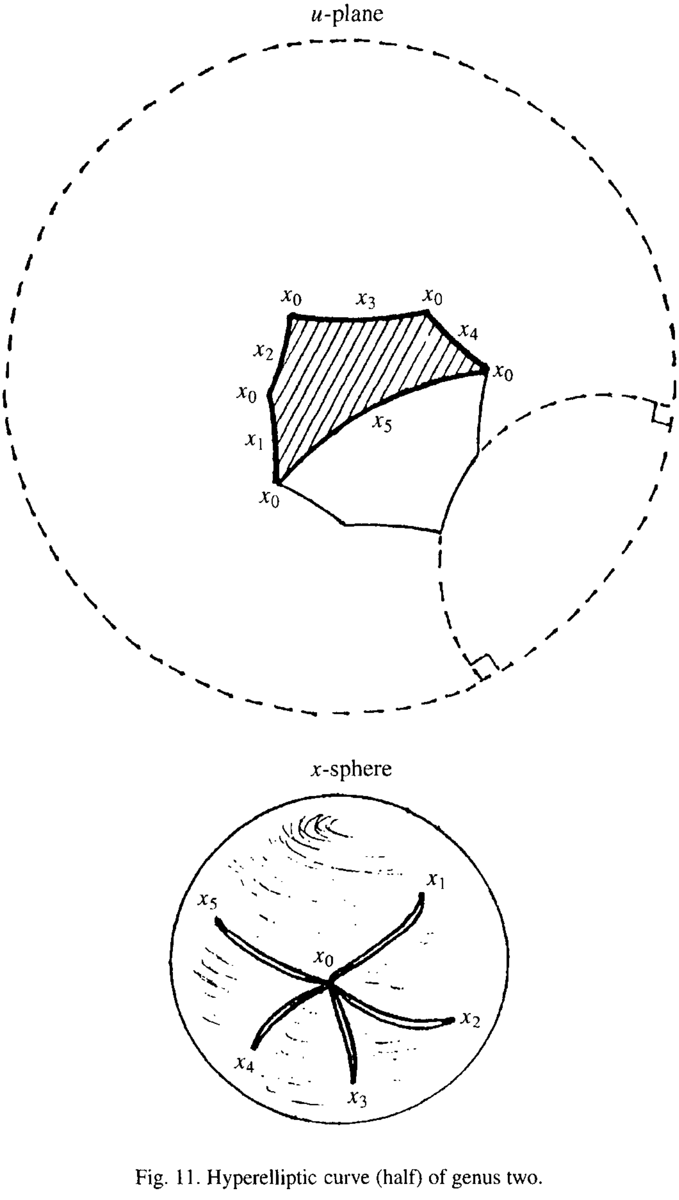


Fig. 11. Hyperelliptic curve (half) of genus two.

images the parallelogram R onto the equivalent R_1, so for $p = 1$ we have again the variety M_T and the group G_T of **Example 6** of "A.S.", § 11.

Let us consider next the four-dimensional variety V defined by $z^2 = F(x, y)$, with y only constrained to be outside q small circles guarding the singular points $\{A_1, \ldots, A_q\}$. To analyze V we shall join a chosen ordinary point O of the complex y sphere to these points by means of q disjoint cuts $OA_1, \ldots, OA_q$. Indeed we shall think of y as a Fuchsian function of a new auxiliary variable $\zeta + i\zeta' \in \Delta$, invariant with respect to a Fuchsian group Γ generated by a Fuchsian polygon Q of the second type whose q cusps α_i correspond to the

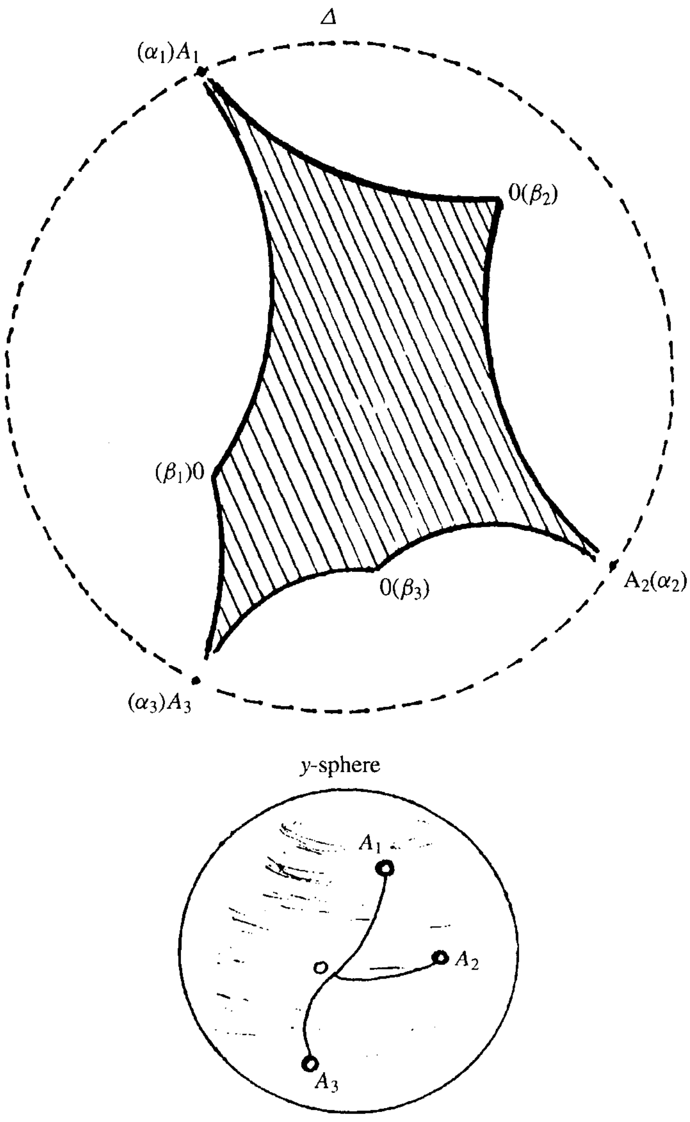


Fig. 12. A Fuchsian tile of the second kind.

singular points A_i, while its remaining vertices β_i all correspond to the ordinary point O. (See Fig. 12.) The sum of the tile's angles is again 2π (the angle at each cusp being zero) and the group Γ, i.e. the free rank $q-1$ fundamental group of $S^2 \setminus \{q \text{ points}\}$, is generated by q motions Σ_i which identity pairs of edges incident to the same cusp, and one has the relation $\Sigma_1 \Sigma_2 \dots \Sigma_q = 1$.

To each point of V assign four variables ξ, η, ζ and ζ', of which ξ and η are the real and imaginary parts of the u_0 above a base point y_0 obtained from u by monodromy over any path from y to y_0 which does not cross the cuts OA_i. This is independent of the path chosen, however, if y makes a loop around A_i resulting in the substitution $\Sigma_i(\zeta, \zeta') = (\kappa_i(\zeta', \zeta''), \kappa_i'(\zeta', \zeta''))$ of Γ, then the variables ξ and η can change, to say $\theta_i(\xi, \eta)$ and $\theta_i'(\xi, \eta)$. Our variety is thus represented by the discontinuous group G of 4-space determined by the $2p + q$ substitutions,

$$(\xi, \eta, \zeta', \zeta'') \mapsto (\phi_k(\xi, \eta), \psi_k(\xi, \eta), \zeta', \zeta''),$$
$$(\xi, \eta, \zeta', \zeta'') \mapsto (\theta_i(\xi, \eta), \theta_i'(\xi, \eta), \kappa_i(\zeta', \zeta''), \kappa_i(\zeta', \zeta''))$$

and this G is its fundamental group (Poincaré checks via the usual argument which identifies π_1 with all covering transformations of a simply connected cover). Note that it has G' as a normal subgroup which generates it together with the last q substitutions $T_i = (\theta_i, \theta_i', \kappa_i, \kappa_i')$. (That $\pi_1(V)$ is an extension of G' by the free group Γ can be seen also by using the *homotopy sequence* of V as a **fibration** over $S^2 \setminus \{q \text{ points}\}$ having the surface of genus p as its fiber.)

*Getting rid of the circles guarding the points A_i and supposing that x and y can take arbitrary complex values we now consider the **algebraic surface** V defined by $z^2 = F(x, y)$.* It will be assumed that as y approaches an A_i some two of the roots, say $x_a(y)$ and $x_d(y)$, approach a common value x_{ad}, but the other $2p$ roots all remain distinct, so the (possible) **singularities** of our V are $(x_{ad}, A_i, 0)$ only. Poincaré shows that V *is simply connected* (as against Picard who had shown $b_1(V) = 0$ for a generic complex projective surface V). For this note that $\pi_1(V)$ is a quotient of the above G. Also that, as y makes a small loop around A_i, while x remains constant, we get a small loop on V, so $T_i \simeq 1 \forall i$. With y moving as before, now let x also make a small loop around both $x_a(y)$ and $x_d(y)$. This augments the angle of $z^2 = F(x, y)$ by 4π, so giving us another small loop on V, this time around the singularity $(x_{ad}, A_i, 0)$. Thus $T_i s_a s_d \simeq 1$, so giving $s_a s_d \simeq 1 \forall a, d$.

*Lastly, let V be the **nonsingular part** of the above complex surface.* Since we can no longer deform past $(x_{ad}, A_i, 0)$ we cannot conclude $s_a \simeq s_d$ in the above manner. We, however, still have $T_i \simeq 1 \forall i$, so $\pi_1(V)$ is at most a quotient of G'. For $p = 1$, G' is Abelian, so then Picard's result implies that $\pi_1(V)$ *is finite*. Poincaré shows this in general by writing down some more relations using the fact that, in $\pi_1(V)$, the monodromy action of T_i must become the identity.

We illustrate this for $p = 2$, first *if $x_a(y)$ and $x_d(y)$ interchange* as y makes a small loop around A_i. Shown in Fig. 13 are the initial (full) and final (dotted) positions of cuts, from an ordinary point 0, to these two roots; the other four cuts do not change as y makes this loop. Now s_b (or just b for short) corresponds to a loop which intersects only one initial cut, viz. *Ob*. Observing in order the final cuts which this loop intersects we get $b \simeq dabad$. Likewise $a \simeq d$, $c \simeq dacad, d \simeq dad$, $e \simeq e$, and $f \simeq f$. So we again have $s_a \simeq s_d$. Indeed $(x_{ad}, A_i, 0)$ is a **removable singularity** of V, i.e. its link is S^3: this follows because, near it, our surface is like $z^2 = y - x^2$ near the origin. *If $x_a(y)$ and $x_d(y)$ remain distinct* then the picture can be as in Fig. 14. The same method now gives $b \simeq dadabadad$, $a \simeq dad$, $b \simeq dadabadad$, $c \simeq dadacadad$, $d \simeq dadad$, $e \simeq e$ and $f \simeq f$. So we only obtain $(s_a s_d)^2 = 1$. Now $(x_{ad}, A_i, 0)$ is a **conical singularity**: near it the surface is like $z^2 = y^2 - x^2$ (or *Heegard's example* $z^2 = xy$) and the link is $\mathbb{R}P^3$. From these considerations it is easy to see that

$$\pi_1(V) \cong (\mathbb{Z}/2)^{n-1} \quad or \quad (\mathbb{Z}/2)^{n-2},$$

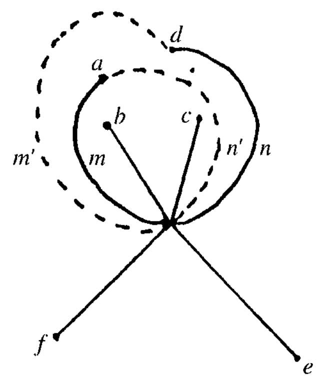


Fig. 13. Removable singularity.

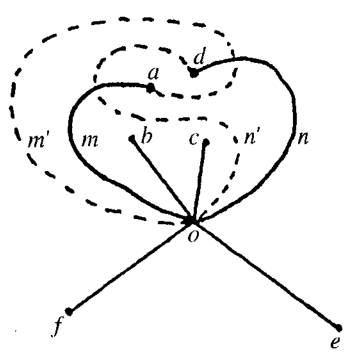


Fig. 14. Conical point.

where n denotes the number of irreducible factors of $F(x, y)$. Indeed if the roots a and b belong to the same factor, then we can identify the corresponding symmetries s_a and s_d, and otherwise identify them up to an ambiguity of order two, with the Fuchsian relation giving one more relation unless all factors are of even degree.

3.4. *The Fourth Complement*

The Fourth Complement opens with a mention of the pioneering "beaux travaux" of Picard, and goes on to show how monodromy can be used to find all the Betti numbers of a smooth complex two-dimensional variety V (Poincaré works over $\mathbb{Q}$ but it is asserted in [65] that this method will also give the torsion invariants). We shall suppose V represented as an algebraic surface $f(x, y, z) = 0$ having only "ordinary singularities", such that for each fixed $y \neq A_1, \ldots, A_q$ our equation determines a smooth complex curve $S(y) \subset V$ of constant genus p, but the genus of the q exceptional curves $f(x, A_i, y) = 0$ can be lower.

A cell subdivision H of V(§ 1): this projects, on the y-sphere, to a $2q$-gon Q with pairs of sides $\beta_i\alpha_i \equiv \beta_{i+1}\alpha_i$ covering the two lips of the cuts OA_i (see a picture above), and induces, for each fixed $y \notin S^2 \setminus \{\text{cuts } OA_i\}$, a subdivision P of $S(y)$ which is preserved by monodromy over paths not crossing the cuts. Further, as y approaches a point M on a cut OA_i from its two sides, P can approach two quite different subdivisions, MP and (MP), of the same Riemann surface $S(M)$: we shall assume that now H induces the common refinement P' of MP and (MP). Likewise, as y approaches O from within any of the q sectors, P can tend to q different subdivisions of $S(O)$: now H induces the common refinement P'' of all q of these. Finally, as M approaches A_i along OA_i, MP and (MP) approach coincidence, and at the same time some cells get identified, to give the cells of H covering the lower genus curve $f(x, A_i, z) = 0$. The faces, vertices, and edges of P (respectively P', respectively P'') are denoted F_i, B_j, C_k (respectively F'_i, B'_j, C'_k, respectively F''_i, B''_j, C''_k), and those of H by pre-multiplying these with the appropriate faces of Q.

Computation of $H_3(V)$ (§ 2). We sketch below the argument which is given to show (in modern terms) that $H_3(V)$ is *isomorphic to the subgroup of* $H_1(S)$ *which remains fixed under the action of the* **Picard group** (so named in [65] by Poincaré), i.e. the image of the monodromy induced group homomorphism $\pi(S^2 \setminus q\text{pts}) \to \text{Aut}(H_1(S)) \cong \text{GL}(2p, \mathbb{Z})$. (Also, the parity of $b_3(V) = b_1(V)$ is always *even*, being double the **irregularity** of V, as was shown by Picard using transcendental methods.)

Let $\omega = \sum_j c_j QB_j + \sum_{k,i} c_{ki}\alpha_i\beta_i F'_k$ be any 3-cycle of H, then – look at terms of left side of $\partial\omega = 0$ involving cells with first factor $Q - \Omega = \sum_j c_j B_j$ must be a 1-cycle of P, and it is easily seen that $\omega \simeq \omega'$ implies $\Omega \simeq \Omega'$. Also – look at the remaining terms of $\partial\omega = 0 - \sum_j c_j\alpha_i\beta_{i+1}B_j - \sum_j c_j\alpha_i\beta_i B_j$ is a boundary $\forall i$, which implies, on intersecting with $S(M)$, that the copies of Ω in the subdivisions MP and (MP) of $S(M)$, $M \in OA_i$, must be homologous (in P' after subdivision). In other words the 1-cycle Ω of P is **invariant** (up to homology) under monodromy. Further if Ω bounds then so must ω: to see this note that now we can add a boundary to ω to get a 3-cycle of the type $\sum_{k,i} c_{ki}\alpha_i\beta_i F'_k$, but then it has to be zero, for otherwise, for some i, we are saying that the fundamental cycle of $S(M)$, $M \in OA_i$, goes to 0 as M approaches A_i. Poincaré also checks that every invariant 1-cycle Ω arises from a 3-cycle ω in the above way. For this purpose he chooses on $S(M)$ *a region* R bounded by Ω and $T_i(\Omega)$ – here T_i denotes monodromy about A_i –which approaches 0 as M approaches A_i. Using this it follows that the boundary of $\omega = Q\Omega + \sum_i \alpha_i\beta_i R_i$ contains at most terms involving cells with first factor β_i. *This 2-cycle* $\partial\omega$ *cannot cover all of* $S(0)$ *and so must be zero.* To see this note, because of our choice of R, that it covers the area of $S(O)$ "swept out" by $\Omega \subset S(y)$, monodromed back to 0, as y describes the flower shaped contour of Fig. 15. Our monodromy (cf. Third Complement) results from the movement of the *branch points* $x_k(y)$ – i.e. the common roots of $f(x, y, z) = 0 = \partial f/\partial z$ – with y, so our sweeper curve is at all times "fleeing" away from these moving branch points, and thus $\partial\omega$ cannot cover all of $S(0)$.

As this sketch indicates the argument depends heavily on the nature of H above the points A_i. In § 5 Poincaré elaborates on this by giving two examples: in both cases the subdivisions MP and (MP) are exhibited, the invariant 1-cycle Ω and the aforementioned "vanishing region" R explicitly given, and (in §§ 3 and 4 this is assumed in general) it is shown that a 1-cycle **vanishing** at A_i is necessarily of type $\Omega - T_i(\Omega)$.

Computation of $H_2(V)$ (§ 3). Using arguments similar to those sketched above a complete list of homologically distinct 2-cycles of H is displayed. Their number $b_2(V)$ is given (there are some misprints here) in terms of the numbers of homologically distinct invariant and vanishing cycles of S and the rank of a certain matrix defined using monodromy (the

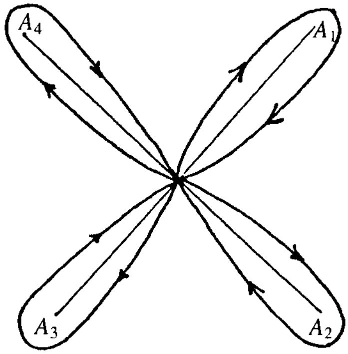


Fig. 15.

correct computation – see [40, p. 40] – identified the **Zeuthen–Segre invariant** of V with its Euler characteristic).

Computation of $H_1(V)$ (§ 4). As Poincaré points out $b_1(V)$ had already been computed by Picard who had shown (in modern terms) that $H_1(V)$ *is isomorphic to* $H_1(S)$ *mod the subgroup generated by the vanishing cycles* $\Omega - T_i(\Omega)$ (cf. a similar result, about π_1 of the smooth part of a surface, in the Third Complement). Poincaré gives another proof of this statement using arguments like those sketched above. Then, using the skewsymmetric intersection form of the genus $2p$ Riemann surface S, he shows that *the sum of the numbers of homologically distinct invariant and vanishing cycles of* S *is* $2p$, thus verifying Poincaré duality $b_1(V) = b_3(V)$ for the orientable smooth 4-manifold V. (In fact a "stronger Poincaré duality" holds for V, viz. the so-called **hard Lefschetz theorem** [40, p. 29]: *there exists a basis of* $H_1(S)$ *represented by cycles which are either invariant or vanishing*. The arguments of Lefschetz's book, which also contains generalizations for smooth complex projective varieties V of arbitrary dimension, are like those sketched above. However – see Lamotke [39] – a complete topological proof of this stronger duality still remains elusive, the best proof being via *Hodge theory*. Incidentally – see [40] – these transcendental methods were also pioneered by Picard and Poincaré.)

3.5. *The Fifth Complement*

The Fifth Complement is mostly about 2- and 3-manifolds but the method used (now called **Morse theory**) is, as Poincaré puts it, "sans doute d'un usage plus général". (For example, Morse [47] and Lusternik and Schnirelmann [42] generalized this method to *path spaces*, furnishing the tool used by Bott [6] to compute $\pi_i(U(n))\forall i < 2n$.)

In § 2 Poincaré sections any smooth $(m+1)$-dimensional manifold $V \subset \mathbb{R}^k$ into m-dimensional subvarieties $W(t)$ by means of a one-parameter family of real hypersurfaces $\phi(x_1, \ldots, x_k) = t$. In general $W(t)$ has no singularities, but for finitely many values t_0 of t it is allowed to have one **singular point**. Poincaré notes that *the diffeomorphism type of* $W(t)$ *changes only when* t *crosses an exceptional value* t_0. If, near its singular point, the section $W(t_0)$ looks like say $\phi_1(y_1, \ldots, y_{m+1}) = 0$ (we shall take $\phi_1 = \phi - \phi(t_0)$) near the origin, then *we can always assume, after perturbing* ϕ *slightly if need be, that the* *second degree terms of ϕ_1 give a* ***nondegenerate*** *quadratic form.* Choosing coordinates which diagonalize this quadratic form, we see thus that near its singularity $W(t_0)$ is, for some $0 \leqslant q \leqslant m+1$, like the hypersurface

$$y_1^2 + \cdots + y_q^2 - y_{q+1}^2 - \cdots - y_{m+1}^2 = 0,$$

near the origin (so $\lambda = m+1-q$ is the **index** of the singularity). *When $q = 0$ or $q = m+1$ the* **singular link** *C of $W(t_0)$ is empty, otherwise it is diffeomorphic to $S^{q-1} \times S^{m-q}$*: this follows because C is given by the above equation and $|y_1|^2 + \cdots + |y_{m+1}|^2 = 1$. (Note that Poincaré had used a similar method even in the last two Complements, viz. sectioning a complex variety by a pencil of hypersurfaces depending on a complex parameter y. For this *holomorphic Morse theory* a singular link is given by the complex equations $z_1^2 + \cdots + z_n^2 = 0$ and $|z_1|^2 + \cdots + |z_n|^2 = 1$, and thus is the *tangent sphere bundle of a sphere*: see Lamotke [39, p. 37], for the rôle which this fact plays in this theory.)

Each $W(t)$ can have many components $w_i(t)$. Poincaré defines the **squelette** (a graph in 3-space) of V by collapsing each $w_i(t)$ to a single point. If $q = 0$ or $m+1$ then one is on a **cul-de-sac**, and if w_i splits into two (or vice versa) as we move past this t, on a **bifurcation** of the squelette. In general there are also other singular values of t which too are marked appropriately on the squelette.

SURFACES V. *Now any singularity must be a cul-de-sac or a bifurcation.* To see this let $W(0)$ have a singularity with $q = 1$ – so C consists of 4 points – near which it is the union of the intersecting arcs 13 and 24. If 1 were **associated** to 3, i.e. joinable to it in $W(0)$ without passing the singularity, then 2 must be associated to 4. Now there is no bifurcation (see Fig. 16 which shows a part of V, which we think of as a polygon with pairwise conjugation of its boundary edges) but V would be *one-sided* (for AB gets conjugated to

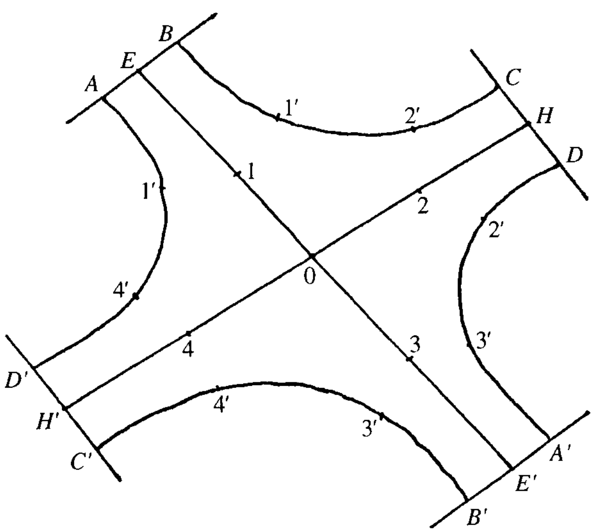


Fig. 16. A one-sided singularity.

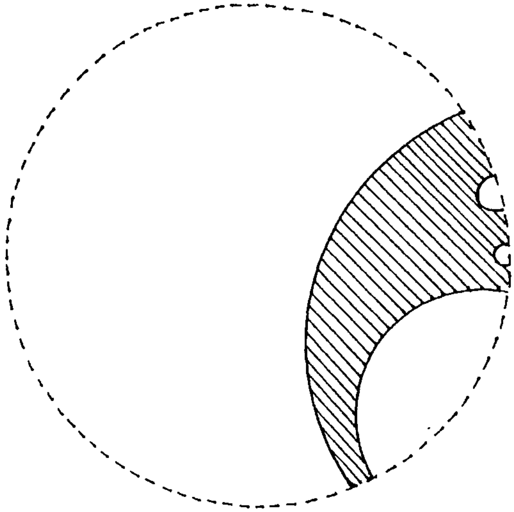


Fig. 17. Half of a Fuchsian polygon of the third kind.

$A'B'$). So 1 can only be associated to 2 or 4, and so 3 to 4 or 2, respectively, and in either of these cases we get a bifurcation of the squelette.

In § 3 this is used to sketch *a Morse theory proof of the classification of surfaces*. Choose p points on the squelette whose removal would get rid of all its circuits but keep it connected. From the above discussion one can deduce that if we cut V along the $w_i(t)$'s corresponding to these p points then we would be left with *a planar region R bounded by $2p$ circles*. The uniqueness of this model follows because clearly $p = b_1(V)$. One may think of R as one of infinitely many congruent **Fuchsian polygons of the third kind** tiling the plane, with conjugations realized via elements of the Fuchsian group. (See Fig. 17.) Another model of V is a **normal polygon** R' (geometrically a Fuchsian tile of the first kind) of $4p$ sides: e.g., for $p = 2$ it is an octagon 12345678 with boundary identifications giving the sole equivalence $C_1 + C_2 - C_1 - C_2 + C_3 + C_4 - C_3 - C_4 \equiv 0$ between the **fundamental cycles** $C_1 = 12,\ C_2 = 23,\ C_3 = 56,\ C_4 = 67$. For $p = 2$ (see Fig. 18) one can go (§ 4) from R to R' by **cutting** the region between DMD and $-B$ and **pasting** it to $+B$. (An algorithm for **normalizing** any polygonal representation of V was given by Brahana [17]; in many text books the classification of triangulated surfaces is proved via some such algorithm.)

ORIENTABLE 3-MANIFOLDS V. If $w(0)$ has a singular point other than a cul-de-sac, the singular link C is a union of two disjoint circles. We note (see Fig. 19) that the **throat** ("ellipse de gorge") K of $w(+\varepsilon)$ shrinks (under the **gradient flow** of the Morse function) to the singularity 0 as t decreases to 0 and then disappears. In case the two circles of C are not in the same component of $w(0) \setminus 0$, then K disconnects $W(+\varepsilon)$ and so is a boundary, *now there is bifurcation but $w(+\varepsilon)$ and $w(-\varepsilon)$ have the same b_1*. In case the two circles of C are in the same component of $w(0) \setminus 0$, then *there is no bifurcation but the b_1 of $W(+\varepsilon)$ is 2 more than that of $W(-\varepsilon)$*. A reduction by 1 occurs because the throat K, which is now

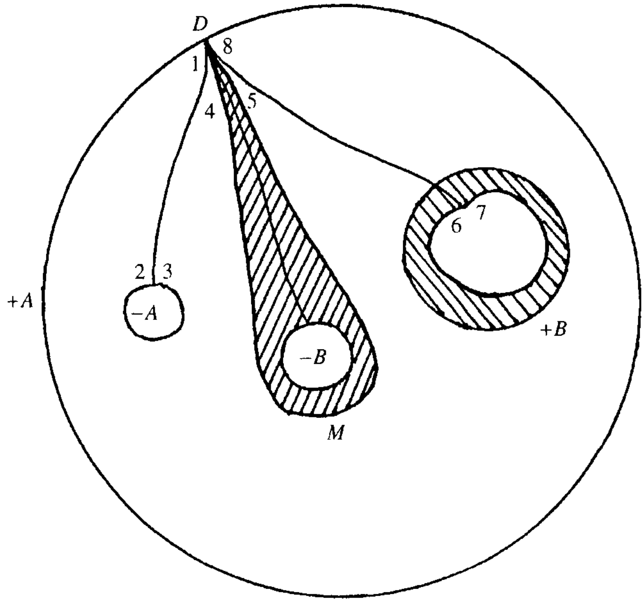


Fig. 18. Cutting and pasting.

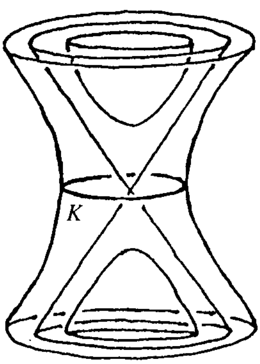


Fig. 19. Throat.

homologically nontrivial, disappears, and by another 1 because a cycle C of $W(+\varepsilon)$ with $N(C, K) \neq 0$ also disappears. There is no further reduction because if $N(C_1, K) = k_1$ and $N(C_2, K) = k_2$, then $k_2C_1 - k_1C_2$ is homologous to a cycle not cutting K, and so cannot disappear. (One obtains $W(+\varepsilon)$ from $W(-\varepsilon)$ by doing a **surgery** of type λ, or equivalently $W(\leqslant +\varepsilon)$ from $W(\leqslant -\varepsilon)$ by attaching a **handle** of index λ.)

To motivate the questions which Poincaré tackles next in §§ 3 and 4 we note that in § 5 he is going to fix (via the gradient flow), for each singular value t_q a copy of its throat on $W(t)\forall t > t_q$. Thus one needs to look at systems of *non self intersecting* ("non bouclé") cycles K_i of $W(t)$ which *do not intersect each other.* (In higher dimensions too, to simplify a Morse function, one needs to analyse systems of *spherical cycles*, of at most half dimension, on the generic level surfaces.)

For example, he checks in § 3 that *a 1-cycle of a surface W is homologous to a non self intersecting cycle iff it is a combination, with relatively prime coefficients, of the fundamental cycles*, i.e. iff it represents a **primitive element** of $H_1(W) \cong \mathbb{Z}^{2p}$. This is deduced as a corollary of a theorem which, in modern terms, says that **the map** $\mathrm{Diff}(W) \to \mathrm{Aut}_F(H_1(W))$, $f \mapsto f_*$, **is surjective**. Here $\mathrm{Aut}_F(H_1(W))$ consists of all automorphisms of $H_1(W)$ which preserve the **intersection form**. Recall that *any integral skewsymmetric matrix having determinant* 1 *is congruent over* $\mathbb{Z}$ *to*

$$F = \mathrm{diag}\left(\ldots, \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \ldots\right).$$

Choosing such a basis $\mathrm{Aut}_F(H_1(V))$ is same as $\mathrm{Symp}(2p, \mathbb{Z})$, i.e. all $A \in GL(2p, \mathbb{Z})$ such that $AFA' = F$. Poincaré lists some matrices over $\{-1, 0, +1\}$ which he asserts – this was verified by Brahana [8] – generate $\mathrm{Symp}(2p, \mathbb{Z})$. The theorem is proved by a long cutting and pasting argument which shows that these generators of $\mathrm{Aut}_F(H_1(W))$ arise from diffeomorphisms of W.

In § 4 Poincaré deals with some analogous questions for equivalences, e.g., *when is a given cycle of the surface equivalent to one which is non self intersecting*? Considering W as Δ/G, where $G \cong \pi_1(W)$ is a Fuchsian group of the first kind, he lifts the given cycle C to an arc of Δ going from say M to SM, $S \in G$, and denotes by α, $\beta \in \partial\Delta$ the two fixed points of this hyperbolic transformation S. He shows that *C is improperly equivalent to a non self intersecting cycle iff the non-Euclidean line $\alpha\beta$ does not intersect the corresponding line $\alpha'\beta'$ of any conjugate S' of S*. Here improper equivalence $A \equiv B$ (impr.) means that base point can move (i.e. the loops are **freely homotopic**). Poincaré points out that $A + B + C \equiv B + C + A$ (impr.), so now *cyclic reordering* is allowed, as against equivalences when no reordering may be valid, or as against homologies when all reorderings are valid.

He also gives *a rule to check if a combination of the fundamental cycles of W is equivalent to a non self intersecting cycle.* The complicated details are written out only for $p = 2$ for which case it shows, e.g., that *of all the combinations involving C_1 and C_3, only C_1, C_3, $C_1 + C_3$ and $C_3 + C_1$ are equivalent to non self intersecting cycles.* (As against this any $aC_1 + bC_3$ with $(a, b) = 1$ was homologous to a non self intersecting one; this anomaly between homologies and equivalences disappears when one uses Morse theory in dimensions $\geqslant 5$.)

The next § 5 examines an orientable 3-manifold V (with boundary $W = W(1)$) generated by connected $W(t)$'s, $0 \leqslant t \leqslant 1$, with p exceptional t_i's, at each of which b_1 increases by 2. The q-th throat fixes a non self intersecting cycle K_q on each $W(t)$ with $t > t_q$ and these cycles K_q of $W(t)$ do not intersect each other. As t increases from t_q each K_q sweeps out a ball B_q around the q-th singularity, whose final position at $t = 1$ is called A_q. Two parallel disjoint 2-balls B'_q and B''_q (which approach coincidence as t approaches 1) are then taken on either side of B_q and we denote by K'_q and K''_q their intersections with $W(t)$. We cut from $W(t)$ the small area S_q between K'_q and K''_q and paste to these two circles the 2-disks B'_q and B''_q. This new surface $W_1(t)$ *is a 2-sphere for all t bigger than* 0. To see this note that cutting out the S_q's from $W(t)$ gives a planar region R bounded by some circles and by pasting the disks we have filled in all the holes including that of the outer circle. The variety U generated by $W_1(t)$'s is thus a 3-ball U with $2p$ **scars** ("cicatrices") on its boundary, the 2 lips of the **cut** A_q which have to be identified in pairs to make V. It follows that V *is diffeomorphic to the genus* p **handlebody**, i.e. the region bounded by a genus p surface embedded in 3-space, and that this is independent of the embedding of the surface (i.e. that surfaces do not **knot** in 3-space).

Poincaré checks that any cycle of V is equivalent to one on W and that any equivalence of V is a consequence of $K_1 \equiv 0, \ldots, K_p \equiv 0$ (i.e. that $\pi_1(V)$ is the free group on p generators). Also he checks that p non self intersecting cycles $K'_1, \ldots, K'_p$ of W, which do not intersect each other, can arise in the above way only if they are equivalent to a combination of conjugates of the cycles $K_1, \ldots, K_p$ (alternatively cutting along them should give a planar region bounded by $2p$ circles).

The final § 6 considers an orientable 3-manifold V generated by connected $W(t)$'s, $0 \leqslant t \leqslant 1$, with $2p$ exceptional values of t; at the first p of these, which lie in $(0, 1/2)$, b_1 increases by 2, and at the remaining p, which lie in $(1/2, 1)$, it decreases by 2. Our V thus decomposes into two handlebodies V' and V'', the first over $[0, 1/2]$, the other over $[1/2, 1]$. The manifold is determined by the genus p surface $W = W(1/2)$ together with the two systems of **principal cycles** $K'_1, \ldots, K'_p$ and $K'_1, \ldots, K''_p$ of these handlebodies. (*Every* 3*-manifold admits such a Morse function*, i.e. a **Heegaard decomposition** into two handlebodies of some genus p. The least such p is called its *Heegaard genus*, and a two-dimensional description of the kind mentioned a *Heegaard diagram* of V: see [25]. A manifold has Heegaard genus 1 iff it is one of the L_T's of Notes 18 and 24, but *classification is unknown for any Heegaard genus* $\geqslant 2$.)

Poincaré shows that any cycle of this closed 3-manifold V is equivalent to one lying on W and that any equivalence is a consequence of the obvious equivalences $K'_1 \equiv 0, \ldots, K'_p \equiv 0$, and $K''_1 \equiv 0, \ldots, K''_p \equiv 0$ (this determines $\pi_1(V)$). Writing the principal cycles as combinations of the fundamental cycles and reordering one gets the homologies

$$m'_{i,1}C_1 + \cdots + m'_{i,p}C_{2p} \simeq 0,$$
$$m''_{i,1}C_1 + \cdots + m''_{i,p}C_{2p} \simeq 0,$$

which determine the Betti number and torsion coefficients of V. So these are the same as a 3-sphere, i.e. V is a **homology sphere**, *iff the* $2p \times 2p$ *determinant formed by the above integer coefficients is* ± 1. However, as the example below shows this need not be a *homotopy sphere*.

Poincaré defines his homology 3*-sphere via a Heegaard diagram*: $p = 2$ and W is represented as a planar region R bounded by four circles, he takes $K'_1 = C_1$, $K'_2 = C_3$ while K''_1 and K''_2 are given, respectively, by the unions of the full and dotted segments. (See Fig. 20.)

He computes using the above method to see that $\pi_1(V)$ is generated by C_2 and C_4 subject to the equivalences $4C_2 + C_4 - C_2 + C_4 \equiv 0$ and $-2C_4 - C_2 + C_4 - C_2 \equiv 0$. The corresponding homologies $3C_2 + 2C_4 \simeq 0$ and $2C_2 - C_4 \simeq 0$ have determinant 1. On the other hand $\pi_1(V)$ is nonzero because on adjoining the first of the following equivalences one has

$$-C_2 + C_4 - C_2 + C_4 \equiv 0, \qquad 5C_2 \equiv 0, \qquad 3C_4 \equiv 0,$$

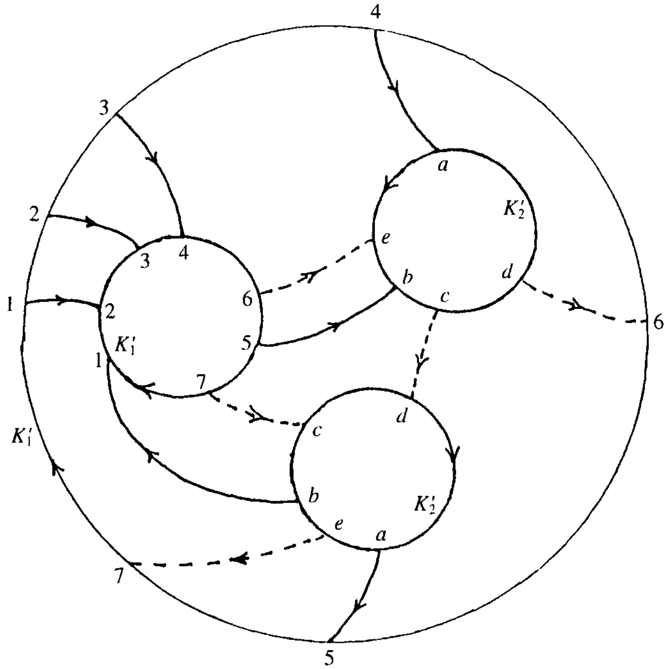


Fig. 20. Poincaré's homology 3-sphere.

which are the defining relations of the **icosahedral group**. (This, and Example 3 of "Analysis Situs", already suggest what Kneser [36] later checked: V can be obtained by conjugating facets of a dodecahedron.)

Then comes *the famous query*: "is it possible that the fundamental group of V reduces to the identity substitution, and yet V is not diffeomorphic to a sphere?"... "*But this question will drag us too far*". (**Poincaré's conjecture** still seems to be open, but we note that Poincaré's method, i.e. Morse theory, did enable Smale [80] to show, *in dimensions $n \geqslant 5$*, that any homotopy n-sphere is necessarily *homeomorphic* to the n-sphere; by Milnor [44] it need not be diffeomorphic.)

Bibliography

[1] S. Akbulut and J.D. McCarthy, *Casson's Invariant for Oriented Homology 3-Spheres*, Princeton, NJ (1990).

[2] J.W. Alexander, *A proof of the invariance of certain constants in analysis situs*, Trans. Amer. Math. Soc. **16** (1915), 148–154.

[3] P. Appell, *Henri Poincaré, en mathématiques spéciales à Nancy*, Acta Math. **38** (1921) 189–195; *Oeuvres de Henri Poincaré*, Vol. XI, Gauthier-Villars, Paris (1951), 139–145.

[4] E. Betti, *Sopra gli spazi un numero qualunque di dimensioni*, Ann. Mat. Pura Appl. **4** (1871), 140–158.

[5] G.D. Birkhoff, *Proof of Poincaré's last geometric theorem*, Trans. Amer. Math. Soc. **14** (1913), 14–22.

[6] R. Bott, *The stable homotopy of the classical groups*, Ann. of Math. **70** (1959), 313–337.

[7] H.R. Brahana, *Systems of circuits on two dimensional manifolds*, Ann. of Math. **23** (1921-22), 144–168.

[8] H.R. Brahana, *A theorem concerning unit matrices with integer elements*, Ann. of Math. **24** (1922-23), 265–270.

[9] L.E.J. Brouwer, *Beweis der Invarianz der Dimensionenzahl*, Math. Ann. **70** (1911), 161–165.
[10] S.S. Cairns, *On the triangulation of regular loci*, Ann. of Math. **35** (1934), 579–587.
[11] S.S. Cairns, *A simple triangulation method for smooth manifolds*, Bull. Amer. Math. Soc. **67** (1961), 389–390.
[12] É. Cartan, *Sur les nombers de Betti des espaces de groupes clos*, C. R. Acad. Sc. **187** (1928), 196–198.
[13] É. Cartan, *Sur les invariants intégraux de certains espaces homogènes clos et les propriétes topologiques de ces espaces*, Ann. de la Soc. Pol. Math. **8** (1929), 181–225.
[14] A. Connes, *Cohomologie cyclic et foncteurs* Ext^n, C. R. Acad. Sc. **296** (1983), 953–958.
[15] G. Darboux, *Éloge historique d'Henri Poincaré* (1913), Oeuvres d'Henri Poincaré, Vol. II, Gauthier-Villars, Paris (1952), VII–LXXI.
[16] M. Dehn, *Die Eulersche Formel in Zusammenhang mit dem Inhalt in der nicht-Euklidischen Geometrie*, Math. Ann. **61** (1905), 561–586.
[17] J.D. Dixon, *Structure of Linear Groups*, Van Nostrand Reinhold, London (1971).
[18] S.K. Donaldson, *An application of gauge theory to four-dimensional topology*, J. Diff. Geom. **18** (1983), 279–315.
[19] W. Dyck, *Beitrage zur Analysis Situs*, Math. Ann. **32** (1888), 457–512; **37** (1890), 273–316.
[20] R.D. Edwards, *The double suspension of a certain homology* 3*-sphere*, Notices Amer. Math. Soc. **22** (1975), A–334.
[21] S. Eilenberg, *Singular homology theory*, Ann. of Math. **45** (1944), 407–444.
[22] S. Eilenberg, *Singular homology in differentiable manifolds*, Ann. of Math. **48** (1947), 670–681.
[23] M.H. Freedman, *The topology of four-dimensional manifolds*, J. Diff. Geom. **17** (1982), 357–453.
[24] C. Golé and G.R. Hall, *Poincaré's proof of Poincaré's last geometric theorem*, Twist Mappings and Their Applications, R. McGehee and K.R. Meyer, eds, Springer, Berlin (1992), 135–151.
[25] P. Heegaard, *Forstudier til en topologisk teori for de algebraiske Fladers Sammenhäng*, Thesis, Copenhagen (1898). Translation: *Sur l'analysis situs*, Bull. Soc. Math. France **44** (1916), 161–242.
[26] F. Hirzebruch and D. Zagier, *The Atiyah–Singer Theorem and Elementary Number Theory*, Publish or Perish, Boston (1974).
[27] H. Hopf, *Vektorfelder in n-dimensionalen Mannigfaltigkeiten*, Math. Ann. **96** (1926), 427–440.
[28] H. Hopf, *Eine Verallgemeinerung de Euler–Poincaréschen Formel*, Nach. Ges.-Wiss, Göttingen (1928), 127–136.
[29] H. Hopf, *Über die Abbildungen der dreidimensionalen sphäre auf die Kugelfläche*, Math. Ann. **104** (1931), 637–665.
[30] J.F.P. Hudson, *Piecewise Linear Topology*, Benjamin, New York (1969).
[31] W. Hurewicz, *Beiträge zur Topologie der Deformationen*, Proc. Akad. Wet. Amsterdam **38** (1935), 112–119.
[32] É. de Jonquières, Note sur un point fondamental de la théorie des polyèdres, C. R. Acad. Sc. **60** (1890), 110–115.
[33] C. Jordan, *Mémoire sur les équations différentielles linéaires à intégrale algébrique*, J. für Reine und Angew. Math. **84** (1878), 89–215.
[34] F. Klein, *Vergleichende Betrachtungen über neure geometrische Forschungen*, Eintritts-Programm Erlangen (1872).
[35] F. Klein, *Ueber binäre Formen mit linearen Transformationen in sich selbst*, Math. Ann. **9** (1875), 183–208.
[36] H. Kneser, *Geschlossene Flächen in dreidimensionale Mannigfaltigkeiten*, Jahr. Deutsch. Math.-Verein **38** (1929), 248–260.
[37] W. Kühnel and T.F. Banchoff, *The* 9*-vertex complex projective plane*, Math. Intelligencer **5** (1983), 11–22.
[38] N.H. Kuiper, *The quotient space of* $\mathbb{C}P(2)$ *by complex conjugation is the* 4*-sphere*, Math. Ann. **208** (1974), 175–177.
[39] K. Lamotke, *The topology of complex projective varieties after S. Lefschetz*, Topology **20** (1981), 15–51.
[40] S. Lefschetz, *L'Analysis Situs et la Géométrie Algébriques*, Gauthier-Villars, Paris (1924).
[41] S. Lefschetz, *Manifolds with a boundary and their transformations*, Trans. Amer. Math. Soc. **29** (1927), 429–462.
[42] L. Lusternik and L. Schnirelmann, *Méthodes Topologiques dans les Problèmes Variationnels*, Hermann, Paris (1934).
[43] W.S. Massey, *The quotient space of the complex projective plane under conjugation is a* 4*-sphere*, Geom. Dedic. **2** (1973), 371–374.
[44] J.W. Milnor, *On manifolds homeomorphic to the* 7*-sphere*, Ann. of Math. **64** (1956), 399–405.
[45] J.W. Milnor, *Topology from the Differentiable Viewpoint*, Univ. of Virginia, Charlottesville (1966).

[46] J.W. Milnor, *Singular Points of Complex Hypersurfaces*, Princeton, NJ (1968).
[47] M. Morse, *The foundations of a theory in the calculus of variations in the large*, Trans. Amer. Math. Soc. **30** (1928), 213–274.
[48] M.H.A. Newman, *On a foundation of combinatory analysis situs, II, Theorems on sets of elements*, Proc. Amster. Acad. **29** (1926), 627–641.
[49] É. Picard and G. Simart, *Théorie des fonctions algébriques de deux variables indépendant*, I, II, Gauthier-Villars, Paris (1897).
[50] H. Poincaré, *Sur les courbes définies par une équation différentielle*, C. R. Acad. Sc. **90** (1880), 673–675; *Oeuvres de Henri Poincaré*, Vol I. Gauthier-Villars, Paris (1951), 1–2.
[51] H. Poincaré, *Mémoire sur les courbes définis par une équation différentielle*, Jour. de Math. **7** (3) (1881), 375–422; **1** (1885), 167–244; **2** (1886), 151–217; *Oeuvres*, Vol. I, 3–84, 90–158, 167–222.
[52] H. Poincaré, *Théorie des groupes Fuchsiens*, Acta Math. **1** (1882), 1–62; *Oeuvres*, Vol. II, 108–168.
[53] H. Poincaré, *Mémoire sur les groupes Kleinéens*, Acta Math. **3** (1883), 49–62; *Oeuvres*, Vol. II, 258–299.
[54] H. Poincaré, *Sur les résidus des intégrales doubles*, Acta Math. **9** (1887), 321–380.
[55] H. Poincaré, *Sur le problème des trois corps et les equations de la dynamiques*, Acta Math. **13** (1890), 1–270; *Oeuvres*, Vol. VII, 262–479.
[56] H. Poincaré, *Sur l'analysis situs*, C. R. Acad. Sc. **118** (1892), 663–666; *Oeuvres*, Vol. VI, 189–192.
[57] H. Poincaré, *Les Méthodes Nouvelles de la Méchanique Céleste*, Vols I, II, III, Gauthier-Villars, Paris (1892, 1893, 1899); Dover, New York (1957).
[58] H. Poincaré, *Sur la généralisation d'un théorème d'Euler relatif aux polyèdres*, C. R. Acad. Sci., **117** (1893), 144–145; *Oeuvres*, Vol. XI, 6–7.
[59] H. Poincaré, *Analysis situs*, J. Ec. Poly. **1** (1895), 1–121; *Oeuvres*, Vol. VI, 193–288.
[60] H. Poincaré, *Sur les nombres de Betti*, C. R. Acad. Sc. **128** (1899), 629–630; *Oeuvres*, Vol. VI, 289.
[61] H. Poincaré, *Complément à l'analysis situs*, Rend. Circ. Math. d. Pal. **13** (1899), 285–343; *Oeuvres*, Vol. VI, 290–337.
[62] H. Poincaré, *Second complément à l'analysis situs*, Proc. Lond. Math. Soc. **32** (1900), 277–308; *Oeuvres*, Vol. VI, 338–370.
[63] H. Poincaré, *Analyse des travaux scientifiques de Henri Poincaré faite par lui-même* (1901), Acta Math. **38** (1921), 1–135.
[64] H. Poincaré, *Sur l'analysis situs*, C. R. Acad. Sc. **133** (1901), 707–709; *Oeuvres*, Vol. VI, 371–372.
[65] H. Poincaré, *Sur la connexion des surfaces algebriques*, C. R. Acad. Sc. **133** (1901), 969–973; *Oeuvres*, Vol. VI, 393–396.
[66] H. Poincaré, *Sur certaines surfaces algebriques; troisième complément à l'analysis situs*, Bull. Soc. Math. France **30** (1902), 49–70; *Oeuvres*, Vol. VI, 373–392.
[67] H. Poincaré, *Sur les cycles des surfaces algèbriques; quatrième complément à l'analysis situs*, J. de Math. **8** (1902), 169–214; *Oeuvres*, Vol. VI, 397–434.
[68] H. Poincaré, *La Science et l'Hypothèse*, Flammarion, Paris (1904). English translation: *Science and Hypothesis*, Dover, New York (1952).
[69] H. Poincaré, *Cinquième complément à l'analysis situs*, Rend. d. Circ. math. di Pal. **18** (1904), 45–110; *Oeuvres*, Vol. VI, 435–498.
[70] H. Poincaré, *Sur un théorèm de géométrie*, Rend. d. Circ. math. di Pal. **33** (1912), 375–407; *Oeuvres*, Vol. VI, 499–538.
[71] K. Reidemeister, *Homotopieringe und Linsenraüme*, Abhand. Sem. Hamburg **11** (1935), 102–109.
[72] G. de Rham, *Sur l'analysis situs des variétés à n dimensions*, J. Math. Pure Appl. **10** (1931), 115–200.
[73] B. Riemann, *Grundlagen für eine allgemeine Theorie der Functionen einer veränderlichen complexen Grösse*, Inauguraldissertation, Göttingen, 1851; Collected Works of Bernhard Riemann, H. Weber, ed., Dover, New York (1953), 3–45.
[74] B. Riemann, *Theorie der Abel'schen Functionen*, J. fur Reine und Angew. Math. **54** (1857); Collected Works, 88–142.
[75] B. Riemann, *Fragment aus der Analysis Situs*, Collected Works, 479–482.
[76] K.S. Sarkaria, *From calculus to cyclic cohomology*, I. H. E. S. (1995) M/95/82.
[77] K.S. Sarkaria, *A look back at Poincaré's Analysis Situs*, Henri Poincaré Science et Philosophie, Akademie, Berlin (1996), 251–258.
[78] H. Seifert and W. Threlfall, *Lehrbuch der Topologie*, Teubner, Leipzig (1934). Translation: *A Text Book of Topology*, Academic, New York (1980).
[79] I.R. Shafarevich, *Basic Algebraic Geometry*, Springer, Berlin (1977).

[80] S. Smale, *Generalized Poincaré's conjecture in dimensions greater than* 4, Ann. of Math. **74** (1961), 391–406.

[81] H.J.S. Smith, *On systems of linear indeterminate equations and congruences*, Phil. Trans. Roy. Soc. **151** (1861), 293–326; *Collected Works*, Vol. 1, Chelsea, New York (1965), 367–409.

[82] D.M. Y. Sommerville, *The relations connecting the angle-sums and volume of a polytope in space of n dimensions*, Proc. Roy. Soc. Lond. A **115** (1927), 103–119.

[83] R.P. Stanley, *The number of faces of simplicial polytopes and spheres*, Ann. N. Y. Acad. Sci. **440** (1988), 212–223.

[84] N.E. Steenrod, *The Topology of Fibre Bundles*, Princeton, NJ (1951).

[85] D.P. Sullivan, *Infinitesimal computations in topology*, Publ. I. H. E. S. **47** (1977), 269–332.

[86] W.P. Thurston, *The Geometry and Topology of 3-Manifolds*, Princeton, NJ (1980).

[87] H. Tietze, *Über die topologischen Invarianten mehrdimensionaler Mannigfaltigkeiten*, Monat. für Math. und Phys. **19** (1908), 1–118.

[88] H. Weyl, *Die Idee der Riemannschen Flächen*, Teubner, Leipzig (1918).

[89] J.H.C. Whitehead, *On C^1-complexes*, Ann. of Math. **41** (1940), 809–824.

[90] J.H.C. Whitehead, *On incidence matrices, nuclei and homotopy types*, Ann. of Math. **42** (1941), 1197–1239.

[91] J.H.C. Whitehead, *On simply connected, 4-dimensional polyhedra*, Comment. Math. Helv. **22** (1949), 48–92.

[92] H. Whitney, *The self intersections of a smooth n-manifold in 2n-space*, Ann. of Math. **45** (1944), 220–246.

This Page Intentionally Left Blank

CHAPTER 7

Weyl and the Topology of Continuous Groups

Thomas Hawkins

Department of Mathematics, Boston University, 111 Commington St., Boston, MA 02215, USA

Introductory overview

The idea of a general theory continuous groups goes back to Sophus Lie (1842–1899) who developed the theory in the decade 1874–1884, published a three-volume synthesis of his theory with the help of his assistant Friedrich Engel [1888–93] and created a school of mathematicians devoted to the theory and application of continuous groups – and including among its members the great French mathematician Élie Cartan (1869–1951).[1] Lie's work is the common origin of both the modern theory of Lie groups and the more general theory of topological groups. During Lie's lifetime the topological considerations that nowadays seem essential in such theories were not a part of the theory, for in Lie's time topology was in its infancy. Consequently the theory was developed by purely analytical means, although these were inadequate in certain respects. Even as topology developed, however, Lie's students did not initiate the application of topology to his theory.

Hilbert was the first to introduce a topological viewpoint into the theory of continuous groups, and during 1909–1910 his work was carried on in a sense by L.E.J. Brouwer; but this did not lead anywhere at the time. One reason was that topology itself first required further development and refinement, and Brouwer was a pioneer in this direction. His profound contributions to topology in 1911–1912 did much to stimulate a ground swell of mathematical research on topological questions during the following twenty-five years.

With the rise of topology as a major mathematical discipline it was perhaps inevitable that it would eventually be applied to deal with some of the questions that had arisen in Lie's theory of continuous groups. Indeed, in [1925] Otto Schreier (1901–1929) proposed to deal topologically with a question apparent to any critical reader of Lie's treatises but ignored by Lie and his students: what is the relation of groups which are locally isomorphic? Schreier dealt rigorously and abstractly with groups – called continuous groups by him – that would now be described as topological groups which are Hausdorff, connected and locally Euclidean. His main result was undoubtedly motivated by Lie's misleading pronouncements (discussed in Section 1) regarding the global isomorphism of groups with isomorphic Lie algebras. Schreier's Theorem may be stated as follows: *Consider the class*

[1] See my papers [1982, 1987, 1989, 1992, 1994, 1998].

HISTORY OF TOPOLOGY
Edited by I.M. James

Hermann Weyl (1885–1955)

of all continuous groups $\mathfrak{G}$ which are locally isomorphic. Then there exists within this class a group $\widetilde{\mathfrak{G}}$ with the following properties: If $\mathfrak{G}$ is any group of this class then $\mathfrak{G}$ is isomorphic to a factor group $\widetilde{\mathfrak{G}}/\mathfrak{D}$, where $\mathfrak{D}$ is a discrete subgroup of the center of $\widetilde{\mathfrak{G}}$ and is isomorphic to the Poincaré fundamental group of $\mathfrak{G}$, $\pi_1(\mathfrak{G})$. The group $\widetilde{\mathfrak{G}}$, which he showed to be simply connected and unique up to global isomorphisms, was called a covering group. The name was suggested to Schreier by his colleague at the University of Hamburg, Emil Artin, and reflects the connection with the development of the notion of a covering space that arose in the context of the uniformization problem of complex function theory (Section 3).[2]

Schreier was the first to lay down precisely the foundations for the modern theory of topological groups. However, slightly earlier, in 1924, Hermann Weyl (1885–1955) had already, in a more informal manner, introduced topological considerations into the theory of the structure and representation of semisimple Lie groups which are closely related to Schreier's theorem. Schreier, who did his work without knowing of Weyl's,[3] thus inadvertently laid the foundations for some of the central topological considerations in Weyl's great paper on the representation of semisimple continuous groups [1925]. However, it was Weyl's work, namely the paper [1925] and its sequel [1927] (written with F. Peter), with their wealth of new ideas and viewpoints and impressive theorems that provided the main impulse for the development of the global, topological aspects of mathematical theories related to continuous groups. In this connection I have in mind particularly: (1) the topology of Lie groups; (2) the more general development of topological groups in conjunction with Hilbert's fifth problem; and (3) the development of harmonic analysis on groups. These three interrelated areas of research provided the main thrust behind the development of topological notions within the context of continuous groups. And Weyl's work, in one way or another, motivated them all. Had Schreier not proved the above-mentioned theorem, someone else would have done so, motivated by Weyl's work. On the other hand, it is not clear what would have happened without the work of Weyl.

The following essay has two primary goals: (i) to sketch the historical background to Weyl's work with emphasis on those developments which led Weyl to bring topological considerations to bear upon the problems with which he was dealing; (ii) to indicate how and why Weyl's work was so influential in promoting the lines of development (1)–(3). The first four sections are related to goal (i) and the fifth to goal (ii). Sections 1–3 also provide the background to Schreier's work, but here the emphasis is upon Weyl.[4] For a broader view of the development of topological groups with special emphasis on Lie groups, see [Freudenthal, 1968].

1. Lie's theory of transformation groups

The continuous groups that Lie spent his life studying were groups of transformations, which he conceived of in intuitive geometrical terms as acting upon or transforming the

[2] Schreier's construction of $\widetilde{\mathfrak{G}}$ in [1925] is quite different from the geometrically intuitive construction of a universal covering surface in uniformization theory, but in [1927] he adopted this approach and extended the scope of his theorems by replacing the local Euclidean hypothesis by weaker assumptions.

[3] Weyl's work is not mentioned in Schreier's papers [1925, 1927] but is discussed in his survey article [1928].

[4] This essay was written thanks to the Resident Fellowship granted to me by the Dibner Institute for the History of Science and Technology located at MIT.

elements of an n-dimensional "manifold" which might represent points or lines or some other type of geometrical object coordinatized by n variables, $x_1, \dots, x_n$. From the outset of his work with such groups their characteristic feature was that the "finite" transformations $x' = Tx$ of the group $\mathfrak{G}$ were generated by infinitesimal transformations. For Lie, an infinitesimal transformation was defined by a differential operator $X = \sum_{i=1}^{n} \xi_i(x)\partial/\partial x_i$, the associated infinitesimal transformation being $x_i' = x_i + X(x_i)\delta t$. (Nowadays X would be regarded as a vector field.) The infinitesimal transformation X generates a one parameter family of transformations $x' = T_t x$, where $T_t = \exp(tX) = \sum_{j=0}^{\infty} t^j X^j / j!$; and to say that any $T \in \mathfrak{G}$ is generated by an infinitesimal transformation is to say that T belongs to such a one parameter family. Initially for Lie it was this property that made the group $\mathfrak{G}$ continuous. Thus in [1883, p. 314] he wrote: "A group is called *continuous* when all of its transformations are generated by repeating infinitesimal transformations infinitely often ...". Since T_t with $t = 0$ corresponds to the identity transformation, this means that every T is connected via a one parameter family T_t to the identity transformation. Thus Lie's notion of the continuity of a group involved a kind of connectivity.

In the first volume of his treatise *Theorie der Transformationsgruppen*, Lie began with some generalities about what it meant for a group to be continuous and presented the following modified characterization: "A transformation group is called *continuous* if it is possible for any given transformation belonging to the group to specify certain other transformations of the group which differ only infinitely little from the given transformation, [and] if, on the other hand, it is not possible to decompose the totality of transformations in the group into individual discrete families" [1888–93, vol. 1, p. 3]. Here it is not explicitly claimed that all transformations of the group are generated by infinitesimal ones, and the second part of the definition serves to guarantee the sort of connectedness that had been implicit in the earlier definition. Lie provided no example that would illustrate the need for the new definition so that it is uncertain what he had in mind in formulating it. However, two years later, in 1890, Engel discovered what amounts to an example of a group continuous in the latter sense but not in the former. That is, he discovered that not every transformation of the special linear group $\mathbf{SL}(2, \mathbb{C})$ is generated by an infinitesimal transformation. Hence $\mathbf{SL}(2, \mathbb{C})$ is not continuous in accordance with Lie's definition of 1883 but it is continuous in the sense of the later definition. Perhaps Lie had some such example in mind when he proposed the new definition, although if he had $\mathbf{SL}(2, \mathbb{C})$ in mind, he evidently did not reveal his thoughts to Engel.

A prime example of a continuous transformation group is the projective group of the line which is defined by the equation

$$x' = \frac{ax + b}{cx + d}, \quad ad - bc \neq 0. \tag{1}$$

As an example of group which is not continuous, Lie gave the group of all coordinate changes in the plane that take one (right-angled) Cartesian coordinate system into another [1888–93, vol. 1, p. 7], i.e. all translations, rotations and reflections in planes. The transformations of the group are given by two different sets of equations, defining two discrete 'continuous' families of transformations which are defined, respectively, by

$$\left\{ \begin{aligned} x' &= a + x\cos\alpha - y\sin\alpha \\ y' &= b + x\sin\alpha + y\cos\alpha \end{aligned} \right\} \quad \text{and} \quad \left\{ \begin{aligned} x' &= a + x\cos\alpha + y\sin\alpha \\ y' &= b + x\sin\alpha - y\cos\alpha \end{aligned} \right\}.$$

The family on the left involves an orthogonal transformation of determinant +1 while that on the right by one with determinant −1.

The above described discussion of what it means for a group to be continuous, given in the introductory pages of the first volume of *Theorie der Transformationsgruppen*, was as far as Lie ever ventured into a discussion of the global, topological properties of his continuous groups. Although he was a geometer in spirit, when it came to developing his theory of groups he saw no choice but to use the standard analytical language of the period. As a consequence he was not compelled to develop his ideas on the continuity of a group because, at least for the groups on which he concentrated in *Theorie der Transformationsgruppen*, he saw it as guaranteed by the assumption that the group is defined by a *single* set of equations given by analytic functions. That is, Lie considered groups of transformations defined by a single system of equations involving a finite number of parameters:

$$x_i' = f_i(x_1, \ldots, x_n, a_1, \ldots, a_r), \quad i = 1, 2, \ldots, n. \tag{2}$$

It is assumed that the functions in (2) are analytic in the n variables $x_1, \ldots, x_n$ and the r parameters $a_1, \ldots, a_r$ [1888–93, vol. 1, p. 11], and this was seen as guaranteeing that the family defined by (2) was continuous [1888–93, vol. 1, p. 311].

A group $\mathfrak{G}$ of transformations defined by a single set of equations of the form (2) is thus the principal object of study in *Theorie der Transformationsgruppen*. These are the 'finite continuous transformation groups' of the theory. However, an additional assumption on such equations is necessary before they can provide the theoretical starting point. In (2) all the parameters $a_1, \ldots, a_r$ need not be 'essential'. According to Lie they are essential "if it is impossible to introduce as new parameters independent functions of $a_1, \ldots, a_r$ so that as a result the equations $x_i' = f_i(x, a)$ contain fewer than r parameters" [1888–93, vol. 1, p. 12]. I have given Lie's definition *verbatim* to illustrate the manner in which he typically masked the *local* nature of what he was discussing. For example, the four parameters a, b, c, d of the projective group of the line (1) are not essential; there are for Lie only three essential parameters involved. For example, in a sufficiently restricted neighborhood of (a_0, b_0, c_0, d_0) where $d_0 \neq 0$, one will have $d \neq 0$ and one can divide through by d and express (1) in terms of three parameters as $x' = (\alpha x + \beta)/(\gamma x + 1)$, where $\alpha = a/d$, and so on. In Lie's actual definition of essential parameters, however, there is no mention of the fact that new parameters need only be defined locally as functions of the original parameters.

Indeed, most of Lie's theory is developed in a neighborhood of the identity element of the group. That is all that is needed to show that the r-dimensional group $\mathfrak{G}$ defined by (2) possesses r linearly independent infinitesimal transformations $X_1, \ldots, X_r$ such that their linear span $\mathfrak{g}$ comprises all infinitesimal transformations of $\mathfrak{G}$ and, in addition, that for all i, j $[X_i, X_j] = X_iX_j - X_jX_i \in \mathfrak{g}$. This means, in modern terms, that $\mathfrak{g}$ is an r-dimensional Lie algebra. Much of Lie's theory involves $\mathfrak{g}$ rather than $\mathfrak{G}$; $\mathfrak{g}$ was not thought of as separate from $\mathfrak{G}$ but as comprising the infinitesimal transformations of $\mathfrak{G}$. One could say that Lie's theory of transformation groups is a theory of group germs, but this would be misleading if taken to mean that Lie explicitly and carefully formulated his theory in such terms, for he did not. Quite the contrary, as in his definition of essential parameters, he tended to suppress much mention of the truly local nature of the reasoning underlying his theory. Indeed, he went further: he articulated his theorems in a form that made them appear to be globally true. Thus Lie stated that every $T \in \mathfrak{G}$ is of the form $T = \exp(X)$,

$X \in \mathfrak{g}$ [1888–93, vol. 1, p. 75], even though the proof preceding it was local in nature. I will refer to this as Lie's exponential mapping theorem.

Another example is given by Lie's isomorphism theorem [1888–93, vol. 1, p. 418]. Suppose that $\mathfrak{G}$ and $\mathfrak{H}$ are r-dimensional continuous groups with Lie algebras $\mathfrak{g}$ and $\mathfrak{h}$, respectively. Lie said that the groups have the same structure or composition if bases $X_1, \ldots, X_r$ and $Y_1, \ldots, Y_r$ of their respective Lie algebras could be chosen so that for all i, j $[X_i, X_j]$ and $[Y_i, Y_j]$ are the same linear combination of the respective bases. This of course means, in modern terminology, that $\mathfrak{g}$ and $\mathfrak{h}$ are isomorphic as Lie algebras. His isomorphism theorem states that $\mathfrak{G}$ and $\mathfrak{H}$ are isomorphic as groups if and only they have the same structure. In his statement of this theorem he even emphasized the one-to-one, onto nature of the correspondence between $\mathfrak{G}$ and $\mathfrak{H}$, even though his proof can be seen to establish nothing more than a local isomorphism in the neighborhood of the identity elements.

The apparent global mode of statement used by Lie may have been merely a loose manner of speaking, but I suspect it was perhaps more than that. I suspect that he regarded his theorems as generic statements, i.e. "generally" true in the global form of his statements. He was imitating the analysts of the 18th and early 19th centuries (e.g., Lagrange, Jacobi, Clebsch) whose work on differential equations he drew on extensively in developing his theory. They preferred to reason and state their results in terms of what they saw as the "general" case. In such a frame of mind, possible exceptions to the generic theorems are not something of concern and may not even be expected.

That is why Engel was somewhat surprised by his discovery that not every transformation of $\mathbf{SL}(2, \mathbb{C})$ is generated by an infinitesimal transformation so that Lie's exponential mapping theorem does not hold globally. When he communicated the discovery to his friend, the mathematician Eduard Study, who was working at the time within Lie's sphere on the applications of Lie's theory to the theory of invariants, Study replied that this property of $\mathbf{SL}(2, \mathbb{C})$ was new to him, although he had always found Lie's exponential mapping theorem "amazing" (*wunderbar*).[5] This shows that Study interpreted Lie's statement of his theorem literally, albeit with suspicion. Most of Study's letter was devoted to a geometrical explanation of why $\mathbf{SL}(2, \mathbb{C})$ had this surprising property, and he suggested they write a paper on the matter with the title "A Paradox in the Theory of Groups". In a postcard the next day he added that such a paper was a good idea because most mathematicians "know nothing about such things".

Such a joint paper was never written, but Engel went on to write a two part paper on the matters raised by his discovery [1892, 1893]. It did not, however, bear Study's suggested title because that title was not in keeping with the main thrust of Engel's paper. In this connection it should be noted that although in Lie's theory "linear" groups play the expected central role, by virtue of Lie's background in projective geometry, he tended to couch everything "linear" in terms of projective space and projective transformations. In his paper Engel showed that for the general projective group $\mathbf{PGL}(n, \mathbb{C})$ and two of its most important subgroups, namely the projective orthogonal group $\mathbf{PSO}(n, \mathbb{C})$ and the projective symplectic group $\mathbf{PSp}(2n, \mathbb{C})$, Lie's exponential mapping theorem is globally valid. These results formed the bulk of Engel's papers, which thus downplayed the "paradoxical" phenomenon displayed by $\mathbf{SL}(2, \mathbb{C})$.

Nonetheless, in his second paper Engel did present Study's geometrical explanation of the basis for that phenomenon. He also called attention to the implication the examples

[5] Letter to Engel dated 9 November 1890. The original is located in the Engel archive at the University of Giessen, as are the two postcards Study sent the next day by way of a postscript.

SL(2, $\mathbb{C}$) and **PGL**(2, $\mathbb{C}$) had for Lie's isomorphism theorem. That is, it was well known that these two groups have isomorphic Lie algebras, but they are not globally isomorphic as Lie's isomorphism theorem literally asserts. Engel put it this way: since the two groups have isomorphic Lie algebras "it follows, as Lie has proved ..., that both groups are also isomorphic in the sense of the theory of permutations *so long as one restricts oneself to the transformations in a certain neighborhood of the identical transformation*. This sort of limitation is assumed in all the theories of ... [Lie, 1888–93, vol. 1] ... and only under this assumption is it there proved that groups with the same structure are also isomorphic" [1893, p. 695]. In particular, Engel pointed out that **SL**(2, $\mathbb{C}$) and **PGL**(2, $\mathbb{C}$) are not isomorphic "in the strictest sense of the word", i.e. on a global level.

Of course Engel's blanket assumption that everything in *Theorie der Transformationsgruppen* is only true for group elements in a certain neighborhood of the identity was not explicitly made there, and the global nature of the statements of the theorems themselves continued to leave readers unclear as to their intended import. Indeed a decade later, Engel [1902] had once again to explain that Lie's theorems were only to be understood as local theorems based on local reasoning after an American mathematician, Steven Slocum, having studied *Theorie der Transformationsgruppen* but oblivious to this point, claimed to have discovered a flaw in Lie's proof of one of his fundamental theorems.

A question raised by Engel's remarks is: what can be said about the relation of groups which have the same structure in the sense that they have isomorphic Lie algebras and so are locally isomorphic? This is, of course, precisely the question posed by Schreier and also raised, as it applies to group representations, by Weyl. To my knowledge, however, no one within Lie's school considered this problem. This is not surprising since evidently a global approach would be needed whereas the analytical methods used by Lie and his students gave only local results, as Engel emphasized in the above quoted passage. The initial impetus for a global, topologically oriented approach to continuous groups came from without, from Hilbert at Göttingen.

2. Hilbert's fifth problem

Hilbert's interest in the foundations of geometry turned his attention to what became known as the "space problem" of Hermann von Helmholtz (1821–1894), a distinguished physiologist and physicist who in 1868 proposed to deduce the geometrical nature of space from observed facts of experience having to do with the properties of mobile rigid bodies. Using calculus, he argued that his facts lead to the conclusion that metric relations in space correspond either to Euclidean geometry or to the geometry of Lobachevsky. In [1887] Poincaré used Lie algebra techniques to solve the analog of Helmholtz's problem in two dimensions, and Lie [1890] did the same in n dimensions. Hilbert questioned the necessity of assuming, as was done by Helmholtz and his successors, the differentiability of the transformations they considered, a *sine qua non* if one were to solve the problem using Lie's theory. In particular, the assumption that the group is generated by infinitesimal transformations did not fit in easily or naturally with the other geometrical axioms. Hilbert wondered whether this assumption might actually follow as a consequence of the continuity of the transformations defining the rigid motions, together with the group property and the other axioms of geometry. Thus in his famous lecture on mathematical problems Hilbert posed as his fifth problem the more general question as to what extent Lie's theory, with its differentiability assumptions, could be recovered from seemingly weaker continuity assumptions about the transformations of the group and their multiplication. "Hence there arises the question whether, through the introduction of suitable new variables and parameters, the group can always be transformed into one with differentiable defining functions; or whether at least with the help of certain simple assumptions a transformation is possible into groups admitting Lie's methods" [1900, p. 452].

Shortly thereafter, Hilbert himself dealt with this problem within the limited context of the analog of Helmholtz's problem for the plane. Here we find the first attempt to deal with continuous groups by topological means. Hilbert stressed the fact that his methods of proof were completely different from Lie's, that he would mainly use the concepts of Cantor's theory of point sets and the Jordan Curve Theorem for the plane [1902, p. 234]. Of course, the context of Hilbert's topological approach to groups was a special and limited one, but it showed that results could be achieved and suggested the possibility of doing something similar with respect to the more formidable fifth problem. Although the details of Hilbert's paper need not concern us, it should be noted that in seeking to characterize the plane as a two-dimensional manifold [1902, pp. 234–235], Hilbert introduced the approach that eventually led (through Weyl, as indicated in §3) to the modern concept of a manifold. "These stipulations", Hilbert wrote, "contain for the case of two dimensions, it seems to me, the rigorous definition of the concept which Riemann and Helmholtz designated by 'multiply extended manifold' and Lie by 'number manifold' and which is at the basis of their entire investigation" [1902, p. 235].

In dealing with the special groups of motions of the plane, Hilbert had no need to characterize them as manifolds as well. That step was taken by L.E.J. Brouwer, who undertook the topological study of continuous groups with an eye towards the fifth problem [1909a, 1909b, 1910]. However, he did not follow Hilbert's lead on how to define a manifold and gave instead his own definition [1909b, p. 247], which was completely verbal and expressed so succinctly that its precise meaning must have been difficult for most readers to fathom. It was Hilbert's notion, as developed by Weyl (Section 3), that ultimately became the standard approach. In terms of his own definition, Brouwer then defined a finite continuous group to be a group of transformations acting on an n-dimensional manifold such that the transformations themselves may be identified with a p-dimensional 'parameter manifold'. Unlike Lie's finite continuous groups, no differentiability assumptions were involved in Brouwer's definition.

The goal of Brouwer's papers was to determine, using the theory of point sets, all such groups with $n = 1, 2$, the idea being to then see if indeed they could all be realized as groups in Lie's sense, thereby answering the question posed by Hilbert's fifth problem in these special cases. For the case $n = 1$ treated in [1909b] he succeeded with the classification and was able to give an affirmative answer to Hilbert's fifth problem, but for $n = 2$ [1910] he only developed the theory to a point which made it possible, he claimed, in a future paper to push through the complete classification. Brouwer, however, published no more in this vein. Undoubtedly a major reason for his termination of research on continuous groups was lack of an adequate topological theory. For the plane he had relied on the work of Schoenflies, which he discovered to be seriously flawed, and for higher dimensions little had been done. Consequently Brouwer focused his attention on topological questions and became one of the founders of modern topology.

Brouwer's work on continuous groups was reviewed for the abstracting journal *Fortschritte der Mathematik* by Engel. In his review of [Brouwer, 1909b] Engel justifiably complained about the overly succinct definition of a continuous group which, in so far as he could understand it, did not seem to cover all the possibilities for groups in Lie's sense [1912, p. 194]. Brouwer objected to Engel's unspecific criticisms and so began a brief correspondence with Engel, which Freudenthal has described as "a discussion between people living in different worlds: Engel, the co-author of Lie's great treatise, who could not grasp a group except in its analytical setting, and Brouwer, who had shaken off the algorithmic yoke and from his conceptual viewpoint could not comprehend his correspondent's difficulties."[6] Although Freudenthal's evaluation is essentially correct, Engel's difficulty in comprehension was certainly magnified by Brouwer's excessively compressed writing style. But even after Brouwer patiently explained the meaning and implications of his terms, Engel remained overwhelmed by and suspicious of Brouwer's topological approach. Thus in his review of Brouwer's second paper [1910], Engel began with some clarifying remarks about the first: "I am still of the opinion that everyone who is not an inveterate set-theoretician will find, as I did, that the general assumptions of §1 are not worded clearly enough .... I cannot conceal the fact that, in general, the vast generality of the investigation and the great number and multiplicity of the necessary lines of reasoning strikes me with a slight dread. It is actually inconceivable to me that on the first try everything should have been settled" [1913, p. 182]. As we shall see in Section 5, a similar sentiment was expressed by Élie Cartan, Lie's greatest disciple, when in 1925 he advocated avoiding the use of topological reasoning in dealing with Lie groups because of the great "delicacy" of its arguments.

Although Brouwer had stressed the fact that his approach had the advantage over Lie's of providing insight into the global structure of transformation groups [1909a, p. 303, 1909b, p. 267], it undoubtedly at the time appeared too difficult for others to emulate. In any case, the topological study of continuous groups was not taken up again until the work of Weyl and Schreier (1924–1925). Brouwer's dimension-by-dimension approach to determining all finite-dimensional continuous groups is reminiscent of Lie's effort in the 1870's to do the same, albeit by analytical means. Lie gave up on his project after solving it for $n \leqslant 3$ and only published his results for $n \leqslant 2$. However, during 1888–1913, Killing and Cartan developed the deep and powerful algebraic tools needed to completely resolve the more limited problem of classifying complex semisimple Lie groups and their linear representations up to local isomorphisms. Of course the local nature of these results was not emphasized or fully recognized within Lie's school. It first became apparent when Weyl became interested in the theory, as we shall see in Section 4. His greater sensitivity to the global aspects of Lie's theory as well as the approach he brought to bear upon the resulting questions grew out of his involvement with the uniformization problem of complex analysis, to which I now turn. One of the key notions to emerge from the consideration of this problem was that of a covering space, a notion which may have inspired Schreier's work as well.

3. The uniformization problem and covering spaces

The story behind the uniformization problem begins with the work of Poincaré. Poincaré had become involved in a friendly competition with Klein involving what is now known

[6] The correspondence, which occurred in 1912, is included in [1976, pp. 141–155]; Freudenthal's quoted editorial comment is from p. 142.

as the theory of automorphic functions.[7] One type of automorphic function studied by Poincaré he had called "Fuchsian functions," and in a note [1881, p. 31] he claimed he could prove that: "the coordinates of the points of any algebraic curve can be expressed by Fuchsian functions of an auxiliary variable." That is, as Poincaré explained in [1882, p. 101], if $f(x, y)$ is a polynomial in the complex variables x and y, then Fuchsian functions F, G defined in an open circular disc D exist such that the nonsingular points on the algebraic curve $f(x, y) = 0$ are parametrized by the equations $x = F(\zeta)$, $y = G(\zeta)$ so that $f(F(\zeta), G(\zeta)) = 0$ for all $\zeta \in D$. Poincaré's assertion can also be expressed as follows. Imagine that the equation $f(x, y) = 0$ defines, say, y as a multiple-valued analytic function of x, what was called at the time a *nonuniform* function of x. For example $f(x, y) = y^2 - x = 0$ defines the two-valued square root function y. Then since $x = F(\zeta)$ and $y = G(\zeta)$ are single-valued, or uniform, functions of ζ, they may be regarded as uniformizing the original multiple-valued function. Thus $x = \zeta^2$, $y = \zeta$ uniformizes the two-valued square root function.

Klein was impressed by Poincaré's uniformization theorem and, stimulated by Poincaré's achievements, he discovered a remarkable theorem of his own, which he called the fundamental theorem of the theory of automorphic functions [1883, pp. 698–699] and which had the uniformization theorem as a corollary.[8] Poincaré and Klein discovered that the proofs of their respective theorems involved a similar type of argument which Poincaré called "the method of continuity" [1884, p. 329]. The method involved reasoning of an essentially topological nature at a time when topology was still in its infancy. Consequently the method was fraught with difficulties. Klein made no claims of a proof but merely sketched out his intuitive ideas for such a proof. Poincaré, while suggesting that Klein's approach involved a difficulty "which cannot be overcome in a few lines" [1884, p. 332], sought to develop his own continuity arguments more carefully by establishing lemmas which "permit us to apply the method of continuity with all rigor" [1884, p. 368]. Eventually in the light of advances in set theory and topology, the arguments of both Klein and Poincaré were seen to be inadequate.[9]

Hilbert's Paris lecture of 1900, in which he focused attention on uniformization theorems, was an important factor in triggering renewed interest in these matters. Although Hilbert may have privately doubted the adequacy of the proofs of Poincaré's uniformization theorem and Klein's Fundamental Theorem, he focused his critical attention upon a far more general uniformization theorem which Poincaré set forth in a paper of 1883. There, without any explicit reference to his uniformization theorem for algebraic curves, he proposed to demonstrate the following theorem: "Let y be any nonuniform analytic function of x. A variable z can always be found so that x and y are uniform functions of z" [1883, p. 57]. To followers of Riemann it was clear that by considering the Riemann surface of the nonuniform function, x and y could be regarded as uniform functions of a variable tracing out this surface, but Poincaré's variable z traces out a portion of the complex plane and thus makes x and y uniform functions of a complex variable.

Poincaré was not very well versed in Riemannian principles and had developed his own ways of looking at Riemann surfaces [Gray, 1986, p. 299]. In particular, the demonstra-

---

7 An engaging account of this competition can be found in [Gray, 1986].

8 For a discussion of Klein's Fundamental Theorem and the closely related Limit Circle Theorem (*Grenzkriestheorem*) see [Gray, 1986, pp. 297–316].

9 A detailed comparative analysis of Klein's and Poincaré's use of the method of continuity is given by Scholz [1980, pp. 205–216].

tion of his theorem is based upon his own construction of a simply connected Riemann surface for several multivalued analytic functions. The construction involves the now familiar idea behind the construction of a universal covering space and is worth quoting in its entirety [1883, pp. 58–59]:

> We consider m analytic functions of x,
>
> $$y_1, y_2, \ldots, y_m, \tag{1}$$
>
> which are in general not uniform. These functions will be completely defined when not only the value of x is known but also the path, starting from an initial point O, by which the variable has attained this value.
>
> We do not consider the variable x as moving on a plane but on a Riemann surface S. This surface will be formed of superimposed plane sheets as in the Riemann surfaces by means of which algebraic functions are studied: only here the number of sheets will be infinite.
>
> We trace in the plane an arbitrary closed contour C which begins and ends at at the same point x. The surface will be completely defined if we state the conditions under which the initial and final point of this contour must be regarded as belonging to the same sheet or to different sheets.
>
> Now there are two sorts of contours C:
>
> (1°) Those which are such that at least one of the m functions y does not return to its initial value when the variable describes the contour C;
>
> (2°) Those which are such that the m functions y return to their initial values when the variable x describes the contour C.
>
> Among the contours of the second sort, I will distinguish two species:
>
> (1°) C will be of the first species if, by deforming this contour in a continuous manner, one can pass to an infinitesimal contour so that the contour never ceases to be of the second sort.
>
> (2°) C will be of the second species in the contrary case.
>
> Well, the initial and the final point of C will belong to different sheets if this contour is of the first sort or of the second species of the second sort. They belong to the same sheet if C is of the first species of the second sort....
>
> The Riemann surface is then defined completely. It is simply connected and does not differ, from the viewpoint of the geometry of position, from the surface of a circle, from a spherical cap or from one sheet of a hyperboloid of two sheets.

Although the above description of S is rather vague by present day standards, in view of the first paragraph of the quotation, it seems safe to say that Poincaré's remarks suggest the idea of conceiving of the points $\tilde{x}$ of the surface S which lie over the point x in the complex plane as corresponding to paths from O to x which lie in the domain D of analyticity of the functions $y_1, \ldots, y_m$, where two such paths, α and β, are regarded as determining the same point $\tilde{x}$ of S if the closed curve $-\alpha + \beta$ (α traced in reverse direction from x to O followed by β from O to x) is of the first species of the second sort. Furthermore, if the functions $y_1, \ldots, y_m$ are all assumed to be uniform and so irrelevant to the construction of S, then two paths α and β from O to x would determine the same point $\tilde{x}$ over x precisely when $-\alpha + \beta$ can be continuously deformed within D to a point. In this case S would be the universal covering space of D. As we shall see, these ideas were set forth explicitly by Poincaré in 1907 when he returned to his general uniformization theorem, spurred on by Hilbert's Paris lecture which called attention to an unsatisfactory quirk in Poincaré's demonstration and resultant theorem.

Before describing the quirk, it is necessary to say something about the overall proof idea, which was a good one: to establish the existence of what is now called a Green function on the surface S, and then using the simple connectivity of S, to deduce the existence of a one-to-one conformal mapping from S into the unit disk $|z| < 1$. This idea is still central to modern proofs of uniformization.[10] If $\tilde{x} = \Phi(z)$ denotes the inverse of this mapping and $p_*(\tilde{x}) = x$ is the projection mapping from S to D, then, since each y_i is a single-valued function on S, say $y_i = f_i(\tilde{x})$, the desired uniformization of $x, y_1, \dots, y_m$ is given by $x = p_*(\Phi(z)), y_i = f_i(\Phi(z))$.

The quirk arises in the following manner. In order to construct the Green function, Poincaré used a composite of linear fractional transformations and an elliptic modular function to define a function $\psi(\zeta)$ satisfying $|\psi(\zeta)| < 1$ which is analytic, except at three points a, b, c corresponding to the points $0, 1, \infty$ where the elliptic modular function is not analytic. Then by taking $y_m = \psi$ in the construction of the Riemann surface S, he could assume ψ was defined on S and consider the function $t = \log|1/\psi|$ on S which is then used to construct a Green function on S. However, because ψ enters into the construction of S the points a, b, c are excluded from the common domain of analyticity of $y_1, \dots, y_{m-1}, \psi$. Thus, for example, to uniformize a single function y of x, as is asserted in Poincaré's theorem, one would consider the Riemann surface S for the two functions y, ψ which excludes the points covering a, b, c even though a, b, c have nothing to do with the function y and may well be points at which y is analytic. Thus in the uniformization $x = p_*(\Phi(z)), y = f(\Phi(z))$ some perfectly good values of x and y may be excluded due to the introduction of the function ψ. Poincaré himself called attention to this artificial restriction by remarking in a note at the end of his paper that the points a, b, c "being singular points, are outside the Riemann surface."

Although Poincaré's paper [1883] seems to have been fairly well known, it was not until the turn of the century that attention was focused on its defects and on the desirability of remedying them. The first to do so was W.F. Osgood in lectures of 1898, but it was Hilbert who refocused attention on the matter of uniformization. On August 8, 1900 in his Paris lecture on mathematical problems, Hilbert posed, as one of the ten problems included in his talk, the problem of uniformization, noting that Poincaré's 1883 Theorem on the matter was subject to limitations due to the above-mentioned points a, b, c. In the published version of his talk the problem is the twenty-second [Hilbert, 1900, p. 323].

It was Poincaré himself and Paul Koebe, a *Privatdozent* at Göttingen, who in 1907 independently provided what is generally regarded as the first satisfactory resolution of the issues raised by Hilbert. Poincaré's paper [1907] is of particular interest because in it he decided to separate the construction of the Riemann surface of a nonuniform function from that of its universal covering surface. To construct the Riemann surface of such a function, Poincaré followed the lead of Weierstrass and conceived of the surface as composed of "function elements" (x, y) corresponding to pairs of series in powers of $\zeta - \zeta_0$ but, unlike Weierstrass, he added elements to correspond to poles and branch points as well [1907, pp. 73–77]. Weyl was to develop this approach systematically in his book [1913].

Having indicated how to describe in this fashion a Riemann surface for nonuniform functions $y_1, \dots, y_m$, which he called the "principal domain" D of $y_1, \dots, y_m$, Poincaré pointed out that other constructions were possible which, instead of leading to D would

[10] See in this connection the expository account by Abikoff [1981], especially §5. A more detailed description of Poincaré's own construction of the Green function and the associated conformal mapping is given by Gray [1994, §2].

lead to a "multiple domain" Δ in which every point of D corresponds to several, possibly infinitely many, points of Δ. He came closest to the modern notion of a covering space with his definition of a "regular multiple" of D [1907, p. 90]:

> We will say that Δ is a regular multiple of D if it satisfies the following condition. Let M be a point of D; let $M_1, M_2, \ldots$ be the corresponding points of Δ. Let M' be another point of D, infinitely close to M; we suppose that among the points of Δ which correspond to M', there is one which is infinitely close to M_1, one which is infinitely close to $M_2, \ldots$. If this condition is fulfilled, that is to say, if Δ is a regular multiple of D, it is clear that if to a certain point M of D a finite number n of points of Δ correspond, then there will correspond to any other point of D the *same* finite number n of points of Δ.

With these preliminaries out of the way, Poincaré returned to the ideas he had presented in his paper [1883], now presenting them without the added complications of a Riemann surface construction. The result is essentially the modern construction of a universal covering space, which he gave to justify his claim that there is a regular multiple Δ of D which is simply connected. If M_0, M are points of D:

> One can go from M_0 to M on D by many paths. Consider two of these paths. They could be *equivalent*, that is they could bound a continuous area situated on D; but they may not be, at least if D is not simply connected. That given, let us define the domain Δ. A point of this domain will be characterized by the point M of D to which it corresponds and by the path by which one proceeds from M_0. In order that two points so characterized be identical it is necessary and sufficient that one has come from M_0 by equivalent paths. It is clear that Δ is simply connected [1907, p. 90].

Poincaré then sketched an argument to the effect that Δ is a regular multiple of D.

Several months before Poincaré's paper [1907] appeared, Klein presented to the Göttingen Academy a paper by Paul Koebe [1907] which also established the general uniformization theorem without any restrictions. Koebe (1882–1945) had been a student of H.A. Schwarz at Berlin, where he received his doctorate in 1905. He became *Privatdozent* in Göttingen in 1907. Following in the footsteps of Schwarz, Koebe was interested from the outset of his career in conformal mapping problems. His proof of the uniformization theorem was based upon his theorem that any simply connected Riemann surface can be mapped conformally onto one of the following 3 regions of the Riemann sphere: a spherical cap, the sphere minus a point or the entire sphere [1907, p. 198]. (In terms of the complex plane $\mathbb{C}$, the three regions may be taken as $|z| < 1$, $\mathbb{C}$ and $\mathbb{C}$ plus the point at infinity.) To establish the uniformization theorem for nonuniform analytic functions $y_1, \ldots, y_m$ of x, Koebe considered the Riemann surface for $y_1, \ldots, y_m$. Then, citing Poincaré's original paper [1883], he concluded that if this surface is not simply connected, "it is then transformed into a simply connected surface B by means of a definite covering process. The construction of this surface offers no major difficulties" [1907]. If $\Phi : D \to B$ denotes the inverse of the conformal mapping into the Riemann sphere posited by Koebe's theorem, then on B $x, y_1, \ldots, y_m$ are all uniform functions of $b \in B$, say, $y_i = F_i(b)$, $i = 0, 1, \ldots, m$, where $y_0 = x$, and so $y_i = (F_i \circ \Phi)(\zeta)$ gives the uniformization.

During 1907–1911 Koebe published a total of 19 papers on various aspects of uniformization, including the algebraic case where he provided the first proof of Klein's "Fundamental Theorem" and many detailed studies of the various cases that can arise due to the nature of the group of linear fractional transformations defining the automorphic functions.

Although he developed other methods as well, Koebe also stressed the importance of the "Method of Covering Surfaces." Thanks to Koebe there was no danger that the subject of uniformization or covering surfaces would soon be forgotten around Göttingen. Indeed, he appears to have generated considerable interest in uniformization and conformal mappings among the young mathematicians at Göttingen *circa* 1909, which is not surprising given the notoriety he had achieved for sharing with a mathematician of the stature of Poincaré the honor of solving one of Hilbert's problems.

According to Weyl's later recollections Koebe's work on the fundamental uniformization theorems and "Hilbert's establishment of the foundation on which Riemann had built his structure and which was now available for uniformization theory, the Dirichlet Principle," were two of the "three events" that "had a decisive influence" on the form of his lectures on Riemann surfaces (presented at Göttingen in the Winter Semester 1911–1912 and published as [1913]). The other "event" was "the fundamental papers of Brouwer on topology from 1909 onwards".[11] In §2 I mentioned that in his work on continuous groups Brouwer had relied on a version of point set topology that was still somewhat naive in its reasoning. Brouwer pointed this out impressively with an *Annalen* paper of 1910 containing, among other things, an example of a curve which divides the plane into three open connected sets but is the complete boundary of each. This example and the others Brouwer constructed virtually undermined point set topology as it then stood and revealed the need for a more rigorous treatment [Johnson, 1987, pp. 67–71]. Brouwer then went on to show in a succession of innovative papers how point set topology could be developed rigorously. In particular he solved a problem of considerable interest in Göttingen, that of invariance of dimension under one-to-one continuous mappings of Euclidean space $\mathbb{R}^n$. Brouwer's theorem on invariance of dimension and the related theorem on invariance of domain were of special interest because the method of continuity utilized by Klein and Poincaré in the uniformization of algebraic curves by automorphic functions had taken such invariance properties for granted. Brouwer's proof of these invariance properties appeared in the *Annalen* in 1911, and in 1912 he applied his results to vindicate the method of continuity.

It is clear from the preface to Weyl's lectures on Riemann surfaces that the developments in point set topology had convinced him that the theory of Riemann surfaces should be developed in a form that "completely satisfies all modern requirements of *rigor*", and that Brouwer's work, with its many innovative ideas of an algebraic topological nature, had convinced him that it could be so developed: "To a far greater extent than follows from the citations, I have been encouraged by the fundamental topological investigations of Brouwer" [1913, iii]. Although Weyl shared with Klein a conviction that the essence of mathematics is intuitive and that there is a real danger in pushing the rigor and abstraction too far, he also saw the theory of Riemann surfaces as being by its nature especially in need of a rigorous formulation: "A rigorous set theoretical founding of the topological concepts and theorems which come into question in Riemannian function theory is all the more necessary since the 'points' of the basic configurations (curves and surfaces) in this case are not points in space in the usual sense but can be arbitrary mathematical entities of a different sort (e.g., function elements)" [1913, iv].

The influence of both Klein and Hilbert can be seen also in Weyl's definition of a Riemann surface. Following the lead of Klein, he accepted the idea that the theory should

[11] Quoted from the Preface of the English translation of the third edition of Weyl's book [1955]. The importance of these three events is already implicit in the preface to the 1913 edition.

be developed within the conceptual framework of the theory of surfaces rather than in terms of a wrapping of the complex plane over itself so as to make an analytic function single-valued. And it was to Hilbert's paper [1902] that Weyl turned for his approach to the concept of a surface. Thus he began, sounding very much like Hilbert: "Let a totality of things, which are called 'the points of the manifold $\mathfrak{F}$,' be given. To every point $\mathfrak{p}$ of the manifold $\mathfrak{F}$ let certain sets of points be defined as 'neighborhoods of $\mathfrak{p}$ on $\mathfrak{F}$' ". The neighborhoods thus define what would now be called a basis for the topology on $\mathfrak{F}$, and Weyl then went on to stipulate further properties of these neighborhoods to make $\mathfrak{F}$ a surface [1913, pp. 17–18]. Thus associated with each neighborhood is a bijective mapping onto an open disk in the plane satisfying certain conditions. Weyl's definitions contained in spirit if not in detail the modern formulation of a 2-dimensional manifold. In addition, $\mathfrak{F}$ is assumed to be arcwise connected. Such an abstract surface $\mathfrak{F}$ then becomes a Riemann surface if it possesses additional structure that makes it possible to speak of an analytic function of points $\mathfrak{p} \in \mathfrak{F}$ [1913, pp. 35–36].[12]

From the viewpoint of this essay what is of interest in Weyl's presentation is that the theory of Riemann surfaces is developed within the broader context of a theory of real two-dimensional manifolds, a theory that readily generalizes to r-dimensional manifolds. This is true in particular of Weyl's extensive discussion of covering surfaces [1913, pp. 47–51], which marks an important step towards the development of a purely topological theory of covering spaces. Likewise his construction of a "universal covering surface" $\widetilde{\mathfrak{F}}$ (as he called it) associated to a given surface $\mathfrak{F}$ is capable of broad generalization. As with Poincaré, the points of $\widetilde{\mathfrak{F}}$ consist of "curves" γ – i.e. continuous mappings $\gamma : [0, 1] \to \mathfrak{F}$ – emanating from a fixed point $\gamma(0) = \mathfrak{p}_0 \in \mathfrak{F}$, where two such curves γ, γ' define the same point of $\widetilde{\mathfrak{F}}$ if $\gamma(1) = \gamma'(1)$ and if they satisfy a condition that nowadays would be described by saying they are homotopic.[13] Thus in effect each point $\tilde{\mathfrak{p}} \in \widetilde{\mathfrak{F}}$ can be identified with an equivalence class $[\gamma]$ of homotopic curves γ which start at $\mathfrak{p}_0$ and end at the same point $\mathfrak{p} = \gamma(1)$ and is thereby covered by $\tilde{\mathfrak{p}}$. The neighborhoods $\widetilde{\mathfrak{U}}$ of $\widetilde{\mathfrak{F}}$ are then defined as follows: "Let γ_0 be a curve in $\mathfrak{F}$ from $\mathfrak{p}_0$ to $\mathfrak{p}$ and $\mathfrak{U}$ a neighborhood of $\mathfrak{p}$. Attach to γ_0 all possible curves γ which start from $\mathfrak{p}$ and lie within $\mathfrak{U}$ and say that the points of $\widetilde{\mathfrak{F}}$ defined by all such curves $\gamma_0 + \gamma$[14] form a 'neighborhood $\widetilde{\mathfrak{U}}$ of $\tilde{\mathfrak{p}}$" ' [1913, p. 51]. With these words Weyl thus gave what is now the customary description of a basis for the topology of the universal covering space.

If we consider the points $\tilde{\mathfrak{p}}_0$ which cover the point $\mathfrak{p}_0$ defined above, each $\tilde{\mathfrak{p}}_0$ corresponds to an equivalence class of curves which begin and end at $\mathfrak{p}_0$ and thus correspond to the elements of Poincaré's fundamental group of $\mathfrak{F}$ at $\mathfrak{p}_0$, $\pi_1(\mathfrak{F}, \mathfrak{p}_0)$. Weyl never mentioned $\pi_1(\mathfrak{F}, \mathfrak{p}_0)$ but spoke instead of the group Γ of covering transformations T (*Decktransformationen*), which he defined to be homeomorphisms of $\widetilde{\mathfrak{F}}$ with the property that for any $\tilde{\mathfrak{p}} \in \widetilde{\mathfrak{F}}$, $T(\tilde{\mathfrak{p}})$ covers the same point as $\tilde{\mathfrak{p}}$, i.e. as is still done today. Considered abstractly, the group Γ, Weyl declared, expresses completely the relation between $\widetilde{\mathfrak{F}}$ and $\mathfrak{F}$, "in so far as it possess an Analysis situs character" [1913, p. 50]. It is not difficult to see that Γ is isomorphic to $\pi_1(\mathfrak{F}, \mathfrak{p}_0)$, for if $T \in \Gamma$ sends the equivalence class $\tilde{\mathfrak{p}} = [\gamma] \in \widetilde{\mathfrak{F}}$ to $\tilde{\mathfrak{p}}' = [\gamma']$, then the characteristic property of T requires that $\gamma(1) = \gamma'(1)$. Hence, curve

[12] For a detailed, critical discussion of Weyl's definitions, and their relation to ideas of Hilbert and Klein, see [Scholz, 1980, pp. 193–198].

[13] Weyl's definitions of simple connectivity and of the universal covering surface avoided using homotopic equivalence and were apparently guided by convenience of application to function theory [1913, p. 47 (n. 2)].

[14] That is, the curve which first follows γ_0 and then γ.

$\gamma' - \gamma$ begins and ends at $\mathfrak{p}_0$ and defines an equivalence class $[\alpha]$ of $\pi_1(\mathfrak{F}, \mathfrak{p}_0)$ such that $T[\gamma] = [\alpha + \gamma]$. The correspondence $T \to [\alpha]$ gives the isomorphism. Presumably Weyl realized all this but preferred not to introduce fundamental groups *per se*.

In view of the use Weyl later made of universal covering spaces in his work on group representations (Section 4), it is helpful to make a few observations at this point because they follow readily from what has been said, although they were not made by Weyl until a decade later – and then in the context of the higher dimensional manifolds defined by Lie groups. To this end, let us suppose that $\mathfrak{F}$ denotes the r-dimensional generalization of Weyl's notion of a surface.

The first observation involves the situation in which $\mathfrak{F}$ is compact in the sense employed by Weyl and his contemporaries, namely every infinite subset of $\mathfrak{F}$ has an accumulation point. Suppose that $\mathfrak{F}$ has finite connectivity, i.e. that $\mathfrak{F}$ has the property that there are N nonhomotopic curves from $\mathfrak{p}_0$ to $\mathfrak{p}$, so that each point $\mathfrak{p} \in \mathfrak{F}$ has N points of $\widetilde{\mathfrak{F}}$ lying over it which is sometimes expressed by saying that $\widetilde{\mathfrak{F}}$ consists of N sheets. Then if $\widetilde{\mathfrak{K}} \subset \widetilde{\mathfrak{F}}$ is infinite, it follows that the set $\mathfrak{K} \subset \mathfrak{F}$ of points covered by points in $\widetilde{\mathfrak{K}}$ must also be infinite. Thus $\mathfrak{K}$ has an accumulation point $\mathfrak{q}$. If $\tilde{\mathfrak{q}}_1, \ldots, \tilde{\mathfrak{q}}_N$ are the points which cover $\mathfrak{q}$, then it follows readily from Weyl's definition of the topology of $\widetilde{\mathfrak{F}}$ that at least one of the points $\tilde{\mathfrak{q}}_1, \ldots, \tilde{\mathfrak{q}}_N$ must be an accumulation point of $\widetilde{\mathfrak{K}}$ so that $\widetilde{\mathfrak{F}}$ is also compact.

The second observation is that if the product $\mathfrak{p}\mathfrak{p}'$ of elements of $\mathfrak{F}$ is defined so that $\mathfrak{F}$ is a group, then it easy to extend the multiplication from $\mathfrak{F}$ to the universal cover $\widetilde{\mathfrak{F}}$ so that the projection mapping p_*, which sends $\tilde{\mathfrak{p}}$ into the point $\mathfrak{p}$ it covers, is a group homomorphism from $\widetilde{\mathfrak{F}}$ onto $\mathfrak{F}$. Indeed, this can be done in more than one way. To see this, choose as the point $\mathfrak{p}_0$ in the above construction of $\widetilde{\mathfrak{F}}$ the identity element $\mathfrak{e}$ of $\mathfrak{F}$. Then if γ, γ' denote two curves with initial point $\mathfrak{e}$, let the product curve $\gamma \cdot \gamma'$ be defined by $(\gamma \cdot \gamma')(t) = \gamma(t)\gamma'(t)$ for all $t \in [0, 1]$. Then $\mathfrak{e}$ is also the initial point of $\gamma \cdot \gamma'$, and if $\tilde{\mathfrak{p}} = [\gamma]$ and $\tilde{\mathfrak{p}}' = [\gamma']$ are two elements of $\widetilde{\mathfrak{F}}$, we may define the product $\tilde{\mathfrak{p}}\tilde{\mathfrak{p}}'$ to be the element of $\widetilde{\mathfrak{F}}$ determined by $\gamma \cdot \gamma'$. With multiplication so defined, $\widetilde{\mathfrak{F}}$ becomes a group as well, and $p_* : \widetilde{\mathfrak{F}} \to \mathfrak{F}$ is a homomorphism. If $\widetilde{\mathfrak{K}} \subset \widetilde{\mathfrak{F}}$ denotes the kernel of this homomorphism then, as noted above, since $\widetilde{\mathfrak{K}}$ consists of the points covering $\mathfrak{p}_0 = \mathfrak{e}$ its *elements* can be identified with those of the fundamental group $\pi_1(\mathfrak{F}, \mathfrak{e})$, although the multiplication defined in $\widetilde{\mathfrak{K}}$ need not coincide with the multiplication defined in $\pi_1(\mathfrak{F}, \mathfrak{e})$. This can be made to happen, however, by defining multiplication in $\widetilde{\mathfrak{F}}$ as follows. Given $\tilde{\mathfrak{p}} = [\gamma]$, $\tilde{\mathfrak{p}}' = [\gamma'] \in \widetilde{\mathfrak{F}}$, let $\tilde{\mathfrak{p}} \times \tilde{\mathfrak{p}}' = [\gamma + \gamma(1) \cdot \gamma']$, where $\gamma + \gamma(1) \cdot \gamma'$ denotes the curve which follows γ from $\mathfrak{e}$ to its end point $\gamma(1)$ and then follows the curve $t \to \gamma(1)\gamma'(t)$ from its starting point at $\gamma(1)\mathfrak{e} = \gamma(1)$ to its end point at $\gamma(1)\gamma'(1)$. (Thus $p_*(\tilde{\mathfrak{p}}\tilde{\mathfrak{p}}') = p_*(\tilde{\mathfrak{p}} \times \tilde{\mathfrak{p}}')$.) Then for $\tilde{\mathfrak{p}} = [\gamma]$, $\tilde{\mathfrak{p}}' = [\gamma'] \in \widetilde{\mathfrak{K}}$, $\tilde{\mathfrak{p}} \times \tilde{\mathfrak{p}}' = [\gamma + \gamma']$, in accordance with the definition of multiplication in $\pi_1(\mathfrak{F}, \mathfrak{e})$. As we shall see in §4, Weyl simply observed that $\widetilde{\mathfrak{F}}$ could be made into a group without pausing to define the multiplication, and either definition sufficed for his purpose.

4. The contributions of Weyl: 1925–1926

The universal covering group of a continuous group in Lie's sense became relevant to Weyl's research interests about a dozen years after his brilliant lectures on Riemann surfaces. By that time (1924) Weyl was deeply involved with Einstein's general theory of relativity, including its mathematical foundations and possible extensions, and this interest led him to an interest in the linear representation of continuous groups [Hawkins, 1998].

For finite groups, the Berlin mathematician Georg Frobenius had developed in 1896–1903 a remarkable theory of linear representations, as Weyl realized. A fundamental theorem of Frobenius' theory was what became known as the *complete reducibility theorem* for finite groups. It may be stated as follows: Let $\varphi : \mathfrak{G} \to \overline{\mathfrak{G}} \subset \mathbf{GL}(\mathcal{V})$ be a representation of $\mathfrak{G}$, i.e. a group homomorphism, where $\mathcal{V}$ is a complex, finite dimensional vector space. Then $\mathcal{V} = \bigoplus_{k=1}^{n} \mathcal{V}_k$ where each vector subspace $\mathcal{V}_k$ is left invariant by the transformations of $\overline{\mathfrak{G}}$ and $\overline{\mathfrak{G}}$ acts irreducibly on each $\mathcal{V}_k$ in the sense that $\mathcal{V}_k$ contains no proper, nontrival subspace which is left invariant by the transformations of $\overline{\mathfrak{G}}$. The action of $\overline{\mathfrak{G}}$ on $\mathcal{V}_k$ defines an irreducible representation $\varphi_k : \mathfrak{G} \to \overline{\mathfrak{G}}_k \subset \mathbf{GL}(\mathcal{V}_k)$, and the complete reducibility theorem shows that the problem of determining all representations of $\mathfrak{G}$ reduces to that of determining the irreducible ones.

Weyl's interest in the scope of the tensor calculus, the fundamental mathematical apparatus of Einstein's theory, led him to seek to establish the complete reducibility theorem for the special linear group $\mathbf{SL}(n, \mathbb{C})$, which is a simple Lie group. In this connection he knew the work of Cartan [1913], which could be interpreted as determining, on the Lie algebra level, all irreducible representations of any complex simple or semisimple group. I say "could be interpreted" because it was not so interpreted by Cartan, who did not relate his work to Frobenius' theory, and saw his result as solving the problem of determining all projective geometries (in the sense of Klein's *Erlanger Programm*) which leave no points, lines, planes, etc., invariant. Both Cartan and Study privately conjectured that for semisimple groups, a geometrical theorem would hold which is tantamount to the complete reducibility theorem for these groups. However, no one had succeeded in giving a proof except, on the Lie algebra level, for $\mathbf{sl}(2, \mathbb{C})$. In the third volume of *Theorie der Transformationsgruppen*, Lie had conjectured the same result would hold for $\mathbf{sl}(n, \mathbb{C})$. Weyl was unaware of this interest in Lie's school in what amounts to complete reducibility theorems. His motivation came from tensor calculus and the possibility of such a theorem for $\mathbf{SL}(n, \mathbb{C})$ was inspired by Frobenius' theory.

It was Frobenius' student, Issai Schur (1875–1941), who provided Weyl with the means to prove a complete reducibility theorem for $\mathbf{SL}(n, \mathbb{C})$. In a paper [1924] on the theory of invariants, Schur called attention to the fact that in [1897], Adolf Hurwitz had showed how to define an integral over the rotation group $\mathfrak{D}_n = \mathbf{SO}(n, \mathbb{R})$ which is what is now called (right) 'translation invariant', i.e. if f is any continuous function defined on $\mathfrak{D}_n$, then $\int_{\mathfrak{D}_n} f(RS)\,\mathrm{d}m(R) = \int_{\mathfrak{D}_n} f(R)\,\mathrm{d}m(R)$ for any $S \in \mathfrak{D}_n$. Making use of this integral in a manner analogous to how Hurwitz had used it – as a continuous analog of summation over a finite group – Schur observed that Frobenius' theory could be extended to $\mathfrak{D}_n$. In particular, it was known that the complete reducibility theorem was equivalent to a theorem due to Maschke, which asserts that if $\overline{\mathfrak{G}} \subset \mathbf{GL}(\mathcal{V})$ has the property that its transformations leave a proper subspace $\mathcal{W} \neq 0$ invariant then there is a complementary subspace $\mathcal{W}'$ also left invariant by $\overline{\mathfrak{G}}$ such that $\mathcal{V} = \mathcal{W} \oplus \mathcal{W}'$. Maschke's theorem follows by defining an inner product on $\mathcal{V}$ by $\langle v, w\rangle = \sum_{R \in \overline{\mathfrak{G}}} (Rv, Rw)$, where (,) denotes any (complex) inner product on $\mathcal{V}$. Clearly by "translation invariance" of group sums $\langle Sv, Sw\rangle = \langle v, w\rangle$. This says that the $S \in \overline{\mathfrak{G}}$ may be regarded as unitary transformations relative to this inner product. Thus given the proper subspace $\mathcal{W}$, the orthogonal complement $\mathcal{W}' = W^{\perp}$ is easily seen to be the desired complementary subspace. As Schur observed, the same proof yields Maschke's theorem, and hence complete reducibility, for $\mathfrak{D}_n$ since we may use the Hurwitz integral to define the inner product $\langle v, w\rangle = \int_{\mathfrak{D}_n} (Rv, Rw)\,\mathrm{d}m(R)$. A crucial property of the Hurwitz integral used here is that since $\mathfrak{D}_n$ is closed and bounded, the Hurwitz integral of any continuous function is finite.

Schur extended the basic theorems of Frobenius' theory to $\mathfrak{D}_n$ in order to solve a counting problem regarding the invariants of $\mathfrak{D}_n$ that had been solved for the classical binary case of $\mathbf{GL}(2, \mathbb{C})$ by Cayley in 1858. But he also called attention to two other ideas that Hurwitz had introduced in 1897. The first is that a translation invariant integral can be defined on any Lie group $\mathfrak{G}$. Thus if $\mathfrak{G}$ is closed and bounded so that the 'Hurwitz integral' (as I will call it) converges for all continuous functions on $\mathfrak{G}$, Frobenius' theory, including the complete reducibility theorem may be extended to $\mathfrak{G}$. The second idea was what Weyl later termed the 'Unitarian Trick'. Hurwitz had used his integral on the rotation group $\mathfrak{D}_n$ to prove that the invariants of $\mathfrak{D}_n$ have a finite basis, i.e. that a finite number of invariants $I_1, \ldots, I_N$ exist such that every invariant is expressible as a polynomial in $I_1, \ldots, I_N$. Hilbert had proved that the invariants of $\mathbf{SL}(n, \mathbb{C})$ have a finite basis, but was unable by his methods to do the same for $\mathfrak{D}_n$ except in the case $n = 3$. Hurwitz, however, showed that his methods would work for $\mathbf{SL}(n, \mathbb{C})$ as well. At first glance this seems doubtful because, as Hurwitz observed, $\mathbf{SL}(n, \mathbb{C})$ is unbounded, which meant that the integral method could not be directly applied. Enter the 'Unitarian Trick'! The group $\mathbf{SU}(n)$ of unitary linear transformations of determinant $+1$ is a subset of $\mathbf{SL}(n, \mathbb{C})$ forming a real, closed and bounded group. Hurwitz's integral method applied to $\mathbf{SU}(n)$ and established the existence of a finite basis. The same would follow for $\mathbf{SL}(n, \mathbb{C})$ if it could be shown that every invariant relative to $\mathbf{SU}(n)$ is actually an invariant relative to the larger $\mathbf{SL}(n, \mathbb{C})$. This reduced, in Schur's version of the 'Unitarian Trick', to showing that if a homogeneous polynomial in n^2 variables t_{jk} vanishes for all t_{jk} defining elements of $\mathbf{SU}(n)$, then it vanishes identically.

The fertile ideas conveyed in Schur's paper provided Weyl with the leading ideas for proving the complete reducibility theorem for $\mathfrak{G} = \mathbf{SL}(n, \mathbb{C})$ and then for any semisimple group. In what follows, for the sake of brevity a group $\mathfrak{G}$ or a Lie algebra $\mathfrak{g}$ will be said to have the *complete reducibility property* (CRP) if the complete reducibility theorem holds for its representations. (For a Lie algebra, a representation is a Lie algebra homomorphism $\varphi : \mathfrak{g} \to \bar{\mathfrak{g}} \subset \mathbf{gl}(V)$.) Weyl realized that the group $\mathfrak{G}_u = \mathbf{SU}(n)$ has the CRP since being closed and bounded, the Hurwitz integral may be used to prove Maschke's theorem. Now a version of the 'Unitarian Trick' was needed to push this result to $\mathbf{SL}(n, \mathbb{C})$. To obtain it Weyl dropped down to the computationally simpler Lie algebra level and easily proved: If $\mathfrak{g}_u = \mathbf{su}(n)$ has the CRP, then so does $\mathfrak{g} = \mathbf{sl}(n, \mathbb{C})$. Basically the proof boiled down to showing that if a *linear* homogeneous function $f = \sum_{jk} a_{jk} t_{jk}$ vanishes for all t_{jk} defining an element $(t_{jk}) \in \mathbf{su}(n)$, then f vanishes identically.

On the basis of Weyl's version of the Unitarian Trick, it might have been tempting for him to jump to the conclusion that he had proved the complete reducibility theorem for the special linear group. That is, to conclude that, since $\mathfrak{G}_u = \mathbf{SU}(n)$ has the CRP, so does its Lie algebra $\mathfrak{g}_u = \mathbf{su}(n)$ and thus by his Unitarian Trick so does $\mathfrak{g} = \mathbf{sl}(n\mathbb{C})$ and hence $\mathfrak{G} = \mathbf{SL}(n, \mathbb{C})$. Indeed, Lie and his students usually never made any clear distinction in their mathematical discourse between Lie groups and Lie algebras. But Weyl realized that a representation of a Lie algebra such as $\mathfrak{g}_u$ need not correspond to a representation of the Lie group from which it was derived. Hence even though $\mathfrak{G}_u$ has the CRP, $\mathfrak{g}_u$ need not have it, which means the unitarian trick cannot be directly applied.

Weyl seems to have arrived at this insight by perceiving an analogy with analytic continuation and covering surfaces, a subject with which he was well acquainted (Section 3). Given a representation $X \to \varphi(X)$ of the Lie algebra $\mathfrak{g}_u$, Weyl reasoned [1925, pp. 561–

562], one obtains by integration, following Lie, a matrix $\Phi(x)$ for every $x \in \mathfrak{G}_u$ in a neighborhood $\mathfrak{U}$ of the identity element e. For $x_0 \in \mathfrak{U}$ consider the translated neighborhood $x_0\mathfrak{U}$ of x_0. Since all elements in this neighborhood are of the form $x_0 u$ where $u \in \mathfrak{U}$ the representation can be continued by setting $\Phi(x_0 u) = \Phi(x_0)\Phi(u)$. Repeating this process of continuation could lead to a multi-valued representation on $\mathfrak{G}_u$ which first becomes single-valued on a covering manifold over $\mathfrak{G}_u$. Referring to the universal covering surface of uniformization theory (Section 3), Weyl declared: "This universal covering manifold $[\widetilde{\mathfrak{G}}_u]$ over $[\mathfrak{G}_u]$ is the true abstract group whose representations are under consideration; $[\mathfrak{G}_u]$ is only *one* of its representations, and indeed it is [unfaithful] ... when the covering manifold is many-sheeted." Weyl did not pause to explain what he meant by this "true abstract group," leaving it to his readers, who were assumed to know about covering manifolds from uniformization theory, to see how to define a group multiplication in $\widetilde{\mathfrak{G}}_u$. As was indicated in Section 3, this is not difficult to do.

A simple but important example of a covering group which was known to Weyl had been implicit in Hilbert's proof of the finite basis theorem for the invariants of $\mathfrak{D}_3 = \mathbf{SO}(3, \mathbb{R})$, the rotation group of ordinary space. Hilbert had used the well-known connection between rotations in ordinary space and quaternions. That is, if $\mathbf{a} = a_0 + a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}$ is a real quaternion such that $\sum a_i^2 = 1$, then the transformation $(x_1, x_2, x_3) \to (y_1, y_2, y_3)$ of $\mathbb{R}^3$ defined by $\mathbf{y} = \mathbf{a}^{-1}\mathbf{x}\mathbf{a}$, where $\mathbf{x} = x_1\mathbf{i} + x_2\mathbf{j} + x_3\mathbf{k}$, $\mathbf{y} = y_1\mathbf{i} + y_2\mathbf{j} + y_3\mathbf{k}$ is an orthogonal transformation $T(\mathbf{a})$ with determinant $+1$. It is readily seen that the above type quaternions $\mathbf{a}$ form a group $\widetilde{\mathfrak{D}}_3$ under quaternion multiplication and that $\Psi : \beta \to T(\beta)$ is a homomorphism from $\widetilde{\mathfrak{D}}_3$ to $\mathfrak{D}_3$ with kernel consisting of -1 and $+1$. Topologically, $\widetilde{\mathfrak{D}}_3$ looks like the three-sphere S^3 which is simply connected. Hence $\widetilde{\mathfrak{D}}_3$ is the universal covering group for $\mathfrak{D}_3$. As Weyl realized, this shows that $\mathfrak{D}_3$ is not simply connected and that $\widetilde{\mathfrak{D}}_3$ is a two sheeted covering [1925, p. 598]. Here then was a well-known example showing that a group need not coincide with its covering group.

To apply the "integration method" of Hurwitz to establish the CRP for a Lie algebra $\mathfrak{g}$ of a group $\mathfrak{G}$, it was thus necessary that the covering group $\widetilde{\mathfrak{G}}$ be compact. (In the above example, of course, both $\mathfrak{D}_3$ and $\widetilde{\mathfrak{D}}_3$ are compact.) As noted at the conclusion of Section 3, by virtue of his work on Riemann surfaces, Weyl could see that the compactness of $\widetilde{\mathfrak{G}}$ would follow from that of $\mathfrak{G}$ provided $\widetilde{\mathfrak{G}}$ is a covering with finitely many sheets (as occurs in the above example). Weyl realized that a universal covering manifold of a compact group need not be compact. One has only to consider, as Weyl did, the rotation group of the plane, which in terms of complex variables can be thought of as the group $\mathfrak{G}$ consisting of all transformations $T_\theta : z \to e^{i\theta}z$ of the complex variable z. As a topological object $\mathfrak{G}$ can be identified with a circle and its covering group $\widetilde{\mathfrak{G}}$ thought of as an infinite spiral. Hence $\widetilde{\mathfrak{G}}$ is not compact even though $\mathfrak{G}$ is. Using this simple example, he defined a representation of the Lie algebra $\mathfrak{g}$ of $\mathfrak{G}$ which generates a many-valued representation on $\mathfrak{G}$ that first becomes single-valued on $\widetilde{\mathfrak{G}}$. In other words, the representation of $\mathfrak{g}$ does not correspond to one of $\mathfrak{G}$ but only of $\widetilde{\mathfrak{G}}$. He also pointed out that $\widetilde{\mathfrak{G}}$ fails to have the CRP.

Weyl could thus see something that had eluded Cartan, who, like Lie, tended to deal with groups on the Lie algebra or group germ level: the irreducible representations of the infinitesimal group $\mathfrak{g}$ associated to a given group $\mathfrak{G}$ do not necessarily correspond to representations of $\mathfrak{G}$ but rather to its simply connected universal covering group $\widetilde{\mathfrak{G}}$. This meant that in order to utilize his Unitarian Trick, Weyl had to show that the covering group $\widetilde{\mathfrak{G}}_u$ of $\mathfrak{G}_u = \mathbf{SU}(n)$ is compact. He confirmed that this is the case by proving that $\mathbf{SU}(n)$ is simply connected and hence its own covering group. Now it did follow that $\mathfrak{g}_u = \mathbf{su}(n)$

has the CRP and so (by the Trick) $\mathfrak{g} = \mathbf{sl}(n, \mathbb{C})$ has the CRP. Now it followed that $\mathfrak{G} = \mathbf{SL}(n, \mathbb{C})$ had the CRP as well. In his first announcement of this result in the Göttingen *Nachrichten*, Weyl emphasized that: "Here analysis situs plays a decisive role; the real impediment to the universal application of Hurwitz's method (which would lead to the thoroughly false theorem that every linear group is fully reducible) lies in the realm of topology: the noncompactness of most group manifolds" [1924a, p. 464].

By the time of the *Nachrichten* note Weyl realized that his proof that $\mathbf{SL}(n, \mathbb{C})$ has the CRP could be modified to cover the other main classes of simple groups, the symplectic groups $\mathfrak{G} = \mathbf{Sp}(2n, \mathbb{C})$ and the special orthogonal groups $\mathfrak{G} = \mathbf{SO}(n, \mathbb{C})$. In both cases $\mathfrak{G}_u$ is defined to consist of all $T \in \mathfrak{G}$ which are unitary, and the analog of the Unitarian Trick is proved by similar considerations. For $\mathfrak{G} = \mathbf{SP}(2n, \mathbb{C})$ Weyl showed that that $\mathfrak{G}_u$ is simply connected, and for $\mathfrak{G} = \mathbf{SO}(n, \mathbb{C})$ he showed $\mathfrak{G}_u = \mathbf{SO}(n, \mathbb{R})$, $n > 2$, is doubly connected [1925, pp. 588ff., 598ff.]. Thus in both cases the covering group $\widetilde{\mathfrak{G}}_u$ is compact and the complete reducibility theorem follows. Shortly after his *Nachrichten* note was submitted, Weyl was able to announce a proof of the complete reducibility theorem for any semisimple group [1924b]. He had discovered that the ideas behind his proof for the classical groups could be extended to this more general context. The basic ideas are sketched below.

Let $\mathfrak{G}$ denote a complex semisimple group with Lie algebra $\mathfrak{g}$. Then $\mathfrak{G}$ does not necessarily consist of linear transformations, but, as anyone acquainted with Lie's theory would realize, one can consider the adjoint representation $\mathrm{Ad} : \mathfrak{G} \to \mathfrak{A} \subset \mathbf{GL}(\mathfrak{g})$, and work with the linear group $\mathfrak{A}$. Because $\mathfrak{g}$ is semisimple, the differential of the adjoint representation is an isomorphism from $\mathfrak{g}$ onto the Lie algebra $\mathfrak{a}$ of $\mathfrak{A}$. To get a real subgroup of $\mathfrak{A}$ which might play the role of the groups $\mathfrak{G}_u$ utilized in dealing with the classical groups, Weyl considered what are now called real forms of $\mathfrak{g}$. That is, it was clear from the structure theory of semisimple $\mathfrak{g}$ that bases $X_1, \ldots, X_r$ for $\mathfrak{g}$ exist for which $[X_i, X_j]$ is a real linear combination of $X_1, \ldots, X_r$ for all i and j. The real span of such a basis then defines a real Lie algebra $\mathfrak{g}_r$. Let $\mathfrak{a}_r \subset \mathfrak{a}$ denote the image of $\mathfrak{g}_r$ under the differential of the adjoint representation and let $\mathfrak{A}_r \subset \mathbf{GL}(\mathfrak{g})$ denote the corresponding connected real Lie group. Of course to play the role of $\mathfrak{G}_u$, $\mathfrak{A}_r$ must be compact.

Weyl knew from Cartan's papers that $\mathfrak{A}$ leaves invariant the nonsingular quadratic Killing form $\psi(X) = \mathrm{tr}\, X \circ \mathrm{tr}\, X$, $X \in \mathfrak{g}$. Thus the linear transformations of $\mathfrak{A}_r$ take the real vector space $\mathfrak{g}_r$ into itself and leave $\psi(X)$ invariant. It is easily seen that the group of all real linear transformations which take a nonsingular quadratic form ψ into itself is bounded, and therefore compact, precisely when ψ is definite, so that the group is the familiar real orthogonal group. Since $\mathfrak{A}_r$ is a subgroup of this group it will also be compact since it is closed. Thus to obtain a compact $\mathfrak{A}_r$, Weyl proved there exists a real form of $\mathfrak{g}$, $\mathfrak{g}_r = \mathfrak{g}_u$, such that the Killing form restricted to $\mathfrak{g}_u$ is negative definite. Now $\mathfrak{g}_u$ is called a *real compact form of* $\mathfrak{g}$. Corresponding to the real compact form $\mathfrak{g}_u$ in the above manner is a real linear group $\mathfrak{A}_u$ which is compact.

Thus $\mathfrak{A}_u$ is the candidate to fill the role played by the unitary transformations in $\mathfrak{G}$ when $\mathfrak{G}$ is one of the classical groups. As in Weyl's proofs for those groups, in order to prove that $\mathfrak{g}$ has the CRP, two theorems remain to be proved. For the general semisimple case they take the following form:

FINITE CONNECTIVITY THEOREM. *If $\mathfrak{G}$ is semisimple, then $\mathfrak{A}_u$ has finite connectivity, i.e. $\widetilde{\mathfrak{A}}_u$ is a finite-sheeted covering of $\mathfrak{A}_u$.*

GENERAL UNITARIAN TRICK. *If* $\mathfrak{g}$ *is semisimple and* $\mathfrak{g}_u$ *has the CRP, then so does* $\mathfrak{g}$.

Weyl proved the Finite Connectivity Theorem by showing that among all the curves in $\mathfrak{A}_u$ which start at the identity element and end at another suitably chosen fixed point there are $N \leqslant (r-l)!$ which are homotopically inequivalent, where r and l denote, respectively, the dimension and rank of $\mathfrak{G}$. This means that the universal covering group $\widetilde{\mathfrak{A}}_u$ has at most N points covering a point of $\mathfrak{A}_u$ and so inherits the compactness of $\mathfrak{A}_u$ as indicated in Section 3. Once the compactness of $\widetilde{\mathfrak{A}}_u$ is established it follows via the Hurwitz integral technique that not only it, but also its Lie algebra $\mathfrak{a}_u$ has the CRP. Since $\mathfrak{a}_u$ is isomorphic to $\mathfrak{g}_u$, this means that $\mathfrak{g}_u$ has the CRP and so by the General Unitarian Trick so does $\mathfrak{g}$. Once the CRP is established for $\mathfrak{g}$, it follows for $\mathfrak{G}$. (It is the reverse conclusion that fails, as Weyl was first to observe, when $\mathfrak{G}$ is not simply connected.)

The use of topology by Weyl was not limited to his theorems about the finite connectivity of semisimple groups. He had used it earlier to prove what amounts to the theorem in present-day Lie group theory that all maximal toral subgroups of a Lie group are conjugate, a result he needed to prove the Finite Connectivity Theorem. When the underlying semisimple group $\mathfrak{G}$ is one of the classical simple groups, Weyl explained, "this theorem coincides with known algebraic facts" [1925, p. 629]. For example, in the case of the special linear group, it follows readily from the known fact of linear algebra that a unitary matrix U can be "unitarily diagonalized," i.e. $U = V^{-1}DV$, where V is unitary and D is diagonal. To establish the general theorem, however, he had to resort to more than linear algebra: "Here [the theorem] will be established generally by means of the method of continuity" [1925, p. 629]. By the "method of continuity" Weyl meant a "topological method" a phrase he also could have used as can be seen from the Weyl quotations given earlier. His choice here to speak of the "method of continuity," however, imbued his remark with historical resonance, for, as we saw in §3, that term had evolved out of the work of Klein and Poincaré on the former's fundamental theorem of automorphic functions where topological reasoning had proved to be indispensable – and in need of further development. In the course of working on his lectures on Riemann surfaces, Weyl had occasion to absorb the newly developing topology (especially as found in Brouwer's papers), and now, a dozen years later, he saw that Lie's theory, like complex function theory earlier, was in need of topological concepts and reasoning to further its development.

Weyl's work revealed the important role played by the real compact Lie group $\mathfrak{G} = \widetilde{\mathfrak{A}}_u$ in the representation theory of complex semisimple Lie groups. In effect, the study of the representations of semisimple groups reduced, by the logic of the Unitarian Trick, to the study of the representations of $\mathfrak{G}$, which, being compact, was equipped with a translation invariant integral. The representation theory of real compact groups thus became central to Weyl's quest for a Lie group analog to Frobenius' theorem that every irreducible representation of a finite group $\mathfrak{G}$ is obtained in the complete reduction of the regular representation of $\mathfrak{G}$. The regular representation of $\mathfrak{G}$ is defined as follows. Regard the n elements of $\mathfrak{G}$ as forming the basis for an n-dimensional complex vector space $\mathcal{V}$, and write its elements in the form $x = \sum_{s\in\mathfrak{G}} x(s)s$, $y = \sum_{t\in\mathfrak{G}} y(t)t$. Using the multiplication of $\mathfrak{G}$ we can make $\mathcal{V}$ into a linear associative algebra, called the group algebra of $\mathfrak{G}$, by setting

$$x * y = \sum_{s,t\in\mathfrak{G}} x(s)y(t)st \overset{u=st}{=} \sum_u \Big[\sum_t x(ut^{-1})y(t)\Big]u. \tag{3}$$

The (right) regular representation $s \to T(s) \in \mathbf{GL}(V)$ is then defined by

$$T(s)y = y * s = \sum_{t \in \mathfrak{G}} y(t)ts \overset{u=st}{=} \sum_{u \in \mathfrak{G}} y(us^{-1})u. \tag{4}$$

Frobenius' theorem says that in the complete reduction of $T(s)$, every irreducible representation of $\mathfrak{G}$ occurs as often as its degree.

The analog of (4) is obtained for the real compact group $\mathfrak{G}$ by identifying y with its coefficients $y(s)$ and thinking of the latter as a continuous complex-valued function of $s \in \mathfrak{G}$. (Nowadays one considers y as an element of $L^2(\mathfrak{G})$, but Weyl choose $C(\mathfrak{G})$ because Hilbert's theory was then developed in that context.) This replaces the finite-dimensional vector space by the infinite-dimensional space $C(\mathfrak{G})$ of all continuous functions on $\mathfrak{G}$. In the regular representation, $T(s)$ is now the linear operator on $C(\mathfrak{G})$ which sends $y(t)$ to the function $\tilde{y}(u) = y(us^{-1})$. The only linear operators on $C(X)$, with X compact, that had been studied were the integral operators $Ty = \int_X K(s,t)y(t)\,dt$ of Hilbert where $X = [a, b]$ but with the realization that the theory could be extended to more general closed and bounded subsets of $\mathbb{R}^n$. The operator $T(s)$ is not one of these, but Weyl saw how to reduce the problem of decomposing $T(s)$ into its irreducible components to an application of Hilbert's theory with $X = \mathfrak{G}$.

Weyl realized that the continuous analog of multiplication in the group algebra (3) is what has since become known as the convolution of functions $x, y \in C(\mathfrak{G})$: $(x * y)(s) = \int_{\mathfrak{G}} x(st^{-1})y(t)\,dm(t)$, where the integral is that of Hurwitz. For any $x \in C(\mathfrak{G})$, the operator $A_x y = x * y$ is thus an integral operator with kernel $K(s,t) = x(st^{-1})$. (This does not work for $T(s) = y * s$ since the group element s, as a function, is zero everywhere except at s where it takes the value 1; consequently there is no viable analog of 's' in $C(\mathfrak{G})$ – or in $L^2(\mathfrak{G})$.) The operator A_x is not hermitian symmetric, so that the strongest results of Hilbert's theory do not apply, but a well known technique of the theory was to consider the hermitian symmetric operator $A_x A_x^H$, where A_x^H denotes the hermitian (or conjugate) transpose of A_x with kernel $L(s,t) = \overline{K(t,s)} = \overline{x(ts^{-1})}$. Weyl did this and was able to take the idea one step further because of the special nature of the kernels of his operators. That is, it turns out that $A_x A_x^H = A_z$, where $z = x * \tilde{x}$ and $\tilde{x}(s) = \overline{x(s^{-1})}$. In other words, if we take any $x \in C(\mathfrak{G})$ and set $z = x * \tilde{x}$ we obtain a hermitian symmetric, definite integral operator A_z with kernel $z(st^{-1})$ to which the most extensive results of Hilbert's theory can be applied.

For example, if $\lambda \neq 0$ is an eigenvalue of A_z then $\lambda > 0$ and the space $\mathcal{E}_\lambda$ of solutions to $(A_z - \lambda I)\varphi = 0$ is a finite-dimensional subspace of $C(\mathfrak{G})$. The relevance of A_z to the decomposition of the regular representation $T(s)$ stems from the fact that $T(s)$ commutes with A_z and so takes $\mathcal{E}_\lambda$ into itself thereby determining a finite-dimensional representation of $\mathfrak{G}$. In this way, Weyl proceeded to decompose the regular representation into its irreducible constituents. And by using Hilbert's theory to do it, he in effect, inaugurated Fourier or harmonic analysis on groups. That is, for a hermitian symmetric operator, the eigenfunctions φ_n of nonzero eigenvalues can be chosen to form an orthonormal set with respect to the inner product $\langle x, y \rangle = \int_a^b x(s)\overline{y(s)}\,ds$ and were regarded by Hilbert as generalizations of the sine and cosine functions of ordinary Fourier analysis and the 'Fourier coefficients' $a_n = \langle x, \varphi_n \rangle$ and 'Fourier series' $\sum_n a_n \varphi_n$ considered. Of particular interest in Hilbert's theory was the question as to when the φ_n form a complete set. As formulated by Hilbert this meant that Parseval's equality, $\int_a^b |x(s)|^2 \, ds = \sum_n |a_n|^2$ should hold for every $x \in C[a, b]$. Nowadays one thinks of completeness as saying that for all $x \in L^2[a, b]$, $x = \sum_n a_n \varphi_n$ in the sense of the L^2 norm.

In the course of decomposing the regular representation, Weyl obtained, with the assistance of his student F. Peter, what is now called the Peter–Weyl completeness theorem [1927]. Briefly put, it says that if $s \to E^{(\mu)}(s) = (u_{jk}^{(\mu)}(s))$, $\mu = 1, 2, \ldots$, is a full set of inequivalent irreducible matrix representations of $\mathfrak{G}$, which without loss of generality can be assumed unitary, then, suitably normalized, the functions $u_{jk}^{(\mu)}(s)$ form a complete set. As the most important application of this theorem, the authors singled out the following result: Suppose for $s, t \in \mathfrak{G}$ $E^{(\mu)}(s) = E^{(\mu)}(t)$ for all μ; then $s = t$ [1927, p. 74]. No explanation was given as to why this result was so important, although it became a fundamental tool in subsequent developments. In particular, as we shall see in Section 5, it provided the key to John von Neumann's solution to Hilbert's fifth problem for compact groups. I will refer to it there as the point separation corollary to the Peter–Weyl theorem.

In this connection, it is important to realize that although the Peter–Weyl paper was motivated by Weyl's interest in an analog of Frobenius' regular representation theorem as it applied to the particular compact group $\mathfrak{G} = \tilde{\mathfrak{A}}_u$, it was realized that the reasoning only depended upon having a real compact 'continuous group' on which a translation invariant integral was defined. Hurwitz had showed how to define such an integral, but his definition depended upon the differentiability of the group operations: $dm(t) = \psi(t) \, dt$, where the density function $\psi(t)$ was the Jacobian of the transformation $s \to st$. For this reason at the outset the authors posited $\mathfrak{G}$ to be any compact continuous group such that "Lie's infinitesimal-conceptual apparatus is applicable" to $\mathfrak{G}$ [1927, p. 58]. The Peter–Weyl paper thus suggested the problem of defining a translation invariant integral on a compact continuous group without resorting to a differentiable Lie group structure – a problem solved by Haar in [1933].

5. The influence of Weyl's papers

Beyond the topological considerations described in Section 4, Weyl directly contributed little more to the topological study of continuous groups. But indirectly, through the impact of his work on others, he, more than anyone else, triggered the surge of activity in this area that occurred during the three decades after his papers were published. In this concluding section I consider briefly the main links between Weyl's work and the three major research programs which contributed to that surge: (1) the topology of Lie groups; (2) Hilbert's fifth problem; (3) harmonic analysis on groups.

With his finite connectivity theorem, Weyl had made the earliest contribution to (1) but as can be seen from the introductory overview, Schreier had gone much further in this direction. Nonetheless it was Weyl's work, by virtue of the great advance it made in Lie's theory, that made Lie's foremost successor, Élie Cartan, a believer in the global topological approach to Lie's theory. Between 1894 and 1924, Cartan had advanced Lie's theory profoundly in many respects, but all of his work, like Lie's, was tacitly on the local level, based upon algebraic considerations involving the Lie algebra of the group. After Cartan saw Weyl's 1924 announcements of the complete reducibility theorem – but before he had any idea of how Weyl proved it for semisimple groups in general – he devised his own proof [1925]. It drew on many ideas which were mentioned in Weyl's 1924 announcements such as the idea of utilizing the Hurwitz integral in conjunction with the unitarian trick, but Cartan avoided Weyl's idea of introducing the universal covering group and proving its compactness. The advantage of his approach, Cartan explained in a letter to Weyl, was that complete reducibility could be proved "without being obligated to devote oneself to studies of analysis situs, which are always delicate".[15] However, as Borel has pointed out [1986, p. 76 (n.16)], Cartan's proof takes for granted results which seem to require the sort of topological reasoning he wished to avoid. In his reply to Cartan's letter Weyl defended his use of topology explaining that the "consideration of Analysis situs is very simple and applies to all semisimple groups without distinguishing cases. This approach lies closer to my whole way of thinking than your more algebraic method, which at the moment I only half understand".[16]

After Cartan was able to read Weyl's detailed paper [1925], which shows how the topological approach can be applied to semisimple groups in general and also shows the benefits of such an approach, he became convinced of the merits of the global topological approach to Lie groups. This can be seen in his paper "The geometry of simple groups", where he declared: "The study presented here is no longer local; rather it concerns the properties of space which depend on *Analysis situs* ..." [1927, p. 210].

Cartan's paper is primarily a study of the connections between Riemannian geometries and Lie groups, to which now a global, topological approach is applied. But in the introductory part of this lengthy paper, he made some of his first contributions to the topology of Lie groups by developing the implications of Weyl's work and filling in some of the details. Thus he showed how to construct the covering group $\tilde{\mathfrak{G}}$ of $\mathfrak{G} = \mathfrak{A}_u$. He defined the multiplication in $\tilde{\mathfrak{G}}$ in the second of the two ways indicated in Section 3 so that the kernel $\tilde{\mathfrak{K}}$ of the covering homomorphism $p_* : \tilde{\mathfrak{G}} \to \mathfrak{G}$ could be identified with the fundamental group $\pi_1(\mathfrak{G})$.[17] Cartan also pointed out that $\tilde{\mathfrak{K}}$ is always in the center of $\tilde{\mathfrak{G}}$ so that $\pi_1(\mathfrak{G})$ is abelian. Schreier had of course already proved a more general result, but his work was not yet known to Cartan. For Cartan it was an easy consequence of Weyl's finite connectivity theorem: Since $\tilde{\mathfrak{K}}$ is normal, for any fixed $k \in \tilde{\mathfrak{K}}$ the continuous function $f(x) = xkx^{-1}$ maps $\tilde{\mathfrak{G}}$ into $\tilde{\mathfrak{K}}$; since $f[\tilde{\mathfrak{G}}]$ is connected and $\tilde{\mathfrak{K}}$ is finite, it must be a single point; and since $f(x) = k$ when x is the identity element of $\tilde{\mathfrak{G}}$, $xkx^{-1} = k$ for all x and so k is in the center of $\tilde{\mathfrak{G}}$. Cartan also computed $\pi_1(\mathfrak{G})$ explicitly for each simple type.

The above described results of Cartan's were just elaborations of the topological aspects of Weyl's work, but they are indicative of his new interest in the topology of Lie groups, which he continued to pursue. Indeed in the immediately following years, it was Cartan who through his publications insured a continuing interest in the topology of Lie groups. As Samelson has written in his survey article [1952, p. 6], interest in the topology of Lie groups "is due, besides to H. Weyl, above all to É. Cartan, who in a long series of papers ... came back to this subject again and again, pointed out its importance, made a thorough study of many special cases and went on from there to prove or predict many general results". In particular, Cartan published the first monograph on the topology of Lie groups [1930]

---

15 Letter dated 12 October 1925. The quoted passage reads: "sans être obligé de se livrer à des études d'analysis situs toujours délicates." The original letter is located in the archives of the library of the Eidgenösische Technische Hochschule (ETH) Zürich [Hs 91: 501].

16 Weyl's letter, dated 22 March 1925, is in the possession of Henri Cartan.

17 Cartan called $\pi_1(\mathfrak{G})$ the 'connection group' of $\mathfrak{G}$ because the term 'fundamental group' had another meaning in his group-theoretic approach to geometry.

and initiated [1928] the program of determining the Betti numbers (Poincaré polynomials) of Lie groups, conjecturing in the process the theorems established by de Rham in his important memoir [1931] on the differential topology of manifolds.

One final remark regarding Cartan. As we saw in Section 4, it was the goal of proving the complete reducibility theorem for a semisimple Lie algebra $\mathfrak{g}$ that had led Weyl to his approach involving the compact group $\mathfrak{G} = \mathfrak{A}_u$ and its compact covering group. In 1935 a completely algebraic proof that semisimple $\mathfrak{g}$ have the complete reducibility property was given by Casimir and van der Waerden so that for its original goal Weyl's approach was no longer needed. But, as Cartan showed repeatedly in his own work, Weyl's ideas were extremely fruitful for the topological study of Lie groups. In particular, Cartan drew from Weyl's work the implication that compact groups were the key to the topological study of more general Lie groups. In this connection I will simply mention Cartan's theorem [1936, p. 245] to the effect that any noncompact simply connected Lie group is homeomorphic to a Cartesian product of compact simple groups and a Euclidean space. An analogous result was later established for all connected Lie groups by Malcev and Iwasawa. A good idea of the plethora of results and techniques on the topology of Lie groups in the thirty years after Weyl's paper [1925] can be obtained from the above survey article by Samelson and the sequel by Borel [1955]. Vivid, first-hand accounts of the early influence of Cartan's work on topological developments can be found in essays by de Rham [1981, pp. 641–664].

Research on more general topological groups was also greatly stimulated by John von Neumann's solution to Hilbert's fifth problem for compact groups [1933]. In the language of manifolds, the question behind the problem is: given a locally Euclidean group $\mathfrak{G}$ which acts on an n-dimensional manifold $\mathcal{M}$, is it possible to choose the local coordinates in $\mathfrak{G}$ and $\mathcal{M}$ so that $x \to a \cdot x$ and $(a, b) \to ab^{-1}$, $a, b \in \mathfrak{G}$, $x \in \mathcal{M}$ are continuously differentiable (or analytic)? Von Neumann showed the answer is affirmative in the stronger sense provided $\mathfrak{G}$ is compact and acts transitively on $\mathcal{M}$, i.e. for any $x, y \in \mathcal{M}$ there is a $T \in \mathfrak{G}$ such that $Tx = y$. (He also gave an example of a one-dimensional noncompact group acting intransitively on a two-dimensional manifold for which Hilbert's question has a negative answer). Von Neumann's theorem was the first affirmative answer to Hilbert's problem for an extensive class of groups and offered hope of a complete resolution of the question. Von Neumann was probably the first to divide Hilbert's problem into two. The first problem concerns the case in which $\mathcal{M} = \mathfrak{G}$, where $a \cdot x = ax$, $a, x \in \mathfrak{G}$. Here $\mathfrak{G}$ acts (simply) transitively on itself. This will be referred to as Hilbert's problem for abstract groups and the second problem as Hilbert's problem for transformation groups. Von Neumann first solved the former problem for $\mathfrak{G}$ compact and then used the result to solve the second problem. As we shall see, it was in solving the first problem that he utilized the Peter–Weyl theory.

Von Neumann's first academic position was as an instructor (*Privatdozent*) at the University of Berlin, where Issai Schur was on the faculty. We saw in Section 4 that Schur's work on the representations of the rotation group had been a great inspiration to Weyl. Schur went on to study the representations of the real full orthogonal group as well. In all his work Schur avoided Lie algebra techniques; and the representations he studied were simply assumed to be continuous, i.e. the matrix elements of the representation were assumed to be continuous functions of the elements in the group being represented. By contrast, in their work on representations of a group $\mathfrak{G}$, both Weyl and Cartan made critical use of the existence of what would now be called the differential of the representation, namely the corresponding representation of the Lie algebra $\mathfrak{g}$. Assumptions about the differentiability of the representations of $\mathfrak{G}$ was thus fundamental to their approach. In a letter to Weyl in 1924, when the latter was in the process of creating the results presented in [1925], Schur had emphasized as one of the virtues of his approach that differentiability is not assumed of his representations only integrability in the sense of continuity.[18] Probably with the encouragement of Schur, von Neumann considered the question of whether the continuity of a representation of a *linear* group implied its differentiability. Von Neumann was able to give an affirmative answer in [1927], but his strongest results were given later in [1929].

In [1929] von Neumann considered a group of linear transformations $\mathfrak{G} \subset \mathbf{GL}(n, \mathbb{C})$ for which a continuous representation $A \rightarrow D(A)$ is given. He showed that if $\overline{\mathfrak{G}}$ is the closure of $\mathfrak{G}$ with respect to $\mathbf{GL}(n, \mathbb{C})$, the latter being considered in the relative topology it inherits as a subset of $\mathbb{R}^{2n^2}$, then D has a unique extension to $\overline{\mathfrak{G}}$ and $D(A)$ is a real analytic function of the a_{ij}, where $A = (a_{ij})$ [1929, p. 543 (Satz II)]. From the standpoint of the rigorous theory of functions of a real variable, of course, it was well known that continuity does not imply differentiability, but as von Neumann noted, the existence of the group operation prevents this sort of pathology and thus removes the critical objections raised (possibly by Schur) against the study of groups of linear transformations and their representations based upon consideration of Lie algebras. "Above all, the results of Weyl on the representation of semisimple groups are thereby freed from their far reaching differentiability assumptions" [1929, p. 511].

Von Neumann's theorem was based upon another which is not only of interest in its own right but historically consequential as well [1929, pp. 532–533 (Satz I)]. He showed that $\overline{\mathfrak{G}}$ possesses a real r-dimensional Lie algebra $\mathfrak{g}$ and that if $r > 0$ the exponential mapping from $\mathfrak{g}$ to $\overline{\mathfrak{G}}$ gives an injective mapping from a neighborhood of $0 \in \mathfrak{g}$ onto a neighborhood of the identity element in $\overline{\mathfrak{G}}$ which makes the matrix coefficients of the elements of $\overline{\mathfrak{G}}$ in this neighborhood analytic functions of r variables and implies the matrix multiplication in this neighborhood is likewise analytic. Thus the component of $\overline{\mathfrak{G}}$ containing the identity element "forms a continuous group in the usual sense of the word, with all the expected analyticity properties" [1929, p. 533].

Perhaps von Neumann did not consider the above result as a significant enough answer to Hilbert's question to be worth emphasizing as such, but, stimulated by a paper by Alfréd Haar (1885–1933), he saw how he could use it to obtain the far more significant solution to Hilbert's problem published in [1933]. Both Haar's paper and von Neumann's application of it were motivated by the Peter–Weyl paper [1927]. In the case of Haar, the motivation came from the fact, noted in §4, that the entire Peter–Weyl theory for real compact Lie groups $\mathfrak{G}$ depended only on $\mathfrak{G}$ being a Lie group so that the translation invariant integral of Hurwitz could be introduced. The entire theory would go through for any compact topological group $\mathfrak{G}$ on which a translation invariant integral could be defined. Haar showed the latter could be defined for topological groups $\mathfrak{G}$ which were only assumed to be separable, locally compact metric spaces. As von Neumann pointed out [1934, p. 445 (n. 2)] with a reference to Hausdorff's *Megenlehre* [1927, p. 230], by virtue of the metrization theorem of Urysohn [1925] as sharpened by Tychonoff [1926], it was known that regular topological spaces with a countable basis were precisely the topological spaces homeomorphic to separable metric spaces, so that Haar's results applied to locally compact groups with these topological properties. (Later, using Tychonoff's theorem that the product of compact

[18] Letter to Weyl dated 10 November 1924. The original is located in the library archives of the Eidgenössische Technische Hochschule Zürich [Hs 91:734].

spaces is compact, Weil showed that Haar's ideas could be used to establish the existence of a translation invariant integral on any locally compact group [1940, pp. 33–38].)

As an application Haar briefly indicated how "the beautiful theory of F. Peter and H. Weyl" could be extended to the case in which $\mathfrak{G}$ is assumed, in addition, to be compact. Haar had given a copy of a draft of his paper to von Neumann, who saw how to use Haar's results to obtain his solution to Hilbert's problem for compact groups. The solution was published as [1933] in *Annals of Mathematics*, which von Neumann edited, immediately following Haar's paper [1933]. It is easy to see that compact, locally Euclidean groups $\mathfrak{G}$ satisfy the above topological conditions for the existence of the Haar integral. Thus $\mathfrak{G}$ admits a translation invariant integral such that $\mathfrak{G}$ has finite measure, and the Peter–Weyl Theory goes over to $\mathfrak{G}$, including, in particular, the point separation corollary to the Peter–Weyl theorem (§4). On a heuristic level, the corollary says that the infinite direct sum of all the inequivalent unitary representations is faithful; and von Neumann saw (in a manner sketched below) how to use the compactness of $\mathfrak{G}$ to obtain a finite partial sum which is also faithful. If $D : \mathfrak{G} \to \overline{\mathfrak{G}} \subset \mathrm{GL}(n, \mathbb{C})$ denotes this (continuous) faithful representation, then the compactness of $\mathfrak{G}$ implies that D is a homeomorphism. Hence $\overline{\mathfrak{G}}$ is closed and the above-mentioned results of his paper [1929] apply to $\overline{\mathfrak{G}}$. This showed that Hilbert's problem is solved affirmatively for the compact group $\mathfrak{G}$.

To show that a compact, locally Euclidean topological group $\mathfrak{G}$ has a faithful representation, he proceeded as follows. Let $U^{(\mu)}(x)$, $\mu = 1, 2, 3, \ldots$, denote the inequivalent unitary matrix representations of $\mathfrak{G}$, with $U^{(\mu)}$ being $n_\mu \times n_\mu$. Form the $N_\mu \times N_\mu$ unitary matrix representation $V^{(\mu)}$, $N_\mu = \sum_{\nu=1}^{\mu} n_\nu$, which has the $U^{(\nu)}$, $\nu = 1, \ldots, \mu$, as diagonal blocks and zero blocks elsewhere. Also let $V^{(\infty)}$ denote the infinite matrix obtained by letting $\mu \to \infty$ in $V^{(\mu)}$. Each mapping $x \to V^{(\mu)}(x)$ is a continuous homomorphism from $\mathfrak{G}$ to a topological group of matrices. The same is true of $x \to V^{(\infty)}(x)$ once a suitable metric topology is introduced on the totality of all infinite matrices. In addition, by the point separation corollary $x \to V^{(\infty)}(x)$ is an isomorphism. If $\mathfrak{K}^{(\mu)}$ is the compact kernel of the corresponding homomorphism $x \to V^{(\mu)}(x)$, then $\mathfrak{K}^{(\mu+1)} \subset \mathfrak{K}^{(\mu)}$ and $\bigcap_{\mu=1}^{\infty} \mathfrak{K}^{(\mu)}$ is the kernel of $x \to V^{(\infty)}(x)$, which of course consists solely of the identity element of $\mathfrak{G}$. Using Brouwer's invariance of domain theorem and results on the theory of dimension from the book by Menger [1928], von Neumann showed that μ_0 exists so that $\mathfrak{K}^{(\mu)} = \mathfrak{K}^{(\mu_0)}$ for all $\mu \geqslant \mu_0$. Thus $\mathfrak{K}^{(\mu_0)}$ is also the kernel of $x \to V^{(\infty)}(x)$ and $x \to V^{(\mu_0)}(x)$ is consequently faithful.[19]

As the first affirmative solution to Hilbert's fifth problem for an extensive class of groups, von Neumann's paper inspired further work on it and therewith on the theory of topological groups, eventuating, in the early 1950's, in a definitive resolution for both abstract groups and transformation groups. See in this connection the account by Skljarenko [1971].

Regarding the impact of Weyl's work on the development of harmonic analysis on groups – the third and last line of development mentioned at the beginning of this section – little needs to be said. Thanks to Haar's work, the Peter–Weyl paper could be seen as establishing a Fourier type analysis on compact groups and, of course, suggesting the possibility of an analogous theory for locally compact groups since locally compact groups possess a Haar integral. As Gross has written in his essay on the evolution of harmonic analysis [1978, p. 533]: "Modern harmonic analysis begins in the 1920's .... The date

[19] A simpler proof based on von Neumann's idea of considering a descending chain of kernels can be given; see, e.g., [Bröcker and Dieck, 1985, p. 136].

of birth is 1927, and the official birth certificate is the remarkable paper ... by Peter and Weyl". An idea of the theory and the importance to its development of diverse aspects of Weyl's work can be obtained from Gross' essay and from those of Mackey [1992].

By way of conclusion I should mention that Pontriagin, L.S. (1908–1988), who went on to compose a standard reference on topological groups [1954], was among the first mathematicians to contribute to the research problems (1)–(3) generated by Weyl's work. In the brief period 1934–1935 he made important contributions to Cartan's problem of calculating Betti numbers for simple groups; and, inspired by von Neumann's paper, he solved Hilbert's problem for locally compact abelian groups and also developed for these groups his duality theory by using the fact that the character group associated to such a group is compact in a suitable topology so that the Peter–Weyl theory applies.

Bibliography

Abikoff, W. (1981), *The uniformization theorem*, Amer. Math. Monthly **88**, 574–592.

Borel, A. (1955), *Topology of Lie groups and characteristic classes*, Bull. Amer. Math. Soc. **61**, 397–432. Reprinted in *Papers* **1** (33).

Borel, A. (1986), *Hermann Weyl and Lie groups*, Hermann Weyl 1885–1985, K. Chandrasekharan, ed., Springer, Berlin, 655–685.

Brouwer, L.E.J. (1909a), *Die Theorie der endlichen continuierlichen Gruppen, unabhängig von den Axiomen von Lie*, Atti IV Congr. Intern. Mat. Roma (1908) **2**, 296–303. Reprinted in Brouwer, L.E.J. (1976) *Collected Works 2*, H. Freudenthal, ed., North-Holland, Amsterdam, 109–116.

Brouwer, L.E.J. (1909b), *Die Theorie der endlichen kontinuierlichen Gruppen, unabhängig von den Axiomen von Lie (Erste Mitteilung)*, Math. Ann. **67**, 246–267. Reprinted in Brouwer, L.E.J. (1976) *Collected Works 2*, H. Freudenthal, ed., North-Holland, Amsterdam, 118–139.

Brouwer, L.E.J. (1910), *Die Theorie der endlichen kontinuierlichen Gruppen, unabhängig von den Axiomen von Lie (Zweite Mitteilung)*, Math. Ann. **69**, 181–203. Reprinted in Brouwer, L.E.J. (1976) *Collected Works* **2**, H. Freudenthal, ed., North-Holland, Amsterdam, 156–178.

Brouwer, L.E.J. (1976), *Collected Works 2*, H. Freudenthal, ed., North-Holland, Amsterdam.

Bröcker, T. and Dieck, T.t. (1985), *Representations of Compact Lie Groups*, Springer, New York.

Cartan, E. (1913), *Les groupes projectifs qui ne laissent invariante aucune multiplicité plane*, Bull. Sci. Math. **41** 53–96. Reprinted in *Oeuvres* **1**, 355–398.

Cartan, E. (1925), *Les tenseurs irréductibles et les groupes linéaires simples et semi-simples*, Bull. Sci. Math. **49**, 130–152. Reprinted in *Oeuvres* **1**, 531–553.

Cartan, E. (1927), *La géométrie des groupes simples*, Annali di Matematiche **4**, 209–256. Reprinted in *Oeuvres* **1**, 793–840.

Cartan, E. (1928), *Sur les nombres de Betti des espaces de groupes clos*, Comptes Rendus, Acad. Sci. Paris **187**, 196–198. Reprinted in *Oeuvres* **1**, 999–1001.

Cartan, E. (1930), *La Théorie des Groupes Finis et Continus et l'Analysis Situs*, Gauthier-Villars, Paris. Reprinted in *Oeuvres* **1**, 1165–1225.

Cartan, E. (1936), *La topologie des espaces représentatifs des groupes de Lie*, L'Enseignement Mathématique **35**, 177–200. Reprinted in *Oeuvres* **1**, 1307–1330.

De Rham, G. (1931), *Sur l'analysis situs des variétés à n dimensions*, Jl. des Math. Pures et Appl. **10**, 115–200. Reprinted in De Rham, G. (1981), *Oeuvres Mathématiques*, L'Enseignement Mathématique, Université de Genève, 25–110.

De Rham, G. (1981), *Oeuvres Mathématiques*, L'Enseignement Mathématique, Université de Genève.

Dieudonné, J. (1989), *A History of Algebraic and Differential Topology 1900–1960*, Birkhäuser, Boston.

Engel, F. (1892), *Die Erzeugung der endlichen Transformationen einer projektiven Gruppe durch die infinitesimalen Transformationen der Gruppe. Erste Mitteilung*, Berichte über d. Verh. d. Sächsischen Gesell. der Wiss., Math.-Phys. Klasse **44**, 279–291.

Engel, F. (1893), *Die Erzeugung der endlichen Transformationen einer projektiven Gruppe durch die infinitesimalen Transformationen der Gruppe. Zweite Mitteilung, mit Beiträgen von E. Study*, Berichte über d. Verh. d. Sächsischen Gesell. der Wiss., Math.-Phys. Klasse **45**, 659–696.

Engel, F. (1902), *Review of papers by Slocum*, Jahrbuch über die Fortschritte der Mathematik **31**, 148–150.

Engel, F. (1912), *Review of [Brouwer*, 1909b], Jahrbuch über die Fortschritte der Mathematik **40**, 194.

Engel, F. (1913), *Review of [Brouwer*, 1910], Jahrbuch über die Fortschritte der Mathematik **41**, 181–182.

Freudenthal, H. (1968), *L'algèbre topologique en particulier les groupes topologiques et de Lie*, Revue de Synthèse III^e S., Nos 49–52, 223–243.

Gray, J. (1986), *Linear Differential Equations and Group Theory from Riemann to Poincaré*, Birkhäuser, Boston.

Gray, J. (1994), *On the History of the Riemann Mapping Theorem*, Supplemento ai Rendiconti Circolo Mat. Palermo **34** (2) 47–94.

Gross, K.I. (1978), *On the evolution of noncommutative harmonic analysis*, Amer. Math. Monthly **85**, 359–371.

Haar, A. (1933), *Der Massbegriff in der Theorie der kontinuierlichen Gruppen*, Ann. of Math. **34**, 147–169.

Hausdorff, F. (1927), *Mengenlehre*, Leipzig, Berlin.

Hawkins, T. (1982), *Wilhelm Killing and the structure of Lie algebras*, Archive for History of Exact Sciences **26**, 127–192.

Hawkins, T. (1987), *Non-Euclidean geometry and Weierstrassian mathematics: the background to Killing's work on Lie algebras*, Studies in the History of Mathematics, MAA Studies in Mathematics, vol. 26, E.R. Phillips, ed., 21–36.

Hawkins, T. (1989), *Line geometry, differential equations and the birth of Lie's theory of groups*, The History of Modern Mathematics, Vol. 1, J. McCleary and D. Rowe, eds, Academic Press, New York, 275–327.

Hawkins, T. (1992), *Jacobi and the birth of Lie's theory of groups*, Amphora. Festschrift für Hans Wussing zu seinem 65 Geburtstag, M.F.S. Demidov, D. Rowe and C. Scriba, eds, Birkhäuser, Basel, 289–313.

Hawkins, T. (1994), *The birth of Lie's theory of groups*, Proceedings, Sophus Lie Memorial Conference, Oslo 1992, O.A. Laudal and B. Jahren, eds, Oslo, 23–50.

Hawkins, T. (1998), *From general relativity to group representations. The background to Weyl's papers of 1925–26*, Matériaux pour l'histoire des mathématiques au XX^e Siècle. Actes du colloque à la mémoire de Jean Dieudonné. Société Mathématique de France, Séminaires et Congrès **3**, 69–100.

Hilbert, D. (1900), *Mathematische Probleme*, Nachrichten der K. Gesell. der Wiss. Göttingen, 253–297. Republished, with additions by Hilbert, in Archiv der Mathematik und Physik **1** (3), 44–63; 213–237. The *Archiv* version is reprinted in *Abhandlungen* **3**, 290–329. An English translation was published by M. Newson in Bull. Amer. Math. Soc. **8** (1902), 437–479.

Hilbert, D. (1902), *Über die Grundlagen der Geometrie*, Nachrichten der K. Gesell. der Wiss. Göttingen, 233–241. Expanded version, but without the general definition of a two-dimensional manifold, published in Math. Ann. **56** (1903), 381–422. A modified version of the *Annalen* paper (with the above general definition included) became Anhang IV in the 2nd (1903) and later editions of his *Grundlagen der Geometrie*.

Hurwitz, A. (1897), *Ueber die Erzeugung der Invarianten durch Integration*, Nachrichten der K. Gesell. der Wiss. Göttingen, 71–90. Reprinted in *Werke* **2**, 546–564.

Johnson, D.M. (1987), *L.E.J. Brouwer's Coming of Age as a Topologist*, Studies in the History of Mathematics, MAA Studies in Mathematics, vol. 26, E.R. Phillips, ed., 61–97.

Klein, F. (1883), *Neue Beiträge zur Riemannschen Funktionentheorie*, Math. Ann. **21**, 141–218. Reprinted in *Abhandlungen* **3**, 630–710.

Koebe, P. (1907), *Ueber Uniformisierung beliebiger analytischer Kurven*, Nachrichten der K. Gesell. der Wiss. Göttingen, 191–210.

Lie, S. (1883), *Über unendliche kontinuierliche Gruppen*, Forhandlinger, Christiania. Reprinted in *Abhandlungen* **5**, 314–360.

Lie, S. (1888–93), *Theorie der Transformationsgruppen. Unter Mit wirkung von ... Friedrich Engel*, 3 vols., Leipzig.

Lie, S. (1890), *Über die Grundlagen der Geometrie*, Berichte über d. Verh. d. Sächsischen Gesell. der Wiss., Math.-Phys. Klasse, 1890, 284–321, 355–418. Reprinted in *Abhandlungen* **2**, 380–468.

Mackey, G.W. (1992), *The Scope and History of Commutative and Noncommutative Harmonic Analysis*, American Mathematical Society, Providence, RI.

Menger, K. (1928), *Dimensionstheorie*, Teubner, Leipzig.

Poincaré, H. (1881), *Sur les fonctions fuchsiennes*, Comptes Rendus, Acad. Sci. Paris **93**, 301–303. Reprinted in *Oeuvres* **2**, 29–31.

Poincaré, H. (1882), *Sur les fonctions uniformes qui se reproduisent par des substitutions linéaires*, Math. Ann. **19**, 553–564. Reprinted in *Oeuvres* **2**, 92–105.

Poincaré, H. (1883), *Sur une théorème générale des fonctions*, Bull. Soc. Math. France **11**, 113–125. Reprinted in *Oeuvres* **4**, 57–69.

Poincaré, H. (1884), *Sur les groupes des équations linéaires*, Acta Mathematica **4**, 201–311. Reprinted in *Oeuvres* **2**, 401.

Poincaré, H. (1887), *Sur les hypothèses fondamentales de la Géométrie*, Bull. Soc. Math. France **15**, 203–216. Reprinted in *Oeuvres* **11**, 79–91.

Poincaré, H. (1895), *Analysis situs*, Jl. Ec. Polyt. **1** (2), 1–121. Reprinted in *Oeuvres* **6**, 193–288.

Poincaré, H. (1907), *Sur l'uniformisation des fonctions analytiques*, Acta Mathematica **31**, 1–63. Reprinted in *Oeuvres* **4**, 70–146.

Pontriagin, L.S. (1954), *Nepreryvnye Gruppy*, 2nd edn., Moscow. Translated into German by V. Ziegler as *Topologische Gruppen*, 2 vols., Leipzig, 1957–1958, and into English by A. Brown as *Topological Groups*, New York, 1966.

Samelson, H. (1952), *Topology of Lie groups*, Bull. Amer. Math. Soc. **58**, 2–37.

Scholz, E. (1980), *Geschichte des Mannigfaltigkeitsbegriffes von Riemann bis Poincaré*, Birkhäuser, Boston.

Schreier, O. (1925), *Abstrakte kontinuierliche Gruppen*, Abh. Math. Seminar der Hamburgischen Universität **4**, 15–32.

Schreier, O. (1927), *Die Verwandschaft stetiger Gruppen im grossen*, Abh. Math. Seminar der Hamburgischen Universität **5**, 233–244.

Schreier, O. (1928), *Über neuere Untersuchungen in der Theorie der kontinuierlichen Gruppen*, Jahresber. Deutschen Math.-Vereinigung **37**, 113–122.

Schur, I. (1924), *Neue Anwendungen der Integralrechnung auf Probleme der Invariantentheorie*, Sitzungsberichte Akad. Wiss. Berlin, 189–208. Reprinted in *Abhandlungen* **2**, 440–459.

Skljarenko, E.G. (1971), *Zum fünften Hilbertschen Problem*, Die Hilbertsche Probleme, H. Wussing, ed., Leipzig, 126–144.

Tychonoff, A. (1926), *Über einen Metrisationsatz von P. Urysohn*, Math. Ann. **95**, 139–160.

Urysohn, P. (1925), *Zum Metrisationproblem*, Math. Ann. **94**, 309–315.

von Neumann, J. (1927), *Zur Theorie der Darstellungen kontinuierlicher Gruppen*, Sitzungsberichte Akad. Wiss. Berlin, 76–90. Reprinted in *Collected Works*, **1**, 134–148.

von Neumann, J. (1929), *Über die analytischer Eigenschaften von Gruppen linearer Transformationen und ihrer Darstellungen*, Math. Zeitschr. **30**, 3–42. Reprinted in *Collected Works* **1**, 509–548.

von Neumann, J. (1933), *Die Einführung analytischer Parameter in topologischen Gruppen*, Ann. of Math. **34**, 170–190. Reprinted in *Collected Works* **2**, 366–386.

von Neumann, J. (1934), *Zum Haarschen Mass in topologischen Gruppen*, Compos. Math. **1**, 106–114. Reprinted in *Collected Works* **2**, 445–453.

Weil, A. (1940), *L'intégration dans les Groupes Topologiques et ses Applications*, Paris.

Weyl, H. (1913), *Die Idee der Riemannschen Fläche*, Teubner, Leipzig.

Weyl, H. (1924a), *Das gruppentheoretische Fundament der Tensorrechnung*, Nachrichten der K. Gesell. der Wiss. Göttingen, 218–224. Reprinted in *Abhandlungen* **2**, 461–467.

Weyl, H. (1924b), *Zur Theorie der Darstellung der einfachen kontinuierlichen Gruppen.* (*Aus einem Schreiben an Herrn I. Schur*), Sitzungsberichte Akad. Wiss. Berlin, 338–345. Reprinted in *Abhandlungen* **2**, 453–460.

Weyl, H. (1925), *Theorie der Darstellung kontinuierlicher halbeinfacher Gruppen durch lineare Transformationen. I–III*, Math. Zeitschr. **23**, 271–301, **24**, 328–376, **24**, 377–395. Reprinted in *Abhandlungen* **2**, 544–645.

Weyl, H. (1955), *Die Idee der Riemannschen Fläche*, 3rd edn., Teubner, Leipzig. English translation as *The Concept of a Riemann Surface*, Addison-Wesley, Reading, MA.

Weyl, H. and Peter, F. (1927), *Die Vollständigkeit der primitiven Darstellungen einer geschlossenen kontinuierlichen Gruppe*, Math. Ann. **97**, 737–755. Reprinted in *Abhandlungen* **3**, 58–75.

CHAPTER 8

By Their Fruits Ye Shall Know Them: Some Remarks on the Interaction of General Topology with Other Areas of Mathematics

Teun Koetsier and Jan van Mill

Faculteit Wiskunde en Informatica, Vrije Universiteit,
De Boelelaan 1081a, 1081 HV Amsterdam, The Netherlands
E-mail: teun@cs.vu.nl; vanmill@cs.vu.nl

1. Introduction

In his letter of invitation to contribute to this handbook of the *History of Topology*, Professor James asked us to discuss the role of general topology in other areas of topology. So this paper is *not* a paper on the history of general topology, it is a paper on the history of its *interactions* with other fields of mathematics. Of the many possibilities, we decided to report on the one hand on the genesis of general topology and on the other hand on infinite-dimensional topology and set theoretic topology.[1] For a much more comprehensive description of (parts of) the history of general topology, we refer the reader to [15].

The primary goal in general topology, also sometimes called point set topology, is the investigation and comparison of different classes of topological spaces. This primary goal continues to yield interesting problems and results, which derive their significance from their relevance with respect to this primary goal and from the need of applications. In the history of general topology we distinguish three periods. The first period is *the prehistory* of the subject. It led to the work of Hausdorff, Brouwer, Urysohn, Menger and Alexandroff. The prehistory resulted in a definition of general topology, but it left many questions unanswered. The second period, roughly from the 1920's until the 1960's was *general topology's golden age*. Many fundamental theorems were proved. Many of the results from that period can be viewed as a necessary consequence of the genesis of the subject. However, much work from the golden age was also an investment in the future, an investment that started to yield fruit in the third period lasting from the 1960's until the present. That is why we will call this period *the period of harvesting*.

[1] These two research areas are familiar to us. Other possibilities to report on could have been: Topological Dynamics, Theoretical Computer Science, Topological Groups, Topological Games, Categorical Topology, Dimension Theory, Topological Algebra, Descriptive Set Theory, etc.

HISTORY OF TOPOLOGY
Edited by I.M. James

In this paper we concentrate on the first and the last period: the prehistory and on the period of harvesting. In Section 2, which deals with the prehistory, we describe in particular the historical background of the concept of an abstract topological space. We discuss the contributions of Georg Cantor, Maurice Fréchet and Felix Hausdorff. That discussion is rather informal; it reflects the informal style of that particular period in history. Because a complete description of that work is out of the question, we concentrate in particular on the background and the genesis of some crucial topological notions. Although we repeatedly went back to the original texts, we also relied heavily on secondary sources. We would like to mention Purkert and Ilgauds [145] and Dauben [53] with respect to Cantor, Taylor [170] for information on Fréchet, and Scholz [156] concerning Hausdorff. We also used Moore [136] and Monna [135] with appreciation.[2]

Sections 4 and 5 of the paper are devoted to the period of harvesting. In that period general topology rather unexpectedly succeeded in solving several difficult problems outside its own area of research, in functional analysis and in geometric and algebraic topology. Also here a survey of all significant results is impossible. There were in that period at least two major developments in general topology that revolutionized the field: the creations of *infinite-dimensional topology* and *set theoretic topology*.[3] It was mainly due to the efforts of Dick Anderson and Mary Ellen Rudin that these fields have played such a dominant role in general topology ever since.

There is a well-known pattern that occurs often in mathematics. An established part of mathematics generates nontrivial questions and possible ways to answer these questions that are new, but of little immediate significance. Research in the area is essentially pursued for its own sake. However, if the mathematics is good, after a longer or shorter period, the theories involved significantly contribute to solve external problems. Hilbert [89] wrote:

> The final test of every new mathematical theory is its success in answering pre-existent questions that the theory was not designed to answer. By their fruits ye shall know them – that applies also to theories.[4]

And indeed, there is no doubt that the most convincing test for the value of a theory is its external significance.[5] We will show that the genesis and further development of general topology offer many examples that illustrate this pattern. We believe, frankly, that research in general topology is almost exclusively driven by two things: the existence of difficult, challenging problems, and the beauty of many of the results. Of course, not everything that was and is done in general topology is equally important, as is the case in any other field of mathematics. It is, for example, relatively easy to define variations of the axiomatic bases of the various types of spaces and, as in other fields, it is not always easy to say in advance whether certain lines of research are worth pursuing. Yet, most of the areas of research in

[2] Manheim wrote [123] the first book on the history of general topology and certainly at the time it was a useful contribution. He restricted himself to what we call the prehistory of the field.

[3] Also shape theory was created by Borsuk, see, e.g., [33], but this field was much more motivated from algebraic and geometric topology than infinite-dimensional and set-theoretic topology.

[4] "Wherefore by their fruits ye shall know them", St. Matthew 7:20.

[5] Hallett [83]: It may take a long time before the external significance of a theory becomes clear. For example, when the Greeks were pursuing mathematics entirely for its own sake, independent of applications, they developed an elaborate theory of conic sections. Only many centuries later Kepler applied this theory to describe the orbits of the planets. External significance is a sufficient condition for quality, it proves the value of a theory afterwards. Obviously, a theory may generate and solve such interesting problems that even without definite proof through external significance, the theory should be considered valuable (Koetsier [107, p. 171]).

general topology represent good mathematics. The two examples of infinite-dimensional topology and set theoretic topology illustrate this.

2. The prehistory of general topology

2.1. *Developments in 19-th century analysis*

2.1.1. *Weierstrassian analysis.* Cauchy played a major role in the *first revolution of rigour*, that had turned eighteenth century calculus from a collection of formal methods to solve problems, into a coherent deductive theory based upon definitions of the fundamental concepts of convergence, continuity, the derivative and the integral in terms of the notion of limit (Grabiner [82]). Yet, after Cauchy, a further development and refinement of concepts was inevitable. Cauchy primarily used his new conceptual apparatus to give a solid foundation of existing analysis and in his mathematics a function is still always associated with a formula. In the second half of the nineteenth century the conceptual apparatus itself became the object of investigation. This happened in combination with a much more general concept of function: a function became, in principle, a completely arbitrary correspondence between numbers. In particular the discovery that discontinuous functions can be expressed by means of Fourier series – dating from the beginning of the 19-th century – contributed considerably to this change. For example, in 1854, in his "Habilitationsschrift", Riemann studied the problem of the representation by means of Fourier series of as large a class of arbitrary functions as possible. This automatically led to the problem of the integrability of highly discontinuous functions. Riemann discovered that a function could possess an infinite number of points of discontinuity in any interval and still be integrable (in the sense of Cauchy–Riemann). It became clear that such highly discontinuous functions could be studied and research partially shifted from the investigation of functions defined by a particular formula or classes of formulas to the investigation on a much more general level: abstracting from particular examples that illustrate those relations, the relations themselves between notions like real number, function, series, convergence, limit, continuity, differentiability, integrability became subject of investigation.

From this perspective Cauchy's work showed weaknesses and a *second revolution of rigour* took place in analysis, that is associated with the name of Weierstrass. It became clear that Cauchy had not sufficiently distinguished between, for example, uniform convergence and non-uniform convergence. It also became clear that he had, essentially, taken the real numbers and their properties, for example their completeness, for granted. A proof of a theorem like "A real function that is continuous in a closed and bounded interval attains its maximum value", which we owe to Weierstrass, would have been out of place in Cauchy's work[6] and the same holds for more fundamental theorems like the Bolzano–Weierstrass Theorem, actually due to Weierstrass alone: "Every infinite bounded subset of $\mathbb{R}^n$ has a limit point".[7] This theorem was stated for $n = 2$ by Weierstrass in a course of lectures in 1865. In 1874 he gave a general proof (Moore [136, p. 17]). The theorem is necessary to prove the existence of limits, something that Cauchy had also, at heart, still taken for granted.

---

[6] Cauchy's well-known proof of the intermediate value theorem is in the context of Cauchy's work rather exceptional. But also in this case the first completely satisfactory proof was given by Heine [87].

[7] As far as we know the notion of limit point or accumulation point was first used by Weierstrass.

2.1.2. *Volterra, Ascoli.* As early as 1883[8] Volterra had the idea to create a theory of functionals,[9] or real-valued "functions of lines", as he called the field. Volterra wrote several papers on the subject.[10] The lines are all real-valued functions defined on some interval $[a, b]$. These functions are viewed as elements of a set for which notions like neighbourhood and limit of a sequence can be defined. Volterra gave definitions for the continuity and the derivative of a function of lines and he tried to build up a line-function theory analogous to Riemann's theory of complex functions. These attempts were not motivated by their immediate significance in solving problems in the calculus of variations. Hadamard wrote about Volterra's motivation:

> Why was the great Italian geometer led to operate on functions as the infinitesimal calculus operated on numbers [...]? Only because he realised that this was a harmonious way to complete the architecture of the mathematical building.[11]

Weierstrass' teaching was influential also in Italy. In 1884 Giulio Ascoli (1843–1896) extended the Bolzano–Weierstrass Theorem to sets of functions as follows. He studied a set $\mathcal{F}$ of uniformly bounded functions on $[a, b]$. In order to prove that a sequence of functions $\{f_n\}$ in $\mathcal{F}$ possesses a convergent subsequence $\{g_n\}$, he needed the assumption that the set $\mathcal{F}$ is equicontinuous. The result is known as Ascoli's Theorem. Equicontinuity of $\mathcal{F}$ means then that for every $\varepsilon > 0$, there exists a $\delta > 0$, such that for all $f \in F$ and for all $|x - y| < \delta$, we have $|f(x) - f(y)| < \varepsilon$. The proof-idea is that a subsequence $\{g'_n\}$ of $\{f_n\}$ is chosen, such that $\{g'_n(a)\}$ converges. Then a subsequence $\{g''_n\}$ is chosen from $\{g'_n\}$ such that $\{g''_n(b)\}$ converges. Then a subsequence $\{g'''_n\}$ is chosen from $\{g''_n\}$ such that $\{g'''_n((a + b)/2)\}$ converges. Continuing in this way a sequence of converging sequences is generated that correspond to the elements of a set that is dense in $[a, b]$. The "diagonal sequence" then does the job (Moore [136, p. 81]).

2.1.3. *The Dirichlet-principle and the theorem of Ascoli–Arzelà.* The Italian attempts to extend results from Weierstrass' real analysis to sets of functions and real functions defined on such sets, can certainly be understood as "a harmonious way to complete the architecture of the mathematical building". Yet there were also other reasons. An example is Dirichlet's principle. In 1856–1857 Dirichlet lectured on potential theory in Göttingen. Modelling conductors, he considered a part Ω of $\mathbb{R}^3$, bounded by a surface S on which a continuous function is defined and dealt with the problem of the existence of a function u on Ω that equals f on S and satisfies $\Delta u(x, y, z) = 0$. In order to solve the problem he considered the integral

$$U = \int_\Omega \left[\left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2 + \left(\frac{\partial u}{\partial z}\right)^2\right] \mathrm{d}V,$$

on Ω, which obviously is non-negative for all functions u considered. He concluded that there must be at least one function u on Ω for which the integral reaches a minimum value. One can show that the minimizing function satisfies $\Delta u(x, y, z) = 0$ and Dirichlet thought

[8] According to Whittaker, see Monna [135, p. 108].

[9] The term functional was introduced by Hadamard in 1903 (Monna [135, p. 108]).

[10] Atti della Reale Accademia dei Lincei (4) **3** (1887), 97–105, 141–146, 153–158 = Opere matematiche **1**, 294–314, and other papers of the same and later years. We have not seen these papers.

[11] Quoted by Siegmund-Schultze [163, p. 377].

he had solved the problem (Monna [135, pp. 27–30]). In 1871 Heine criticised Dirichlet for accepting without proof the existence of a minimizing function (Monna [135, p. 41]). And indeed this method, which is sometimes called *Dirichlet's principle*, needs further justification, because the existence of a greater lower bound for the values of the integral does not necessarily imply that there exists a function that corresponds to that greater lower bound. In a paper from 1889 by Cesare Arzelà [14] the author refers to Volterra and his "functions that are dependent on lines" ("funzioni dipendenti dalle linee") and writes that continuity for such functions had been defined but that the existence of maxima and minima still needed investigation. Expressing the hope that his work will lead to a justification of the "Principio di Riemann–Dirichlet" he proceeded to prove what is nowadays usually called the Theorem of Ascoli–Arzelà. First Arzelà generalized Ascoli's theorem from 1884 and proved that an equicontinuous set $\mathcal{F}$ of uniformly bounded functions on $[a, b]$ has a limit-function. By definition a limit-function f of $\mathcal{F}$ is a function that has the property that for every $\varepsilon > 0$, there are infinitely many functions g in $\mathcal{F}$ for which, for all x,

$$f(x) - \varepsilon < g(x) < f(x) + \varepsilon.$$

Then Arzelà turned to continuous real-valued functionals defined on such an equicontinuous set $\mathcal{F}$ of functions – something which Ascoli had not done – and showed that, if the set $\mathcal{F}$ is closed, i.e. contains all its limit-functions, the lower bound of the set of values of the functional, the upper bound and all values in between are taken.

In 1896 Arzelà published a paper in which he applied his results to the Dirichlet principle. He succeeded to prove it only under certain extra conditions (Monna [135, p. 112]). Nowadays the fundamental Ascoli–Arzelà Theorem in analysis is phrased in terms of compactness, a term introduced by Fréchet in 1904. However, in order to understand the background of the ideas of Fréchet, it is necessary to describe the birth of transfinite set theory first.

2.2. *Cantor*

2.2.1. *From Fourier series to derived sets and transfinite counting.* Georg Cantor studied in Berlin under Kummer, Kronecker and Weierstrass. In 1869 he became a Privatdozent at the University of Halle. His doctoral thesis and his Habilitationsschrift were on number theory, but soon Cantor turned to analysis. Edward Heine, one of his colleagues at the University of Halle, had suggested him to study the problem of the uniqueness of the representation of a function by means of a trigonometric series. In the years 1870 through 1872 Cantor published a series of papers on that matter. In 1870 he published a proof of the theorem saying that if

$$\frac{1}{2}a_0 + \sum_{n=1}^{\infty}(a_n \sin nx + b_n \cos nx)$$

converges to $f(x)$ for all x on $(0, 2\pi)$, $f(x)$ cannot be represented by another trigonometric series converging to $f(x)$ for all x on $(0, 2\pi)$. In 1871 he improved the proof and, moreover, showed that the representation remained unique if the requirement of convergence or the convergence to $f(x)$ would be dropped for a finite set of exceptional points.

Soon Cantor's main interest moved from trigonometric series to exceptional sets consisting of infinitely many points. It was this problem that led to the theory of transfinite sets and at the same time to a number of topological results. Cantor realized that the proof of the theorem could be easily modified to hold if the infinite exceptional set contained a finite number of limit points and that it could even be proved if it contained an infinite number of limit points, provided the set of limit points itself possessed at most a finite number of limit points. The argument was easily extended to higher levels of sets of limit points of sets of limit points. However, how could one describe such complex subsets of the continuum? This question led Cantor, in fact, to a definition of the real number system, which was independent of those of Weierstrass, Méray, Heine and Dedekind that were also given in that period. Cantor [38] started with the set of rational numbers, which he called A. He considered the set of all Cauchy sequences of rational numbers (as they are now called – Cantor himself called them fundamental sequences) and defined what we would now call an equivalence relation on that set. The set of equivalence classes is called B. The ordering and the elementary operations are then extended from A to the union of A and B. Cantor now repeats the construction: In precisely the same way by considering Cauchy sequences in $A \cup B$ a set C is generated, then sequences in $A \cup B \cup C$ generate a set D, etc. In this way after λ steps a set L is reached whose elements Cantor called "numbers of type λ". Cantor was aware of the fact that he could identify A with a subset of B and he also knew that B, C, D, etc. are isomorphic (although he does not use that terminology), but he avoided the identification. He needed the hierarchy of number sets to identify point sets on a line. In order to do that he first introduced the notion "derived set" of a point set on a line. The first derived set P^1 of a point set P is by definition the set of all limit points of P and recursively: the derived set P^λ of a set P is the first derived set of $P^{\lambda-1}$. Cantor then called P a set of type ν, iff the ν-th derivative P^ν is finite. The existence of such sets can now be seen by using the above defined hierarchy of number sets, because if we take one point on the line whose coordinate is a number of type ν, we know that this number represents a Cauchy sequence of numbers of type $\nu - 1$, while those numbers all represent Cauchy sequences of numbers of type $\nu - 2$, etc. If in this way, we go all the way back to the rational numbers, we wind up on the line with a point set of type ν.

Applying this new apparatus Cantor proved the uniqueness theorem for trigonometric series for exceptional sets of type ν where ν is an arbitrary natural number. In the same paper he wrote with respect to the hierarchy of number systems defined by means of Cauchy sequences:

> [...] the notion of number, in so far as it is developed here, carries within it the germ of a necessary and absolutely infinite extension.[12]

Although he does not mention it in his paper he had at the time already extended that hierarchy beyond the finite levels. And indeed, the question whether there exist sets that are such that P^ν is infinite for all finite ν, arises naturally. We know that already in 1870 Cantor was aware of the possibility to count beyond the finite (Purkert and Ilgauds [145, p. 39]).The idea of the transfinite ordinal numbers, was born in this context.

2.2.2. *The birth of the transfinite cardinals.* In 1872 and 1873 the nature of the continuum intrigued Cantor more and more. In a letter to Dedekind, dated November 29, 1873, he

[12] "[...] der Zahlenbegriff, soweit er hier entwickelt ist, den Keim zu einer in sich notwendigen und absolut unendlichen Erweiterung in sich trägt" (Cantor [40, p. 95]).

wrote that he had tried to find a one–one correspondence between the natural numbers and the real numbers, but that he had failed. Several days later the transfinite cardinal numbers were born; they still would have to go a long way, but the idea was there. On December 7 of the same year Cantor wrote to Dedekind that he had found the proof that the proposed one-to-one correspondence does not exist. Cantor, who would later use the diagonal method, gave the following simple proof. Let a_1, a_2, a_3, a_4, etc. be the sequence of all real numbers. Consider an interval $[p, q]$. Find in the sequence the first two real numbers that represent an interval $[p_1, q_1]$ inside $[p, q]$. Find then the first two real numbers that represent an interval $[p_2, q_2]$ inside $[p_1, q_1]$, etc. This inevitably leads to a nested sequence of intervals with a non-empty intersection of points that do not occur in the sequence a_1, a_2, a_3, a_4, etc. Cantor published the proof in 1874 (Cantor [40, pp. 115–118]) pointing out that the proof implied the existence of transcendental numbers. It is remarkable that at the time, Cantor and Dedekind both considered these results as interesting but not of great importance (Purkert and Ilgauds [145, p. 45]). The next problem Cantor turned to was the question whether a 2-dimensional continuum could be mapped one-to-one on the real numbers. In 1877 he found the answer: the unit square, yes, even the n-dimensional unit cube can be mapped one-to-one on the interval $[0, 1]$. The paper was published in 1878. Cantor immediately realised that the result created a problem for the traditional view that the number of dimensions of a continuum corresponded to the number of parameters needed to describe it. Here we have the beginning of dimension theory. A survey of its further history was given by Johnson [98, 99]. See also Koetsier and van Mill [108].

2.2.3. *Transfinite set theory and topological notions.* Those first results from the period 1872–1878 gave Cantor the idea that the problem of the nature of the different kinds of point sets could be approached systematically. That is what he did in a famous series of six publications under the title "About infinite linear pointmanifolds" ("Über unendliche lineare Punktmannigfaltigkeiten"), that appeared in the years 1879 through 1884. The papers wonderfully show how Cantor's theory gradually developed; they also show the emergence of several topological notions and results. Right from the start set theory and general topological notions have been intimately connected. A complete discussion of the six publications goes beyond the purpose of this paper. We will mention a few results.

In the first paper Cantor distinguishes point sets of the first kind – the n-th derivative is empty for a finite n – and points sets of the second kind – by definition those that are not of the first kind. The notion of "density in an interval" is introduced and it is shown that sets of the first kind are never dense in an interval. Cantor also shows that all sets of the first kind and also some but not all of the second kind are countable. In the second paper Cantor introduces the sequence: P^{∞}, $P^{\infty+1}$, $P^{\infty+2}, \ldots$ etc., where P^{ν} refers to the ν-th derivative of a set P. The fourth paper contains a number of topological results. He calls a set P of $\mathbb{R}^n$ "isolated" if it contains none of its limit-points. He proves: "Every isolated set in $\mathbb{R}^n$ is at most countable" and some related results.

In the fifth paper the transfinite ordinal numbers, viewed as well-ordered sets, are "constructed", and denoted in the now standard way: $\omega, \omega + 1, \ldots, \omega \cdot \omega, \ldots$ etc. The ordinal numbers are related to the cardinal numbers by means of the notion of number class. The theory developed in this way generated two fundamental problems: the need to prove that every set can be well-ordered (this would guarantee that all cardinal numbers could be reached by means of ordinal numbers) and the continuum hypothesis. The continuum hypothesis is in the last sentence of the fifth paper. In this text from 1878 Cantor writes that his investigations point at the conclusion that among the infinite "linear manifolds", i.e. the subsets of $\mathbb{R}$, there would occur only two cardinal numbers. He added: "we postpone a precise investigation to a later occasion" (Cantor [40, p. 133]).

In this fifth paper Cantor also discusses the question when a subset of $\mathbb{R}^n$ should be called a "continuum". In order to answer that question he defines the notions of a perfect point set and a connected point set. A *perfect* point set is by definition equal to its derivative. A set T is by definition connected if for any two points t and t' of T and for any $\varepsilon > 0$, there exists a finite number of points $t_1, t_2, \ldots, t_n$ of T in such a way that all distances tt_1, t_1t_2, t_2t_3, $\ldots$, $t_{n-1}t_n$, t_nt' are all smaller than ε. A subset of $\mathbb{R}^n$ then is defined as a *continuum* iff it is a perfect, connected set (Cantor [40, p. 194]). In a note Cantor gave the famous example of a set that is perfect and at the same time dense in no interval:

$$\left\{\frac{c_1}{3} + \frac{c_2}{3^2} + \frac{c_3}{3^3} + \cdots : c_i \in \{0, 2\} \text{ for every } i\right\}.$$

The sixth paper contains a number of topological results on point sets in $\mathbb{R}^n$, which Cantor undoubtedly obtained while working towards a proof of the continuum hypothesis. Two examples are: "A perfect set is not countable", and "If a set is not countable, it can be split into a perfect set and a countable set".

2.2.4. *The reception of set theory.* After a period in which he hardly wrote anything Cantor published in 1895 and 1897 in two parts his last important paper on set theory: "Contributions to the foundation of transfinite set theory" ("Beiträge zur Begründung der transfiniten Mengenlehre"). In this paper, primarily devoted to "general" set theory,[13] Cantor's theory got its final form. At the same time the appreciation for set theory among mathematicians was growing slowly. Right from the start Cantor's set theory had been met by sceptical reactions. In particular, Kronecker was very critical. However, significant external applications made many prominent mathematicians understand the value of set theory. In 1872 Heine, Cantor's colleague in Halle, had proved that a real-valued function that is continuous on an interval $[a, b]$ of $\mathbb{R}$ is uniformly continuous. The proof runs roughly as follows. Using the continuity, Heine first aims at constructing for all small $\varepsilon > 0$ a monotonously increasing sequence $\{x_i\}$ with $a = x_1$ for which

$$\left|f(x_{i+1}) - f(x_i)\right| = 3\varepsilon$$

and for all x with $x_i \leqslant x \leqslant x_{i+1} \leqslant b$, one has

$$\left|f(x) - f(x_i)\right| \leqslant 3\varepsilon.$$

If the sequence cannot be constructed (because the function varies less than 3ε) or the construction stops after a finite number of steps because in the remaining interval the function varies less than 3ε, we are done. If the sequence is infinite, it converges to a number X in the interval. Then there exists also an η for which for all x with $X - \eta \leqslant x \leqslant X$ we have

$$\left|f(x) - f(X)\right| \leqslant 2\varepsilon.$$

[13] Zermelo wrote: "viele Hauptsatze der 'allgemeinen' Mengenlehre finden erst hier ihre klassische Begründung" (Cantor [40, p. 351]).

This, however, contradicts the fact that in the interval $[X - \eta, X]$ there are infinitely many points of the sequence $\{x_i\}$ for which $|f(x_{i+1}) - f(x_i)| = 3\varepsilon$.

In his 1894 doctoral thesis Emile Borel (1871–1956) applied Cantorian set theory to problems of analytic continuation in the theory of functions of a complex variable. It was undoubtedly this work which put him on the road to his later contributions to measure theory. However, one of his proofs involved "a theorem interesting in itself [...]: If one has an infinity of sub-intervals on a line (that is a closed interval) such that every point of the line is interior to at least one of them, a finite number of intervals can effectively be determined having the same property" (Quoted and translated by Hawkins; Grattan-Guinness [69, p. 175]). We have here the Heine–Borel Covering Theorem. Borel's proof uses Cantor's transfinite ordinals. He considers a transfinite sequence $\{(a_\lambda, b_\lambda)\colon \lambda < \alpha\}$ of open intervals that covers the interval $[a, b]$ from the left to the right and then by transfinite induction on α proves that the collection can be reduced to a finite collection. Hallett [83] discussed this proof an argued that, although the Heine–Borel Theorem was soon proved without the use of transfinite numbers, Borel's proof still counts as one of the first applications of transfinite ordinal numbers outside of set theory. Soon other applications followed. Hurwitz gave an invited lecture at an international congress of mathematicians in Zürich in 1897 on the development of the general theory of analytical functions in which he summarized Cantor's theory of transfinite ordinal numbers and subsequently applied it to classify analytical functions on the basis of their sets of singular points. Set theoretical methods had arrived in a classical discipline like the theory of complex functions. And both the problem of the Continuum Hypothesis and the Well-Ordering Theorem occur in Hilbert's famous list of problems he considered in 1900 to be the most important for mathematical research in the twentieth century.

2.3. *Maurice Fréchet's "Analyse Générale"*

2.3.1. *Tables, chairs, and beer mugs: another revolution of rigour.* In the history of analysis one distinguishes often the "first revolution of rigour", brought about by Cauchy and the "second revolution of rigour", brought about by Weierstrass. The systematic introduction of the axiomatic method in mathematics (in combination with the language of set theory and first-order predicate logic) could, undoubtedly, also be characterized as a revolution of rigour. There is a famous story told by Constance Reid in her biography of Hilbert:

> In his docent days Hilbert had attended a lecture in Halle by Hermann Wiener on the foundations and structure of geometry. In the railway station in Berlin on his way back to Königsberg, under influence of Wiener's abstract point of view in dealing with geometric entities, he had remarked thoughtfully to his companions: 'One must be able to say at all times – instead of points, straight lines and planes – tables, chairs, and beer mugs' (Reid [146, p. 57]).

In an appendix to the book Weyl writes that according to Blumenthal it must have been 1891 and Wiener's paper was on the role of Desargues's and Pappus's theorems (Reid [146, p. 264]). Hilbert's remark contains in a nutshell an important aspect of the abstract, axiomatic point of view: the theory becomes independent of its intended model; whatever names are used for the undefined terms, the axioms completely determine the way in which these terms are related. In Hilbert's "Foundations of Geometry" ("Grundlagen der Geometrie") of 1899 this point of view is applied to geometry. For many mathematicians Hilbert's book represented the future; after more than 2000 years Euclid had been dethroned. Consequently, in the first decade of the twentieth century the axiomatic method was very much in the air. In 1904 Zermelo published a proof of the Well-Ordering Theorem (Moore [136, p. 159]). The proof contains the first explicit statement of what was later called the Axiom of Choice. The reactions to the paper were such that Zermelo found it necessary to secure the proof even further. The result was Zermelo's axiomatization of set theory (Moore [136, p. 157]).[14]

The axiomatic method is, on the one hand, a method by means of which an already existing theory can be given its final form. However, the axiomatic method is also a powerful research method. Its basic rule is: "The occurrence of analogy between different areas points at the existence of a more general structure that should be defined explicitly by means of a suitable set of axioms". In France, Borel used the axiomatic method, Lebesgue did and also Fréchet, who applied it on a problem suggested to him by Hadamard, whose student he was. At the first International Congress of Mathematicians in 1897 Hadamard lectured briefly on possible future applications of set theory. He remarked that it would be worthwhile to study sets composed of functions. Such sets might have properties different from sets of numbers or points in space. He said:

> But it is primarily in the theory of partial differential equations of mathematical physics that research of this kind will play, without any doubt, a fundamental role[15]

and one of the examples that he implicitly referred to was Dirichlet's principle.

2.3.2. *The genesis of Fréchet's thesis.* In 1904 and 1905 Fréchet published a series of short papers on "abstract sets" or "abstract classes" in the "Comptes Rendus", that layed the groundwork of his thesis: "Sur quelques points du calcul fonctionnel", Rendiconti del Circolo Matematica di Palermo, 1906, pp. 1–74. That thesis is one of Fréchet's most important contributions to mathematics. We will concentrate on the early papers in order to get an idea of the genesis of the thesis. In his first paper [74] the analogy between Weierstrass' theorem: "A real function continuous in a closed and bounded interval attains its maximum value" and the Dirichlet principle is given as the motivation to develop a general theory of continuous real functions (Fréchet said "opérations fonctionnelles") on arbitrary sets that encompasses both theorems. Fréchet did not mention the Italians and it is possible that he only heard about their work in 1905. In his first paper Fréchet introduced an abstract axiomatic theory of limits. The theory refers to a set or class C of arbitrary elements and concerns infinite sequences of elements $A_1, A_2, A_3, \ldots$ of C that may or may not possess a limit element B in C. Fréchet's axioms are

(i) If a sequence has a limit B, then all infinite subsequences have the same limit, and
(ii) If $A_i = A$ for all i, then the limit of the sequence equals A.

In terms of this axiomatically defined notion of limit Fréchet can then define the notions of a closed subset of C, a compact subset of C and of a continuous real function on a subset of C.

---

[14] So Zermelo's primary motivation was not the occurrence of the antinomies. By the way, the so-called Russell's paradox was also found by Zermelo, several years before Russell did so (Moore [136, p. 89]).

[15] "Mais c'est principalement dans la théorie des équations aux dérivées partielles de la physique mathématique que les études de cette espèce joueraient, sans nul doute, un role fondamental" (quoted by Taylor [170, p. 259]).

(iii) A subset E of C is by definition *closed* if every limit element of a sequence of elements of E belongs to E.
(iv) A subset E of C is *compact* if for all sequences E_n consisting of non-empty closed subsets of C, that are such that E_{i+1} is a subset of E_i for all i, the intersection of all the E_n's is non-empty.
(v) The continuity of a function F on C is also defined in terms of sequences: F is *continuous* on a subset E of C if for all sequences $\{A_i\}$ in E that have a limit B in E, the sequence $\{F(E_i)\}$ has the limit $F(B)$.

Fréchet then, without further proof, phrases the generalisation of Weierstrass' theorem as follows: "If E is a closed and compact set in C and U is a continuous functional operation on C, then the values of U are bounded and U assumes an absolute maximum value at some point A of E." In his next note [77] in the "Comptes Rendus" Fréchet answers in the negative the question whether the derived set of a set E is necessarily closed. The counterexample that he gives consists of all real polynomials in the set of all real functions on an interval; a function f is the limit of a sequence of polynomials if there is pointwise convergence. This was a problem for Fréchet, because he felt that in order to get interesting generalisations of existing theorems he would need the property that the derivative of a set is always closed. In the third note [76] the new ideas are applied to the space E^* of infinitely many dimensions, the elements of which are all real sequences $\{a_i\}$. Sequence A is the limit of a sequence $\{A_i\}$ of sequences iff for all p the sequence of p-th coordinates of the A_i converges to the p-th coordinate of A.[16] A set of points A in E^* is bounded iff there are fixed numbers M_i such that for all points in the set for all i the absolute value of the i-th coordinate is smaller than M_i. Fréchet defines a condensation point ("point de condensation") of a set A as a limit-point that remains a limit-point of the set if one removes in an arbitrary way a countable infinite number of points from the set. Fréchet then states, without actually giving proofs, that he succeeded in proving several theorems. Three examples are: "The necessary and sufficient condition for a subset of E^* to be compact is that it is bounded", "The derived set of a subset of E^* is closed" and "Every uncountable and bounded subset of E^* possesses at least one condensation point".

The desire to develop a general theory in which the derived set of a set E is necessarily closed, continued to bother Fréchet. There exists an interesting letter (probably from 1904) concerning this point from Hadamard to Fréchet in which Hadamard suggests the use of an abstract notion of nearness or neighbourhood ("voisinage") (quoted by Taylor [170, pp. 245–246]). Hadamard wrote:

> Would it be good if you started, in general, from the notion of neighbourhood and not from that of limit?[17]

In [75] Fréchet had decided to introduce a generalised notion of "voisinage", assuming that in the classes of arbitrary elements to each couple of arbitrary elements there corresponds a real number (A, B) for which

(i) $(A, B) \geqslant 0$,
(ii) $(A, B) = 0$ iff $A = B$,
(iii) If (A, C) and (B, C) are infinitely small, then so (A, B).

---

[16] So Fréchet considers the space that we now call s, see Section 4.

[17] "Feriez vous bien de partir, en général, de la notion de voisinage et non de celle de limite? [...]" (Taylor [170, p. 246]).

(The last requirement means: if (A, C) and (B, C) are sequences converging to zero, then so (A, B).) Fréchet defines the notion of limit in terms of this abstract notion of distance: $\{A_n\}$ converges to A iff (A_n, A) converges to 0. Without proof Fréchet states that always when the limit can be defined by means of a suitable "écart": 1. Every derived set is closed, 2. A functional operation that is continuous on a compact set is uniformly continuous. Fréchet also refers to Ascoli and Arzelà, remarking that these theorems and the ones in his earlier notes can be seen as generalisations of the Italian work.

2.3.3. *Fréchet's 1906 thesis.* Fréchet's 1906 thesis is based on the papers from the period 1904–1905. We will only discuss the thesis very briefly.[18] In the thesis an abstract class with sequential limits that satisfy the two requirements from his first 1904 note is called "une classe (L)". We shall call them *L(imit)-classes*. In the second chapter of the first part of the thesis Fréchet introduces an abstract notion of distance, which he calls "voisinage". An abstract set is "une classe (V)", or, as we will say, a *V(oisinage)-class* if there exists a real-valued binary function (A, B) on the set which satisfies:

(i) $(A, B) = (B, A) \geqslant 0$,

(ii) $(A, B) = 0$ iff $A = B$,

(iii) There exists a positive real function $f(\varepsilon)$, defined for positive ε, for which

$$\lim_{\varepsilon \to 0} f(\varepsilon) = 0,$$

such that, whenever $(A, B) \leqslant \varepsilon$ and $(B, C) \leqslant \varepsilon$, then $(A, C) \leqslant f(\varepsilon)$.

A V-class can be turned into an L-class by means of the definition: $\{A_n\}$ converges to A iff (A_n, A) converges to 0.

It is remarkable that at one point in the thesis Fréchet replaces the third V-class axiom by a triangle inequality:

(iii^a) for all A, B, C we have $(A, C) \leqslant (A, B) + (B, C)$.

Here is what we nowadays call a metric space (following Hausdorff [84, p. 211]). It is called a "classe (E)" by Fréchet, because here he uses the term "écart" instead of "voisinage". For such écart-classes or E-classes Fréchet proves the theorem: If a subset G of an E-class is such that every continuous functional operation on G is bounded on G and attains on G its least upper bound, then G is closed and compact. The E-class was actually introduced because Fréchet could not prove this theorem for V-classes.[19] The thesis also contains a generalization of the Heine–Borel Covering Theorem: If E is a closed and compact subset of a V-class then every countable covering $\mathcal{M}$ of E contains a finite number of sets that also cover E. In the second part of his thesis Fréchet applies the abstract theory to concrete examples.

2.4. *Hausdorff's definition of a Hausdorff space*

2.4.1. *Hilbert.* While Fréchet was developing a theory of abstract spaces, others were doing similar things. Before Fréchet started working on abstract spaces, Hilbert in 1902 briefly wrote about the possibility to characterize the notion of manifold in an abstract

[18] For a more extensive discussion we refer to Taylor [170].

[19] In 1908 Hahn succeeded in doing so and in 1917 Chittenden [47] turned it into a metrisation theorem.

way, while in 1906 in Hungary, independent of Fréchet, Frigyes Riesz (1880–1956) also attempted to give as general as possible a characterization of the notion of space. In Hilbert's proposal the notion of neighbourhood is central. Hilbert wrote:

> The plane[20] is a system of things that are called points. Every point A determines certain sub-systems of points to which the point itself belongs and that are called neighbourhoods of the point A. The points of a neighbourhood can always be mapped by means of a one-to-one correspondence on the points of a certain Jordan-area in the number plane. The Jordan-area is called the image of that neighbourhood. Every Jordan-area, that contains (the image of) A, and is contained in an image, is also image of a neighbourhood of A. If different images of a neighbourhood are given, then the resulting mapping of the two corresponding Jordan-areas on each other is continuous. If B is any point in a neighbourhood of A, this neighbourhood is also a neighbourhood of B. To any two neighbourhoods of A always corresponds such neighbourhood of A, that the two neighbourhoods have in common. When A and B are any two points of the plane, there exists always a neighbourhood of A that contains at the same time B. These requirements contain, it seems to me, for the case of two dimensions, the sharp definition of the notion that Riemann and Helmholtz denoted as "multiply extended manifold" and Lie as "number manifold", and on which they based their entire investigations. They also offer the foundation for a rigourous axiomatic treatment of the analysis situs.[21]

The quotation, which contains everything that Hilbert wrote about the subject, dates from 1902, that is from before Fréchet started his topological work. Hilbert never continued the line of research that the quotation suggested. He left the further "rigourous axiomatic treatment of the analysis situs" to others. The axioms define an abstract notion of space and the basic concept is the concept of neighbourhood. Some of the axioms only concern the set theoretic properties of neighbourhoods. The other properties of the neighbourhoods are, however, fixed by means of axioms concerning the (continuous) one-one correspondences that are postulated to exist between neighbourhoods and Jordan-areas in the number-plane.

2.4.2. *Riesz.* Riesz' approach to the problem and also his motivation are quite different. In his 1907 paper (a German translation of a Hungarian paper that appeared in 1906) Riesz distinguishes our subjective experience of time and space from the mathematical continua by means of which we describe them. His goal is to give as general a characterisation as possible of mathematical continua and to show the precise relation between our subjec-

---

[20] As will be clear later, Hilbert uses the notion of plane in a generalised sense.

[21] "Die Ebene ist ein System von Dingen, welche Punkte heißen. Jeder Punkt A bestimmt gewisse Teilsysteme von Punkten, zu denen er selbst gehört und welche Umgebungen des Punktes A heißen. Die Punkte einer Umgebung lassen sich stets umkehrbar eindeutig auf die Punkte eines gewissen Jordanschen Gebietes in der Zahlenebene abbilden. Das Jordansche Gebiet wird ein Bild jener Umgebung genannt. Jedes in einem Bilde enthaltene Jordansche Gebiet, innerhalb dessen der Punkt A liegt, ist wiederum ein Bild einer Umgebung von A. Liegen verschiedenen Bilde einer Umgebung vor, so ist die dadurch vermittelte umkehrbar eindeutige Transformation der betreffenden Jordanschen Gebiete aufeinander eine stetige. Ist B irgendein Punkt in einer Umgebung von A, so ist diese Umgebung auch zugleich eine Umgebung von B. Zu irgend zwei Umgebungen eines Punktes A gibt es stets eine solche Umgebung des Punktes A, die beiden Umgebungen gemeinsam ist. Wenn A und B irgend zwei Punkte unserer Ebene sind, so gibt es stets eine Umgebung von A, die zugleich den Punkt B enthält. Diese Forderungen enthalten, wie mir scheint, für den Fall zweier Dimensionen die scharfe definition des Begriffes, den Riemann und Helmholtz als "mehrfach ausgedehnte Mannigfaltigkeit" und Lie als "Zahlenmannigfaltigkeit" bezeichneten und ihren gesamten Untersuchungen zugrunde legten. Auch bieten sie die Grundlage für eine strenge axiomatische Behandlung der Analysis situs." (Hilbert [88, pp. 165–166].)

tive experience and mathematical continua. In a footnote Riesz criticises the way in which philosophers have dealt with notions like continuous and discrete and he repeats Russell's remark about the followers of Hegel: "the Hegelian dictum (that everything discrete is also continuous and vice versa) has been tamely repeated by all his followers. But·as to what they meant by continuity and discreteness, they preserved a discrete and continuous silence; [...]" (Riesz [147]). The relation of our subjective experience of space and time and mathematical continua is described by Riesz as follows. Mathematical continua possess certain properties of continuity, coherence and condensation. On the other hand, our subjective experience of time is discrete and consists of countable sequences of moments. Systems of subsets of a mathematical continuum can be interpreted as a physical continuum when subsets with common elements are interpreted as undistinguishable and subsets without common elements as distinguishable. Riesz [147, p. 111] is an interesting paper in which Riesz, who had read Fréchet's work and appreciated it, developed a different theory of abstract spaces, based on the notion of "Verdichtungsstelle", i.e. "condensation point" or, as we will translate "limit point". In his theory Riesz succeeded in deriving the Bolzano–Weierstrass Theorem and the Heine–Borel Theorem. We will not discuss this paper. We will restrict ourselves to a shorter paper that was presented by Riesz in 1908 at the International Congress of Mathematicians in Rome. In that paper, "Stetigkeit und Abstrakte Mengenlehre" (Riesz [148]), concentrates on the characterisation of mathematical continua. We will briefly describe some of the ideas that Riesz describes in the paper. As we said, Riesz' basic notion is the notion of limit point (Verdichtungsstelle). Riesz did consider Frechet's restriction to limit points of countable sequences as too severe. That is, why in his theory limit points satisfy the following three axioms:

(i) Each element that is a limit point of a subset M is also a limit point of every set containing M.
(ii) When a subset is divided into two subsets, each limit point is a limit point of at least one of the subsets.
(iii) A subset consisting of only one element does not have a limit point.

A mathematical continuum is for Riesz, by definition, any set for which a notion of limit point is defined that satisfies these three axioms. However, in order to be able to develop some theory on the basis of the axioms Riesz is forced to add a fourth axiom:

(iv) Every limit point of a set is uniquely determined through the totality of its subsets for which it is a limit point.

Riesz uses the examples of $\mathbb{R} \setminus \{0\}$ and $\mathbb{R} \setminus [0, 1]$, that exhibit as far as their limit points are concerned precisely the same structure, to show that the four axioms are not enough to characterize properties of "continuity". That is why in his paper Riesz suggests to add the notion of "linkage" (Verkettung). For any pair of subsets of a manifold it should be defined whether they are linked or not. Such a linkage structure must satisfy the following three axioms:

(i) If subsets S_1 and S_2 are linked, then every pair of sets that contain S_1 and S_2 are also linked.
(ii) If subsets S_1 and S_2 are linked and S_1 is split into two subsets, at least one of the two is linked to S_2.
(iii) Two sets that each contain only one element cannot be linked.

Although he believed that the notions of limit point and linkage could be used to develop an abstract theory of sets in the sense of Hadamard's proposal of 1897, Riesz himself did not continue this work. In later publications Fréchet used some of Riesz' ideas.

2.4.3. *Hausdorff.* The different attempts to give an abstract definition of space culminated in the work of Felix Hausdorff (1868–1942). In 1912 Hausdorff, professor at the university of Bonn, taught a class on set theory. Chapter 6 of his notes[22] deals with "Point sets" ("Punktmengen") and is called "Neighbourhoods" ("Umgebungen"). Hausdorff writes:

> Point sets on a straight line (linear), in the plane (planar), in space (spatial), in general in an n-dimensional space $r = r_n$. A *point* is defined by a system of n real numbers $(x_1, x_2, \ldots, x_n)$ and vice versa, that we think as orthogonal coordinates. As *distance* of two points we define
>
> $$x \cdot y = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2} \geqslant 0.$$
>
> The *neighbourhood* U_x of a point x is the collection of all points y for which $x \cdot y < \rho$ (ρ a positive number; the inner area of a "Sphere" with radius ρ).
>
> For the sake of illustration we will usually take the plane $r = r_2$; if the individual numbers of dimensions cause deviations, we will especially emphasize them. The neighbourhoods have the following properties:
>
> (α) Every U_x contains x and is contained in r.
>
> (β) For two neighbourhoods of the same point $U'_x \supseteq U_x$ or $U_x \supseteq U'_x$ holds.
>
> (γ) If y lies in U_x, then there also exists a neighbourhood U_y, that is contained in U_x ($U_x \supseteq U_y$).
>
> (δ) If $x \neq y$, then there exist two neighbourhoods U_x, U_y without a common point: ($\theta(U_x, U_y) = 0$).
>
> The following considerations are based initially only on these properties. They hold very generally, if r is a point set $\{x\}$, if to the points x correspond point sets U_x with these 4 properties. Such a system is, for example, the following: one defines as a neighbourhood of x the system of points where
>
> $$|x_1 - y_1| < \rho, \qquad y_2 = x_2;$$
>
> a neighbourhood is then a horizontal segment (without endpoints) of length 2ρ. Or: one defines as a neighbourhood the system
>
> $$|x_1 - y_1| < \rho, \qquad |x_2 - y_2| < \rho,$$
>
> i.e. the inner area of a square with side-length 2ρ, whose centre is x, etc.[23]

---

[22] Hausdorff, manuscript 1912b, par. 6, Archive Bonn University.

[23] "Punktmengen auf einer Geraden (linear), in der Ebene (ebene), im Raume (räumliche), allgemein in einem n-dimensionalen Raume $r = r_n$. Ein Punkt x ist durch ein System von n reellen Zahlen $(x_1, x_2, \ldots, x_n)$ und umgekehrt definiert, die wir als rechtwinklige Coordinaten denken. Als *Entfernung* zweier Punkte definieren wir $x \cdot y = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2} \geqslant 0$. Unter einer Umgebung U_x des Punktes x verstehen wir den Inbegriff aller Punkte y für die $x \cdot y < \rho$ (ρ eine positive Zahl; Inner[e]s einer "Kugel" mit Radius ρ). Wir werden zur Veranschaulichung in der Regel die Ebene $r = r_2$ nehmen; sollten die Einzelnen Dimensionenzahlen Abweichungen hervorrufen, so werden die besonders hervorgehoben werden. Die Umgebungen haben folgende Eigenschaften: (α) Jedes U_x enthält x und ist in r enthalten. (β) Für zwei Umgebungen desselben Punktes ist $U'_x \supseteq U_x$ oder $U_x \supseteq U'_x$. (γ) Liegt y in U_x, so giebt es auch eine Umgebung U_y, die in U_x enthalten ist $U_x \supseteq U_y$). (δ) Ist $x \neq y$, so giebt es zwei Umgebungen U_x, U_y ohne gemeinsamen Punkt ($\theta(U_x, U_y) = 0$). Die folgenden Betrachtungen stützen sich zunächst nur auf diese Eigenschaften. Sie gelten sehr allgemein, wenn r eine Punktmenge $\{x\}$ ist, wenn Punkten x Punktmengen U_x zugeordnet sind mit diesen 4 Eigenschaften. Ein solches System ist z. B. folgendes: man definiere als ein Umgebung von x das System der Punkte, wo $|x_1 - y_1| < \rho$, $y_2 = x_2$; eine Umgebung ist dann eine horizontale Strecke (ohne Randpunkte) von der Länge 2ρ. Oder: als Umgebung werde das System $|x_1 - y_1| < \rho$, $|x_2 - y_2| < \rho$ definiert, d.h. das Inner eines Quadrates von der Seitenlänge 2ρ, dessen Mittelpunkt x ist, u.s.w."

According to Scholz [156], Hausdorff was led to the four axioms in the spring or the summer of 1912 by a logical analysis of the foundations of complex analysis. In this context it is remarkable that at the same time, Weyl was applying Hilbert's ideas from 1902 in his work his on Riemann surfaces (Weyl [188]). Scholz argues that both were independently influenced by Hilbert. In 1914 Hausdorff's "Grundzüge der Mengenlehre" appeared, one of the first textbooks on set theory. Above we saw how set theory was born from point set theory in $\mathbb{R}^n$ and that Cantor's first papers show a mixture of point set theory and more abstract considerations. Hausdorff carefully distinguishes general set theory from point set theory. The first seven chapters of his book are devoted to general set theory. It is remarkable that he did not include Zermelo's axiomatization. In the first chapter, after mentioning the antinomies, he writes why not:

> E. Zermelo undertook the subsequently necessary attempt to limit the borderless process of set-creation by suitable restrictions. Because so far these extremely shrewd investigations can not yet claim to be finished and an introduction of the beginner in set theory in this way would be connected with great difficulties, we will permit here the naive notion of set, at the same time, however, we will in fact stick to the restrictions that cut off the road to that paradox.[24]

In Chapter 7 of his book, Hausdorff addresses the question of the position of point set theory within the system of general set theory. Point set theory here means abstract point set theory. He briefly discusses three possible approaches to turn a set that is so far treated purely as a system of its elements without considering relations between the elements, into a space. His goal is obviously to define a very general notion of space that encompasses not only the $\mathbb{R}^n$, but also Riemann surfaces, spaces of infinitely many dimensions and spaces the elements of which are curves or functions (Hausdorff [84, p. 211]). He gives two advantages of such a general notion: it simplifies theories considerably and it prevents us from illegitimately using intuition (die Anschauung). The first possibility is to base point set theory on the notion of the distance (Entfernung) of two elements, that is a function that associates with each pair of elements of a set a particular value. Hausdorff remarks that on the basis of the notion of distance the notion of a converging sequence of points and its limit can be defined. Moreover, on the basis of the notion of distance, one can also associate with each point of a set subsets of the space called neighbourhoods of the point.

However, one can also turn a set into a space by circumventing the notion of distance and starting from a function $f(a_1, a_2, a_3, \ldots, a_k, \ldots)$ which maps certain sequences of elements (the converging sequences) of the set M on elements of M (the limits of the sequences). Thirdly, one can also start with the notion of neighbourhood. Formally one then maps every element of a set M on certain subsets of M that are called the neighbourhoods of the element. Which of the three "spatial" notions one chooses as the most fundamental is for Hausdorff to a certain extent a matter of taste (Hausdorff [84, p. 211]). Neighbourhoods and limits can be defined in terms of distances. By means of neighbourhoods one can define limits, but in general no distances. By means of limits one can define neither neighbourhoods nor distances. Hausdorff writes; "Thus the distance theory seems to be

[24] "Den hiernach notwendigen Versuch, den Prozeß der Uferlosen Mengenbildung durch geeignete Forderungen einzuschränken, hat E. Zermelo unternommen. Da indessen diese äußerst scharfsinnigen Untersuchungen noch nicht als abgeschlossen gelten können und da eine Einführung des Anfangers in die Mengenlehre auf diesem Wege mit großen Schwierigkeiten verbunden sein dürfte, so wollen wir hier den naiven Mengenbegriff zulassen, dabei aber tatsächlich die Beschränkungen innehalten, die den Weg zu jenem Paradoxon abschneiden." (Hausdorff [84, p. 2].)

the most special and the limit theory the most general; on the other hand, the limit theory creates immediately a relation with the countable (with sequences of elements), which the neighbourhood theory avoids." (Hausdorff [84, p. 211].)

As a good teacher he now first gives an example. He defines metric spaces by means of the well-known three axioms. $\mathbb{R}^n$ with the Euclidean distance is an example of a metric space. Hausdorff concentrates on the four properties of the spherical neighbourhoods that he had already given in his 1912 lectures. He writes:

> A topological space is a set E in which the elements (points) x are mapped on certain subsets U_x, that we call neighbourhoods of x, in accordance with the following neighbourhood axioms [...].[25]

He then gives the four axioms that occur already in his 1912 lectures and he shows that the spherical neighbourhoods in $\mathbb{R}^n$ satisfy the axioms.

Hausdorff's generalization of the notion of space represented a major contribution to the unification of mathematics. Geometry and analysis had been separate disciplines. Axiomatization ended that. Hausdorff succeeded in picking a set of axioms that was, on the one hand, general enough to handle abstract spaces and, on the other hand, restrictive enough to yield an interesting theory. He succeeded in giving a theory of topological and metric spaces that encompassed the previous results and generated many new notions and theorems.

2.4.4. *L.E.J. Brouwer.* Above we sketched the genesis of the notion of topological space as it was finally defined by Hausdorff. His book was very influential. For years it was an important source for many mathematicians. Yet our story, which is so far restricted to the genesis of the notion of topological space, is very one-sided. In order to do some more justice to the actual development, the contributions of Brouwer must be mentioned. Brouwer's approach to general topology is totally different from Hausdorff's. Also their views of mathematics were completely different. Hausdorff was a great supporter of the axiomatic method. Brouwer rejected the axiomatic method and argued that mathematics ought to be founded in intuition.[26]

In the first decade of this century Arthur Schoenflies had attempted to give a thorough set-theoretic foundation of topology.[27] In Schoenflies' work a central result is Jordan's Theorem: a closed *Jordan curve*, i.e. the one-to-one continuous image of a circle, divides the plane into two domains with the image as their common boundary. A *domain* is an open connected set. At certain points Schoenflies work is quite subtle. For example, he distinguishes between simple closed curves and closed curves that are not simple by means of the notion of accessibility. By definition a point P on the boundary of a domain is *accessible* if it can be reached from an arbitrary point in the domain by a finite or an infinite polygonal path in the domain. A *closed curve* is here by definition a bounded closed point set that divides the plane into two domains with the curve as their common boundary (Schoenflies [155, pp. 118–120]). Closed curves that are such that all their points are accessible from the two domains are called *simple* by Schoenflies. An important result that he proved

[25] "Unter einem topologischen Raum verstehen wir eine Menge E, worin den elementen (Punkten) x gewisse Teilmengen U_x zugeordnet sind, die wir Umgebungen von x nennen, und zwar nach Maßgabe der folgenden Umgebungsaxiome [...]" (Hausdorff [84, p. 213]).

[26] See also Koetsier and van Mill [108].

[27] For a fuller treatment we refer to Johnson [98, 99].

is the following: simple closed curves are closed Jordan curves. In his early work Brouwer relied on Schoenflies' results. However, in the winter of 1908–1909 he discovered suddenly that Schoenflies' results were not reliable. In [35], entitled "Zur Analysis Situs", he gave a series of devastating counterexamples. Brouwer does not criticise Schoenflies' theory of simple closed curves, but attacks his more general theory of closed curves. In the paper he gave the sensational example of a closed curve that splits the plane into three domains of which it is the common boundary. It is also the first example of an indecomposable continuum. Schoenflies' general theory of closed curves and domains had to be rejected entirely. Soon Brouwer produced several other highly original papers. We will mention only two: ("Beweis der Invarianz der Dimensionenzahl", submitted in June 1910 and published in 1911 (Brouwer [36]) and his paper [37]. The first paper marks, according to Freudenthal, the onset of a new period in topology. Although the paper is very short and merely contains a simple proof of the invariance of dimension, "it is in fact much more than this – the paradigm of an entirely new and highly promising method, now known as *algebraic topology*. It exhibits the ideas of simplicial mapping, barycentric extension, simplicial approximation, small modification, and, implicitly, the mapping degree and its invariance under homotopic change, and the concept of homotopy class." (Freudenthal [68, p. 436]).[28] In the second paper Brouwer proved the basic theorem on fixed points: every continuous transformation of the n-simplex into itself possesses at least one fixed point. Although Brouwer's results were reached from a totally different philosophical position his results could be easily incorporated in and considerably enriched the axiomatic framework created by Hausdorff. This led to much further work.

2.4.5. *Functional analysis.* Brouwer's work shows how problems in point set topology led to algebraic topology. We will, however, use the example of Brouwer's fixed point theorems to illustrate another way in which results from general topology penetrated other areas of mathematics.

Between Hilbert's, Fréchet's and Riesz' first attempts and the publication of Hausdorff's book the number of "sets with a spatial character", deviating from Euclidean space, had grown, and with it the potential value of abstract characterisations of the notion of space. For example, between 1904 and 1910 Hilbert had published his six famous "communications" on the foundations of the theory of integral equations. The space ℓ^2 had gradually become the object of investigation. The proof of the isomorphism of ℓ^2 and L^2, the space of quadratic Lebesgue-integrable functions, led to the notion of Hilbert space. In 1910 Riesz introduced the normed linear function space L^p. That work meant also the start of modern operator theory.[29]

In his Lwów dissertation of 1920 Banach introduced the notion of a "Banach space" (the name is Fréchet's). In the 1920's and 1930's the Polish school carefully applied set-theoretic methods to functional analysis and proved fundamental theorems like the Hahn–Banach Theorem and the Banach Fixed-Point Theorem. With Banach's "Théorie des opérations linéaires" of 1932 functional analysis was established as one of the central fields in modern analysis. Banach's student J.P. Schauder (Studia Mathematica **2**, 1930, pp. 170–179) and Schauder and J. Leray (Ann. de l'École Normale Superieure **51**, 1934, pp. 45–78)

[28] For a more extensive treatment of Brouwer's work in dimension theory we refer to Johnson [98, 99]. See also Koetsier and van Mill [108]. For Brouwer's topological work as a whole we refer to Freudenthal [68].

[29] For a more extensive survey of the history of functional analysis at the beginning of the century we refer to Siegmund-Schultze [163].

carried over Brouwer's topological notions into infinite-dimensional spaces and generalized his fixed point theorem in order to establish the existence of solutions of nonlinear differential equations. This work was of great importance for the development of nonlinear functional analysis in the 1950's.

3. Intermezzo: the golden age

The next phase in the history of general topology, its golden age, lasted roughly from the 1920's until the 1960's. Among the main themes in this period were dimension theory, paracompactness, compactifications and continuous selections. The important results and the way in which they are related can all be found in a number of classical textbooks (see below). About this phase we will be very brief.

Dimension theory was fully developed (see Hurewicz and Wallman [94] for a beautiful survey of dimension theory until 1941). In the late 1940's and early 1950's paracompactness, introduced by Dieudonné [57], was the leading theme in general topology. Stone [168] proved that metrizable spaces are paracompact and Nagata, Smirnov and Bing published/proved their metrization theorems in [140, 164, 29], respectively. The work on compactifications in the 1950's culminated in the publication of the beautiful book [80] by Gillman and Jerison. Michael [128–130] developed his theory of continuous selections. For more information, see, e.g., Hu [92], Dugundji [65], Nagata [142, 141], Engelking [71, 72] and Arhangel'skiĭ and Ponomarev [13].

We will turn now to the third period that we distinguish in the history of general topology: the period of harvesting. We will concentrate in Sections 4 and 5 on two major areas of research that developed out of the golden age, infinite-dimensional topology and set theoretic topology, and show how these solved difficult problems outside of general topology.

The style of Section 2 was rather informal, in keeping with the pioneering works of the area. In Sections 4 and 5 we attempt to describe some complex results from the front line of mathematics. In order to do so we will use the much more compressed, conceptual style of modern mathematics.

4. The period of harvesting: infinite-dimensional topology

4.1. *The beginning*

As usual, let a *separable Hilbert space* be the set

$$\ell^2 = \left\{x \in \mathbb{R}^\infty \colon \sum_{i=1}^{\infty} x_i^2 < \infty\right\}$$

endowed with the norm

$$\|x\| = \sqrt{\sum\nolimits_{i=1}^{\infty} x_i^2}. \tag{4.1}$$

The metric derived from (4.1) is complete and hence ℓ^2 is a complete linear space.[30]

[30] A *linear space* in this article is a real topological vector space.

By using convexity type arguments, Klee [106] proved that $\ell^2 \setminus \{pt\}$ and ℓ^2 are homeomorphic. We say that points can be *deleted* from ℓ^2. In fact, he proved in that even arbitrary compact sets can be deleted from any infinite-dimensional normed linear space. This result demonstrates a striking difference between finite-dimensional and infinite-dimensional normed linear spaces. For a finite-dimensional linear space is equivalent to some $\mathbb{R}^n$ and no point can be deleted from $\mathbb{R}^n$, since $\mathbb{R}^n \setminus \{pt\}$ is not contractible.

Klee's results were later substantially simplified by the approach of Bessaga [22, 23] who proved, among other things, that if an infinite-dimensional linear space admits a C^k-differentiable norm (except at 0) which is not complete, then the deleting homeomorphisms can in fact be chosen to be diffeomorphisms of class C^k.

Motivated by the results of Klee, Anderson [9] studied in 1967 the question which sets can be deleted from another classical linear space, namely the countable infinite topological product of real lines $\mathbb{R}^\infty$ (see Section 2.3.2). This space is denoted by s. Its topology is generated by the following complete metric:

$$d(x, y) = \sum_{n=1}^{\infty} 2^{-n} \frac{|x_n - y_n|}{1 + |x_n - y_n|}. \tag{4.2}$$

So s is a locally convex complete metrizable linear space. Such a space is called a *Fréchet space* in the literature. Unfortunately, s has an unpleasant defect: its topology is not *normable* in the sense that there is no norm $\|\cdot\|$ on it so that the metrics

$$\rho(x, y) = \|x - y\|$$

and d in (4.2) are equivalent. This is clear once one realizes that every neighbourhood of the origin contains a nontrivial (linear) subspace of $\mathbb{R}^\infty$.

The linear structure on s is therefore very different from the linear structure on a normed linear space, and so the methods of Klee and Bessaga do not apply if one wishes to prove results on the possibility of deleting sets. But by using a completely different method, Anderson [9] showed that from s one can delete sets as easily as from ℓ^2. In fact, he got the following remarkable result:[31]

THEOREM 4.1.1. *Let X be any separable metrizable space. Then every σ-compact set can be deleted from $X \times s$.*

A new field in topology was born: it was called *infinite-dimensional topology*.

Anderson was motivated by purely intrinsic topological questions. Soon however it turned out quite unexpectedly that his methods could be used to solve a classical open problem, posed by Fréchet [78, pp. 94–96] in 1928. In 1932 in [19, p. 233], Banach stated that Mazur had solved the problem, but this claim turned out to be incorrect. Subsequently it was understood that the question was still open.

To put the question into perspective, let us first make a few remarks. The spaces s and ℓ^2 are both natural generalizations of the finite-dimensional Euclidean spaces $\mathbb{R}^n$, but their

[31] As usual, a space is σ-compact if it can be written as a union of countably many compact subspaces.

linear structures are notably different. There does not exist a homeomorphism $h : s \to \ell^2$ which is *linear*, i.e. has the property that

$$h(\lambda x + \mu y) = \lambda h(x) + \mu h(y)$$

for all $x, y \in s$, and $\lambda, \mu \in \mathbb{R}$. The question therefore naturally arises whether s and ℓ^2 are (topologically) homeomorphic at all. The question of Fréchet and Banach is much more elaborate, it asks whether all infinite-dimensional Fréchet spaces are homeomorphic.

The question had a long history when Anderson considered it in 1966. By several ad hoc methods, homeomorphy of many linear spaces had already been established. The first relevant result is due to Mazur [126] who proved in 1929 that all spaces L^p and ℓ^p for $1 \leqslant p < \infty$ are homeomorphic to ℓ^2. Then Kadec in a series of papers developed an interesting "renorming technique" for separable Banach spaces and finally proved in 1965 that all infinite-dimensional separable Banach spaces are homeomorphic (see Kadec [101]). Kadec's proof used the result of Bessaga and Pełczyński [25] that a separable Banach space containing a linear subspace homeomorphic to ℓ^2 is in fact itself homeomorphic to ℓ^2. This result combined with another result of Bessaga and Pełczyński [26] showed that the homeomorphy of s and ℓ^2 would imply the positive answer to Fréchet's question, i.e. the homeomorphy of all separable infinite-dimensional Fréchet spaces. The proofs of these interesting results combine techniques from functional analysis, especially the geometry of Banach spaces, with various ingenious arguments from general topology.

This final, but crucial, open problem was solved in the affirmative by Anderson [7] using the results from his previous paper [9]. He proved that s and ℓ^2 are homeomorphic and hence settled the question of Fréchet and Banach in the affirmative.

4.2. *The Hilbert cube Q*

Let Q denote the product $\prod_{n=1}^{\infty}[-1,1]_n$ of countably many copies of $[-1,1]$. The topology on Q is the Tychonoff product topology. Alternatively, its topology is generated by the metric

$$d(x,y) = \sum_{n=1}^{\infty} 2^{-n} \cdot |x_n - y_n|.$$

So Q is a compact metrizable space. Geometrically one should think of it as an infinite-dimensional brick the sides of which get shorter and shorter. This can be demonstrated in the following way. Let $x(n) \in Q$ be the point having all coordinates 0 except for the n-th coordinate which equals 1. So $x(n)$ is the "endpoint" of the n-th axis in Q. In addition, let y be the "origin" of Q, i.e. the point all coordinates of which are 0. Intuitively, each $x(n)$ has distance 1 from y and hence $x(n)$ and y are far apart. However, the appearance of the factor 2^{-n} in the definition of d implies that

$$d\big(x(n), y\big) = 2^{-n},$$

whence the sequence $(x(n))_n$ converges to y in Q.

It can be shown that Q is homeomorphic to the subspace

$$\left\{x \in \ell^2 \colon (\forall n \in \mathbb{N})\left(|x_n| \leqslant \frac{1}{n}\right)\right\}$$

of ℓ^2.

The first paper in infinite-dimensional topology is in fact Keller's paper [104] from 1931. In that paper it is shown that all infinite-dimensional compact convex subsets of ℓ^2 are homeomorphic to Q, and also that Q is *topologically homogeneous*, i.e. for all $x, y \in Q$ there exists a homeomorphism $f: Q \to Q$ with $f(x) = y$. This last result is at first glance very surprising since the finite-dimensional analogues $\mathbb{I}^n$ of Q are not homogeneous. For $n = 1$ this is a triviality, and for larger n this boils down to the Brouwer Invariance of Domain Theorem.

A familiar construction in topology is that of the *cone* over a locally compact space X, it is the one-point compactification of the product $X \times [0, 1)$. The compactifying point is called the *cone point* of the cone which itself is denoted cone(X).[32]

Now, it is clear that for each n the cone over $\mathbb{I}^n$ is homeomorphic to $\mathbb{I}^{n+1}$ and so naively one would expect, by taking the "limit" as n goes to infinity, that cone(Q) $\approx Q$. That this is indeed true follows from Keller's first result because we can realize cone(Q) as a compact convex subset of ℓ^2.

Since Q is contractible, the cone point in cone(Q) has arbitrarily small neighbourhoods with contractible boundaries. This is not surprising since every point on the boundary of $\mathbb{I}^n$ has the same property. However, points in the interior of $\mathbb{I}^n$ do not have this property. But since Q is homogeneous, *every* point of Q has arbitrarily small neighborhoods with contractible boundaries. This is again a striking difference with the finite-dimensional situation.

At the time Keller made his fundamental observations, they apparently did not get the credit they deserved for they did not play any significant role for approximately thirty-five years. Maybe, but this is speculation on the part of the authors of the present paper, in the thirties Keller's results were thought of as mere curiosities. Infinite-dimensional topology took approximately thirty-five more years to finally come to real existence. In that process, the work of Anderson was vital.

4.3. *Homeomorphism extension results in Q-manifolds*

In [9], Anderson also proved results on the possibility of extending homeomorphisms in Q. It was known already that if X is any countable closed subset of Q then any homeomorphism $f: X \to X$ can be extended to a homeomorphism of Q (see Keller [104], Klee [105] and Fort [73]). In the subsequent paper [8], Anderson introduced the fundamental concept of a Z-set in Q and proved that any homeomorphism between such sets can be extended to a homeomorphism of Q.

Before we present the definition of a Z-set, we make some remarks. Let K denote the familiar Cantor middle-third set in $\mathbb{I}$. It is known that it is characterized by the follow-

32 There are other constructions of cones that work for general spaces. One then considers the product $X \times \mathbb{I}$ and identifies the set $X \times \{1\}$ to a single point. But this cone is, in general, not metrizable.

ing topological properties: K is a compact, metrizable, zero-dimensional space[33] without isolated points. So it follows easily that $K \times K \approx K$ from which it follows that K contains a nowhere dense closed copy of itself, say X. It also contains a copy of itself having nonempty interior, namely K itself. Consider any homeomorphism $\varphi : X \to K$. Then it cannot be extended to a homeomorphism $\overline{\varphi} : K \to K$ for obvious reasons. One of them being that K is fat, having nonempty interior, and X is small, having empty interior.

A space homeomorphic to K is called a Cantor *set.*

Similar remarks apply to other spaces as well. It is known that any homeomorphism φ between Cantor sets in $\mathbb{R}^2$ can be extended to a homeomorphism of $\mathbb{R}^2$. However, Antoine's necklace X is a Cantor set in $\mathbb{R}^3$ whose complement is not simply connected and so no homeomorphism $\varphi : X \to Y$, where Y is a Cantor subset of the x-axis of $\mathbb{R}^3$, can be extended to a homeomorphism of $\varphi : \mathbb{R}^3 \to \mathbb{R}^3$. This phenomenon occurs in Q as well: in [191] Wong constructed a wild Cantor set in Q.

So for a homeomorphism extension theorem, one needs a class of tame subspaces which are flexible enough to perform the required constructions. For Q this class was identified by Anderson. He called a closed subset A of Q a *Z-set*[34] provided that for every nonempty homotopically trivial open subset $U \subseteq Q$ the set $U \setminus A$ is nonempty and homotopically trivial as well. He proved in [8] the following fundamental homeomorphism extension theorem:

THEOREM 4.3.1. *If $\varphi : A \to B$ is a homeomorphism between Z-sets in Q then there exists a homeomorphism $\overline{\varphi} : Q \to Q$ extending φ.*

In the proof important ideas of Klee [105] were exploited.

Later, Barit [20] observed that if the homeomorphism φ satisfies $d(\varphi, \mathrm{id}) < \varepsilon$ for some $\varepsilon > 0$ then the extension $\overline{\varphi}$ can be chosen to satisfy the same smallness condition.

The final result on the possibility of extending homeomorphisms in manifolds modeled on Q is due to Anderson and Chapman [11]. Let X be a space and let $f : X \to X$ be a function. If $\mathcal{U}$ is an open cover of X then we say that f is *limited by $\mathcal{U}$* provided that for every $x \in X$ there exists $U \in \mathcal{U}$ containing both x and $f(x)$. Here is the Anderson–Chapman Homeomorphism Extension Theorem from 1971:

THEOREM 4.3.2. *Let M be a manifold modeled on Q and let $A, B \subseteq M$ be Z-sets. If $\varphi : A \to B$ is a homeomorphism and $\mathcal{U}$ is an open cover of M such that φ is limited by it, then there exists a homeomorphism $\overline{\varphi} : M \to M$ extending φ which is also limited by $\mathcal{U}$.*

This is a purely topological result belonging to general topology and at the time Anderson and Chapman proved it, they could not have foreseen what potential this theorem turned out to have. We will report on this later.

---

33 Here a space is called zero-dimensional if it has a base consisting of open and closed sets.

34 One of the authors of the present paper once asked Anderson why he chose this terminology. He replied that he had no idea.

4.4. *Identifying Hilbert cubes*

In 1964, Anderson [6] proved that the Q is homeomorphic to any countably infinite product of dendrons.[35] In particular, one gets the curious result that if T denotes

$$(\mathbb{I} \times \{0\}) \cup \left(\left\{\frac{1}{2}\right\} \times \mathbb{I}\right)$$

then $T \times Q$ and Q are homeomorphic. For a published proof of Anderson's result, see West [185]. So the Hilbert cube surfaces at various places, not only as convex objects such as in Keller's theorem cited above. The result started the game of identifying Hilbert cubes. It was a very fascinating game. The tools were from general topology with special emphasis on geometric methods.

The hyperspace 2^X of a compact space X is the space consisting of all nonempty closed subsets of X with topology generated by the Hausdorff metric d_{H} defined by

$$d_{\mathrm{H}}(A, B) = \inf\{\varepsilon > 0\colon\ A \subseteq D_\varepsilon(B) \text{ and } B \subseteq D_\varepsilon(A)\};$$

here $D_\varepsilon(A)$ means the open ball about A with radius ε. Hyperspaces were first considered in the early 1900's in the work of Hausdorff and Vietoris. In 1939 Wojdysławski [190] asked whether for every Peano continuum X the hyperspace 2^X is homeomorphic to Q. At the time of the conjecture this was a rather bold question because the only nontrivial Hilbert cubes that were identified at that time were Keller's infinite-dimensional compact and convex subsets of ℓ^2. In [157] Schori and West proved that $2^{\mathbb{I}}$ is homeomorphic to Q and in [52] Curtis and Schori completed the picture by showing that 2^X is homeomorphic to Q if and only if X is a Peano continuum. This was a spectacular result at that time and fully demonstrated the power and potential of infinite-dimensional topology.

4.5. *Hilbert cube manifolds*

In the early seventies, Chapman began the study of spaces modeled on Q, the so called Hilbert cube manifolds or Q-manifolds. Certain delicate finite-dimensional obstructions turned out not to appear in Q-manifold theory. In some vague sense, Q-manifold theory is a "stable" PL n-manifold theory.

We already mentioned the important homeomorphism extension result Theorem 4.3.2. Using this result, and several ingenious geometric constructions, Chapman developed the theory of Q-manifolds. It was known from previous work that if P is a polyhedron then $P \times Q$ is a Q-manifold. Chapman [43] proved the converse, namely that all Q-manifolds are of this form, a result that turned out to be of fundamental importance later.

Some truly spectacular results were the result of Chapman's efforts. In 1974 he used Q-manifold theory to prove the invariance of Whitehead torsion. This is the statement that any homeomorphism between compact polyhedra is a simple homotopy equivalence. A map $f : X \to Y$ of compact polyhedra is a *simple homotopy equivalence* if it is homotopic to a finite composition

$$X \xrightarrow{f_1} X_1 \xrightarrow{f_2} X_2 \longrightarrow \cdots \xrightarrow{f_n} X_n \xrightarrow{f_{n+1}} Y,$$

[35] A dendron is a uniquely arcwise connected Peano continuum.

where each X_i is a compact polyhedron and each f_i is either an elementary expansion or an elementary collapse. Thus a simple homotopy equivalence is a homotopy equivalence of a very special nature. It is one which can be resolved into a finite number of elementary moves. More specifically, Chapman proved in [44]:

THEOREM 4.5.1. *A map $f : X \to Y$ between compact polyhedra is a simple homotopy equivalence if and only if $f \times \{\mathrm{id}\} : X \times Q \to Y \times Q$ is homotopic to a homeomorphism.*

There is also a version of this result for noncompact polyhedra.

All of Chapman's results quoted here can also be found in his book [45].

4.6. *West's theorem*

Borsuk [32, Problem 9.1] asked whether every compact ANR has the homotopy type of a compact polyhedron. For simply connected spaces, this question was answered by De Lyra [122] in the affirmative. For nonsimply-connected spaces Borsuk's problem stayed a mystery for a long time.

The problem was laid to rest by West [186] who showed, using among other things the technique in Miller [133], that for every compact ANR X there are a compact Q-manifold M and a cell-like map from M onto X.

A *cell-like map* between compacta is one for which point-inverses have the shape of a point; a cell-like map between ANR's is a *fine homotopy equivalence* as proved by Haver [85] and Toruńczyk [177]. As we have seen above, the Q-manifold M is homeomorphic to $P \times Q$ for some compact polyhedron P and so X has the same homotopy type as P.

4.7. *Edwards' theorem*

In 1974, Edwards [45, Chapter 14] improved West's result by showing that $X \times Q$ is a Q-manifold if and only if X is a locally compact ANR. This provides an elegant proof of West's Theorem: for a compact ANR X there is by Chapman's result a compact polyhedron P such that $X \times Q$ and $P \times Q$ are homeomorphic; clearly then X and P have the same homotopy type.

In the proof of Edwards' result and in Toruńczyk's work, which we shall describe momentarily, a crucial role was played by shrinkable maps. A continuous surjection $f : X \to Y$ between compact spaces is said to be *shrinkable* if one can find for every $\varepsilon > 0$ a homeomorphism φ of X onto itself such that $d(f \circ \varphi, f) < \varepsilon$, and $\mathrm{diam}(\varphi(f^{-1}(y))) < \varepsilon$ for all $y \in Y$. So a shrinkable map f is map whose fibers can be uniformly shrunk to small sets by a homeomorphism that looking from Y does not change f too much.

Bing's shrinking criterion from [30] characterizes shrinkable maps as uniform limits of homeomorphisms (so-called *near homeomorphisms*). Thus, in order to prove two compact spaces homeomorphic it suffices to produce a shrinkable map between them.

As an example consider cone(Q). We observed above that from Keller's theorem it follows that cone(Q) $\approx Q$. But this follows also trivially from Bing's shrinking criterion. Since one-point compactifications are unique, it follows that we can also think of cone(Q)

as the space obtained from $Q \times [0,1]$ by identifying the set $Q \times \{1\}$ to a single point. The decomposition map is easily seen to be shrinkable, hence a near homeomorphism, and so cone$(Q) \approx Q \times [0,1] \approx Q$, as desired (for details, see [131, Theorem 6.1.11]).

It is easy to see that if X and Y are compact ANR's and $f : X \to Y$ is a near homeomorphism then f is cell-like. So the method of shrinkable maps only works for cell-like maps.

4.8. *Toruńczyk's theorems (part 1)*

In 1980, Toruńczyk [178] published a remarkable result. He was able to topologically characterize the Q-manifolds among the locally compact ANR's. From Edwards's theorem it was already known that if X is a locally compact ANR then $X \times Q$ is a Q-manifold. Toruńczyk studied the question when the projection

$$\pi : X \times Q \to X$$

is shrinkable, and came to an astounding conclusion. This map is shrinkable if and only if X has the following property: given $n \in \mathbb{N}$ and two maps $f, g : \mathbb{I}^n \to X$ and $\varepsilon > 0$ there exist maps $\xi, \eta : \mathbb{I}^n \to X$ such that

$$\xi[\mathbb{I}^n] \cap \eta[\mathbb{I}^n] = \emptyset$$

while moreover

$$d(f, \xi) < \varepsilon \quad \text{and} \quad d(\eta, g) < \varepsilon.$$

For obvious reasons this property is called the *disjoint cells-property*. So one arrives at the following conclusion, which is called Toruńczyk's theorem:

THEOREM 4.8.1. *Let X be a locally compact* ANR. *Then X if a Q-manifold if and only if X satisfies the disjoint-cells property.*

As in the case of Edwards' theorem, the Bing shrinking criterion and the Z-set Unknotting Theorem 4.3.2 were crucial in the proof of this result.

Toruńczyk's remarkable theorem ended the game of identifying Hilbert cubes. For in order to prove that a given space X is homeomorphic to Q, all one needs to prove is that it is an AR and satisfies the disjoint cells-property. Observe that both properties are trivially *necessary* for a space to be homeomorphic to Q. It is fascinating that these two properties that are stated in simple topological terms are also *sufficient*. In order to demonstrate the power of his topological characterization of Q, Toruńczyk [178] presented a very short and elegant proof of the Curtis–Schori–West hyperspace theorem.

4.9. *The Taylor example*

The above results emphasized the close relationships between infinite-dimensional topology and AR and ANR-theory. As we said above, certain delicate finite-dimensional obstructions turned out not to appear in Q-manifold theory. However, certain delicate *infinite-dimensional* obstructions *do* appear in infinite-dimensional topology. The first result demonstrating this was the result of Taylor [171] which we shall describe briefly.

It is an example of a cell-like map from a compactum X to Q which is not a shape equivalence. The space X is the inverse limit of a sequence of compact polyhedra with special properties. That the desired polyhedra exist follows from work of Adams [2] and Toda [173]. Adams' proof uses complex K-theory.

The Taylor example was widely used in infinite-dimensional topology, shape theory and ANR-theory to obtain all sorts of counterexamples. Daverman and Walsh [55] used it to get an example of a cell-like map $f : X \to Y$ between compacta whose non-degeneracy set is contained in a strongly countably dimensional set and which is not a shape equivalence. They also obtained from the Taylor example new examples of locally contractible continua which are not ANR's. It was also used in 1979 to answer Borsuk's problem [32, Problem V.12.16] in the negative for the construction of an upper semi-continuous decomposition of Q into copies of itself, whose decomposition space is not an ANR.[36] And it was used to give a negative answer to Kuratowski's question [117] from 1951 whether a space with the compact extension property is necessarily an AR; a space X is said to have the *compact extension property* if for every space Y and every *compact* subset A of Y every continuous map from A to X has a continuous extension over X.[37] For the use of the Taylor example in shape theory, see Mardešić and Segal [125].

4.10. *Dranišnikov's example*

Through the work of Edwards and Walsh [184] it was known in 1981 that the following two fundamental problems in dimension theory are equivalent:

1. Does there exist an infinite-dimensional compactum with finite cohomological dimension? (This problem is due to Alexandrov [5].)
2. Does there exist a cell-like map $f : X \to Y$, where X is a finite-dimensional compactum but Y is infinite-dimensional? (This problem, known as the *cell-like dimension raising mapping problem*, grew out of manifold theory and the work of Kozlowski [111]. The first attempt to solve it by proving that every infinite-dimensional compactum contains sets of arbitrarily large finite dimension was shown to lead nowhere by Walsh [183].)

The problem was solved by Dranišnikov [63] in 1988: there exists an infinite-dimensional compactum with cohomological dimension 3. Essential in his construction is that there is a generalized cohomology theory for which the Eilenberg–MacLane complex $K(\mathbb{Z}, 3)$ behaves like a point. For 2 dimensions, such an approach does not work. But there does exist an infinite-dimensional compactum with cohomological dimension 2, as was shown by Dydak and Walsh [66]. Their work is based on the validity of the Sullivan conjecture.

As in the case of the Taylor example, the Dranišnikov example was also used by various authors to obtain counterexamples to a variety of questions. We will mention only one such application of special interest in infinite-dimensional topology. To put this result into

[36] The construction can be found in Topology and its Applications **12** (1981), 315–320.

[37] The construction can be found in Proc. Amer. Math. Soc. **97** (1986), 136–138.

perspective, we will first make some remarks. It was known since 1951 from the work of Dugundji [64] that every locally convex linear space is an AR. Whether the local convexity assumption could be dropped was a fascinating question in infinite-dimensional topology and ANR-theory for a very long time. It was settled in the negative by Cauty [41] in 1994 by a very interesting method and an essential use of the Dranišnikov example. More specifically, he proved that there exists a necessarily nonlocally convex linear space L which is not an AR but which is a closed linear subspace of a linear space which is an AR.

4.11. *Toruńczyk's theorems (part 2)*

So far, we mainly concentrated on (locally) compact spaces. As is to be expected, there are also results for complete spaces. Recall that infinite-dimensional topology started with the investigation of completely metrizable linear spaces. In [179] Toruńczyk characterized the topology of Hilbert spaces in much the same way as he characterized the topology of the Hilbert cube. In this characterization the disjoint-cells property is replaced by the discrete approximation property; this property states that for every open cover $\mathcal{U}$ of the space X and every map f from the topological sum $\bigoplus_{n=1}^{\infty} \mathbb{I}^n$ to X there is another map g from $\bigoplus_{n=1}^{\infty} \mathbb{I}^n$ to X that is $\mathcal{U}$ close to f and is such that the family $\{g[\mathbb{I}^n]\colon n \in \mathbb{N}\}$ is discrete. The characterization reads:

THEOREM 4.11.1. *A separable space is a manifold modeled on ℓ^2 if and only if it is a completely metrizable separable ANR with the discrete approximation property.*

As a consequence ℓ^2 is characterized as the only separable completely metrizable AR with the discrete approximation property.

Toruńczyk has also a similar characterization of manifolds modeled on arbitrary Hilbert spaces, see [179] for more details.

4.12. *Epilogue*

We saw that Anderson, interested in questions in general topology, created a new field in topology called *infinite-dimensional topology* and was at the beginning unaware of its potential. But good mathematics inevitably led to good results in various other disciplines, mostly in algebraic and geometric topology. The highlights of infinite-dimensional topology are the theorems of Anderson on the homeomorphy of ℓ^2 and s, of Chapman on the invariance of Whitehead torsion, of West on the finiteness of homotopy types of compact ANR's and of Toruńczyk on the topological characterization of manifolds modeled on various infinite-dimensional spaces.

A large collection of open problems is West's paper [187]. The subjects that are being touched upon range from absorbing sets and function spaces to ANR theory. We mention two particularly prominent problems:

1. Let $\alpha : Q \to Q$ be an involution with a unique fixed point. Is α conjugated to the standard involution β on Q defined by $\beta(x) = -x$?
2. For $n \geqslant 3$, let $\mathcal{H}_n$ be the group of all homeomorphisms on $\mathbb{I}^n$ endowed with the compact-open topology. Is $\mathcal{H}_n$ homeomorphic to ℓ^2?

5. The period of harvesting: set theoretic topology

In the sixties, general topology renewed its interaction with set theory. In 1878, Cantor's work [39] had created set theory and topology as we saw in Section 2.2. They developed as diverse, complex and independent fields. Soon after their renewed interaction spectacular results surfaced, also in parts of topology where traditionally geometric and algebraic tools were used, or tools from analysis. It is about those results that we wish to report here.

We saw that in [39] Cantor wrote down the *Continuum Hypothesis* (abbreviated CH) that would have a profound effect on set theory in the 20-th century. The CH states that the first uncountable cardinal is $\mathfrak{c}$, the cardinality of the real line (the continuum). The work of Gödel [81] and Cohen [49] has shown that CH is consistent with and independent from the "usual" Zermelo–Fraenkel axioms of set theory. The methods used in these proofs, and especially Cohen's forcing, had a profound effect on the development of a new field in topology called *set theoretic topology*. In that development, the work of Mary Ellen Rudin was vital.

In our report below we will almost exclusively concentrate on independence results in topology, that is, results that are independent from and consistent with the "usual" Zermelo–Fraenkel axioms of set theory. So we will ignore important parts of general topology. Also some problems are being discussed whose solution is very strongly of a set theoretic nature without being an independence result. None of the results mentioned has its roots in general topology.

5.1. *Souslin's problem*

Suppose that S is a connected, linearly ordered topological space without a first or last element. If S is separable then S is isomorphic to $\mathbb{R}$. What happens if one relaxes the separability condition to the condition that any pairwise disjoint collection of nontrivial intervals of X is countable? This is Souslin's problem from [166]. It was posed in 1920 and has fascinated topologists and set theorists ever since.

The requirement that pairwise disjoint collections of intervals (or more general open sets) are countable is called the *countable chain condition* (abbreviated ccc).

A counterexample to Souslin's question, a ccc connected linearly ordered space without first or last element that is not homeomorphic to the real line, is called a *Souslin line*, and *Souslin's Hypothesis* (SH) is the statement that no Souslin lines exist. Jech [96] and Tennenbaum [172] used Cohen's forcing method to show that Souslin lines can exist and Jensen [97] proved that they also exist in Gödel's Constructible Universe, the same universe Gödel used to establish the consistency of the Continuum Hypothesis. In [165], Solovay and Tennenbaum developed the forcing method further and proved the consistency of Souslin's Hypothesis. Their proof established the consistency of a powerful combinatorial principle, which we shall discuss briefly.

The principle, called *Martin's Axiom* (MA), states that no compact Hausdorff space that satisfies the ccc is the union of fewer than $\mathfrak{c}$ nowhere dense sets. Under CH "fewer than $\mathfrak{c}$" means countable and so MA holds by the Baire Category Theorem. However, MA is also consistent with the negation of CH and it is this combination, MA + $\neg$CH, that proved to be very powerful indeed.

Solovay and Tennenbaum [165] showed that under MA + ¬CH there are no Souslin lines, thereby proving that SH is undecidable. Ever since this result, Martin's axiom played a prominent role in set theory and set theoretic topology, as the rest of our story will tell.

5.2. *Alexandroff's problem*

Most mathematicians in geometric topology are only interested in metrizable spaces, and metrizable manifolds in particular. But there are also mathematically important objects that are not always metrizable, for example, CW-complexes, linear spaces, topological groups and manifolds. By a *manifold* we mean a locally Euclidean Hausdorff space. Manifolds are certainly mathematically important, with or without differential or algebraic structure.

Let M be a manifold. If $A \subseteq M$ is closed then one certainly wants to be able to extend every continuous real valued function $f : A \to \mathbb{R}$ to a continuous function $\bar{f} : M \to \mathbb{R}$. By the Tietze–Urysohn theorem, this is equivalent to M being normal. In the process of constructing new continuous functions from old ones (think of homotopies) it is also extremely pleasant if M has the following property: for every closed subset $A \subseteq M$ there is a sequence $\langle U_n \rangle_n$ of open subsets of M such that $A = \bigcap_{n<\omega} U_n$. General topologists say that spaces with this property are *perfect*. If one wants to generalize some of the existing theory on metrizable manifolds to nonmetrizable ones, it becomes clear quite quickly that in many instances it is inevitable to restrict oneself to manifolds that are both normal and perfect, i.e. manifolds that are *perfectly normal*. The question then naturally arises whether there is a perfectly normal manifold which is not metrizable. This question was asked by Alexandroff [4] in 1935 and also by Wilder [189] in 1949.

It seems very unlikely that a set theoretic statement like CH has anything to do with manifolds, let alone with Alexandroff's problem. In [154] however, Rudin and Zenor constructed assuming CH an example of a perfectly normal nonmetrizable manifold. Later, Kozlowski and Zenor [112] even constructed such a manifold that is analytic, again under CH. These provisional solutions to Alexandroff's problem very strongly suggested a positive answer to it.

In [153], Rudin proved that under MA + ¬CH, all perfectly normal manifolds are metrizable; as a consequence, she came to the remarkable conclusion that Alexandroff's problem is undecidable.

5.3. *Dowker's problem*

In [31], Borsuk proved his famous homotopy extension theorem for metrizable spaces. Actually, his result is true for spaces X for which the product $X \times \mathbb{I}$ is normal. This generalisation is due to Dowker and was first published in Hurewicz and Wallman [94]. A necessary condition for $X \times \mathbb{I}$ to be normal is that X is normal. So it is natural to ask whether this condition is also sufficient. This is Dowker's problem. Dowker [62] and Katětov [102] independently gave necessary and sufficient conditions for a space X to have the property that its product with $\mathbb{I}$ is normal.

THEOREM 5.3.1. *Let X be a space. Then $X \times \mathbb{I}$ is normal if and only X is normal and for every decreasing sequence of closed subsets $\langle D_n \rangle_n$ of X with $\bigcap_{n<\omega} D_n = \emptyset$ there exists* *a sequence $\langle U_n \rangle_n$ of open subsets of X such that $D_n \subseteq U_n$ for every n while moreover $\bigcap_{n<\omega} U_n = \emptyset$.*

A normal space X for which $X \times \mathbb{I}$ is not normal is called a *Dowker space* in the literature. So if one wishes to construct a Dowker space, all one needs to do is to construct a normal space X having a sequence of closed subsets $\langle D_n \rangle_n$ of X with $\bigcap_{n<\omega} D_n = \emptyset$ such that if $U_n \subseteq X$ is open and $D_n \subseteq U_n$ for every n then $\bigcap_{n<\omega} U_n \neq \emptyset$. It is surprising that this condition is such a complicated one.

In 1955, Rudin [149] constructed the first example of a Dowker space assuming the existence of a Souslin line. That was a major breakthrough at that time, but as turned out later, had an unpleasant drawback since, as we saw above, SH is undecidable.

But in 1971 it was shown that the solution to Dowker's problem does not depend on set theory: the first example of a *real* (= using no axioms beyond ZFC) example of a Dowker space was constructed again by Rudin [151]. This Dowker space was the only ZFC example of such a space for about twenty years. Balogh [17] constructed another such example only in 1994 (see also his subsequent paper [18]). This very interesting example is "small" while the original Dowker space is "large". It is certainly not the final word on Dowker spaces since it is still unknown whether there can be a first countable Dowker space in ZFC, or one of cardinality ω_1. Using pcf theory, Kojman and Shelah [109] constructed a Dowker subspace of Rudin's example in [151] of size $\aleph_{\omega+1}$. This is a "real" example of a small Dowker space since its cardinality is decided in ZFC, while Rudin's and Balogh's are not.

Ironically, Borsuk's theorem that started all this research, turned out to hold also without the assumption of normality of the product with $\mathbb{I}$, see Morita [137] and Starbird [167].

5.4. *Whitehead's problem*

Whitehead asked whether every compact arcwise-connected Abelian topological group is isomorphic to a product of circles. This is a very natural problem for a topologist. We first translate it into purely algebraic language to turn it into a very natural problem for an algebraist as well. If A and B are Abelian groups then a surjective homomorphism $f : A \to B$ is said to *split* if there is a homomorphism $g : B \to A$ with $f \circ g$ is equal to the identity on B. An Abelian group G is *Whitehead* if for every Abelian group B, every surjective homomorphism $f : B \to G$ with kernel isomorphic to $\mathbb{Z}$ splits. It is clear that all free groups are Whitehead and Whitehead asked whether all Whitehead groups are free. It is a consequence of Pontrjagin duality that both problems we attributed here to Whitehead are equivalent.

Shelah [158, 159] showed that Whitehead's problem in undecidable by showing that under $V = L$ all Whitehead groups are free while under MA + ¬CH there exists a Whitehead group which is not free.

The fact that Whitehead's problem can be formulated both into algebraic and topological language is not an exception for a problem that turns out to be dependent upon one's set theory. These problems can often be translated into several mathematical languages and can therefore be attacked from several directions. There are for example numerous problems in Boolean algebras that can be translated into topology and vice versa. Sometimes such a translation helps.

It is questionable of course whether Whitehead's problem discussed above is a "real" topological problem. We took the liberty of mentioning it because it is such a good example of our point that problems of set theoretic nature can often be attacked from different angles.

5.5. *Choquet's problem*

A BA (= Boolean Algebra) will be identified with its universe. A BA B is called

complete / countably complete / weakly countably complete

if for any two subsets P and Q such that $p \wedge q = 0$ for $r \in P$ and $q \in Q$

without further condition / with $|P| = \omega$ or $|Q| = \omega$ / with $|P| = |Q| = \omega$

there is an $s \in B$ which *separates* P and Q, i.e. $p \leqslant s$ fpr $p \in P$ and $q \leqslant s'$ for $q \in Q$. Consider the following statements:

- FB every weakly countably complete BA is a homomorphic image of a countably complete BA;
- BE every countably complete BA is a homomorphic image of a complete BA;
- FE every weakly countably complete BA is a homomorphic image of a complete BA.

The earliest statement we are aware of where one of these statements is considered is Louveau [121]. Here he attributes the question (or conjecture) of whether FE holds to Choquet, and proves that under CH the restriction of FE to algebras of size $\leqslant \mathfrak{c}$ holds. The question of whether BE holds was raised by Koppelberg [110], who was apparently unaware of Louveau's paper. She proved that the restriction of BE to algebras of cardinality $\leqslant \mathfrak{c}$ holds under CH. The question of whether FB holds was raised by van Douwen, Monk and Rubin [59], who also repeated the question of whether BE holds.

By Stone duality, all these question can be formulated in topological language. They were all solved by topologists. It was shown that FB is not a theorem in ZFC under MA + $\mathfrak{c} = \omega_2$, hence, neither is FE.[38] The problem of whether BE holds turned out to be difficult. It was finally shown in Dow and Vermeer [61] that BE is not a theorem of ZFC. The algebra in question is B, the algebra of Borel sets of the unit interval. They showed that if B is the quotient of some complete Boolean algebra then there is a lifting of the quotient of B modulo the meager sets back into B. An appeal to a result of Shelah [162] that such a lifting need not exist finishes the proof.

5.6. *Binary operations on $\beta\omega$*

Let $\beta\omega$ denote the the Čech–Stone compactification of the discrete space ω. As is well known, the points of this space can be thought of as ultrafilters in $\mathcal{P}(\omega)$. Thinking about the points in $\beta\omega$ in this way, it is easy to extend various binary operations on ω to binary operations on $\beta\omega$. As an example, let us consider ordinary addition on ω.

For $A \subseteq \omega$ and $n \in \omega$ we set

$$A - n = \{k \in \omega\colon\ k + n \in A\}.$$

[38] The construction can be found in Trans. Amer. Math. Soc. **259** (1980), 121–127.

For $p, q \in \beta\omega$ put

$$p + q = \{A \subseteq \omega\colon \{n \in \omega\colon A - n \in p\} \in q\}.$$

Then $+$ is a well-defined binary operation on $\beta\omega$ which extends the ordinary addition on ω and moreover is associative and right-continuous (this is due to Glazer, see [50]). So $(\beta\omega, +)$ is a compact right topological semigroup. By a result of Wallace [181, 182] (see also Ellis [70]), the compactness of $\beta\omega$ implies the existence of a point $p \in \beta\omega$ for which $p + p = p$, i.e. a so-called *idempotent.*

Glazer (see [50]) used the existence of idempotents in the semigroup $(\beta\omega, +)$ to give a particularly simple *topological* proof of Hindman's theorem from [90]: *If the natural numbers are divided into two sets then there is a sequence drawn from one of these sets such that all finite sums of distinct numbers of this sequence remain in the same set.*

This statement was known for some years as the Graham–Rothschild Conjecture.

Several other results from classical number theory can be proved as well by similar methods. In [21] Bergelson, Furstenberg, Hindman and Katznelson again used the semigroup $(\beta\omega, +)$ to present an elementary proof of van der Waerden's theorem from [180]: *if the natural numbers are partitioned into finitely many classes in any way whatever, one of these classes contains arbitrarily long arithmetic progressions.*

5.7. *Strong homology*

Let $Y^{(k+1)}$ be the topological sum of countably many copies of the $(k + 1)$-dimensional Hawaiian earring. The calculation of the strong homology of $Y^{(k+1)}$ is of interest in the question of whether strong homology satisfies the additivity axiom (of Milnor [134]). In [124], Mardešić and Prasolov translated the calculation of the (k-dimensional) strong homology of $Y^{(k+1)}$ into a condition of set theory. They proved that this condition holds under CH, and hence that the (k-dimensional) strong homology of $Y^{(k+1)}$ can be nontrivial. But, as was shown in Dow, Simon and Vaughan [60], there are also models of set theory in which it does not hold, and therefore in such models the (k-dimensional) strong homology of $Y^{(k+1)}$ is trivial.

5.8. *Banach spaces*

In Banach space theory, many results from general topology are applied. The completeness of the real line gives the Hahn–Banach Theorem, Baire's Category Theorem is essential in the proof of the open mapping theorem and the uniform boundedness principle, while Tychonoff's compactness theorem proves the Alaoglou theorem, etc. It is therefore not surprising that set theoretic topology turned out to have very interesting applications in Banach space theory. It is about two of those results that we wish to report here.

For a compact space K we let $M(K)$ denote the space of all finite real-valued regular Borel (or, Baire) measures on K (with $\|\mu\| = |\mu|(K)$, where $|\mu|$ is the total variation of μ).

Pełczyński [144] proved the following result:

THEOREM 5.8.1. *Let α be an infinite cardinal number, X a Banach space, and $\ell^1_\alpha \hookrightarrow X$ an isometric imbedding. Then the space $M(\{0,1\}^\alpha)$ admits an isometric imbedding in the dual X^* of X. In particular,*

$$L^1(\{0,1\}^\alpha) \hookrightarrow X^* \quad \textit{and} \quad \ell^1_{2^\alpha} \hookrightarrow X^*.$$

The question naturally arises whether the converse to this theorem holds, i.e. whether from $M(\{0,1\}^\alpha) \hookrightarrow X^*$ it follows that $\ell^1_\alpha \hookrightarrow X$. Pełczyński conjectured that this is true, and verified the conjecture for $\alpha = \omega$, [144]. The answer to Pełczyński's Conjecture is fascinating. For cardinals $\alpha > \omega_1$ it is true, as was shown by Agryros [3]. So there only remains the cardinal ω_1. For that cardinal number the question is undecidable. Under $\mathrm{MA} + \neg\mathrm{CH}$, Pełczyński's Conjecture is true for $\alpha = \omega_1$ as was also shown by Agryros [3]. But under CH, Haydon [86] constructed a counterexample of a particular nice form since it is of the form $C(K)$ for a certain compact Hausdorff space K. The space K is an inverse limit of an ω_1-sequence of Cantor sets with certain specific properties. Independently, a similar space was also constructed by Kunen [115] motivated by topological questions. In addition, it also surfaced in the work of Talagrand [169]. So Pełczyński's Conjecture turned out to boil down partly to the construction under CH of a very peculiar compact Hausdorff space K. It is precisely in such constructions where set theoretic topology plays such a prominent role and where its techniques are fundamental.

Another application of set theoretic topology to Banach space theory is the following one. If X is a Banach space and $A \subseteq X$ then $\overline{\mathrm{convex}}(A)$ denotes the closed convex hull of A. If X is separable, then for every uncountable subset $A \subseteq X$ there exists an element $a \in A$ such that $a \in \overline{A \setminus \{a\}}$, in particular, $a \in \overline{\mathrm{convex}}(A \setminus \{a\})$. Davis and Johnson asked whether the latter property could hold in a nonseparable Banach space. It was solved in the affirmative by Shelah [160] under the combinatorial principle $\diamondsuit$. But this example is not of the form $C(K)$ for some compact Hausdorff space K. But such a space exists even under the weaker hypothesis CH, as was shown by Kunen [113].

5.9. *Epilogue*

Our overview of set theoretic topology is very much less than complete as a description of what happened in that area (see our remarks at the beginning of the introduction). We have for example not mentioned several very important areas in set theoretic topology such as cardinal functions, S- and L-spaces, the Normal Moore Space Conjecture, βX (including $\beta\omega$), the set theoretic aspects of topological groups, etc. In addition, we could have talked much more about its relation with set theory and we said very little about Boolean Algebras.

For more information on set theoretic topology we refer the reader to Rudin's book [152], the Handbook of Set Theoretic Topology [116] and the book on Open Problems in Topology [132].

Notes

In this section we will give some additional information on the material presented in Sections 4 and 5 that we find useful. No attempt has been made to be complete.

In Section 3 we already mentioned the following books for among other things information on some the results obtained in the golden age of general topology: Hu [92], Hurewicz and Wallman [94], Gillman and Jerison [80], Dugundji [65], Nagata [142, 141], Engelking [71, 72] and Arhangel'skiĭ and Ponomarev [13]. To this list we can add the following books for additional information and later developments, some of which we already mentioned in the other sections: Arhangel'skiĭ [12], Bessaga and Pełczyński [27], Borsuk [32, 33], Comfort and Negrepontis [51], Balcar and Štěpánek [16], Devlin and Johnsbråten [56], Bourbaki [34], Hu [93], Aarts and Nishiura [1], Juhász [100], Kechris [103], Kunen [114], Kuratowski [118, 119], Kuratowski and Mostowski [120], Mardešić and Segal [125], Nadler [139], Shelah [161], Todorčević [175], Daverman [54], Aull and Lowen [15], Chapman [45], Rudin [152], the Handbook of Set Theoretic Topology [116], the book on Topics in General Topology [138], the book on Open Problems in Topology [132], the book on Recent progress in General Topology [95] and Chigogidze [46].

Notes on Section 4. For a different proof that all infinite-dimensional separable Banach spaces are homeomorphic, see Bessaga and Pełczyński [24].

For different proofs of Anderson's theorem that $s \approx \ell^2$, see Anderson and Bing [10] and [131, Chapter 6].

As we observed, Keller proved that the Hilbert cube is homogeneous. This result was later generalized by Fort [73] who proved that the infinite product of compact manifolds is homogeneous if and only if none or infinitely many of the factors have a boundary.

For a proof that any homeomorphism φ between Cantor sets in $\mathbb{R}^2$ can be extended to a homeomorphism of $\mathbb{R}^2$, see Kuratowski [118, 119]. For the cited result about Antoine's necklace, the reader can consult, e.g., Daverman [54, Corollary 5A].

Let X be a space, and let $A \subseteq X$ be closed. Nowadays we call A a Z-set in X if for every $\varepsilon > 0$ and every continuous function $f : Q \to X$ there exists a continuous function $g : Q \to X \setminus A$ such that $d(f, g) < \varepsilon$. This definition is easier to work with than the original one and is equivalent to it in the special case $X = Q$ (but this is not entirely trivial). For detailed proofs of the Z-set homeomorphism extension results in Q-manifolds, see Bessaga and Pełczyński [27], Chapman [45] and van Mill [131, Chapter 6].

See Cohen [48] for more information on the concept of a simple homotopy equivalence.

For a detailed description of the Dranišnikov example, see also Chigogidze [46].

As we remarked, the remaining open problems in infinite-dimensional topology (see West [187]) deal among other things with problems in absorbing sets (see, e.g., [28, 58]), function spaces (see, e.g., [42]) and ANR-theory.

For general information on hyperspaces see Nadler [139].

Notes on Section 5. For a simple proof that metrizable spaces are paracompact, see Rudin [150].

Many of the set theoretic things that we merely touched upon in this section can be found in great detail in Kunen [114].

For more information on SH and many related topics, see Todorčević [174].

For more information on Whitehead's problem, see Ecklof [67].

For more information on the role of topology in Banach spaces and measure theory, see Negrepontis [143], Mercourakis and Negrepontis [127] and Fremlin [79].

For more information on ultrafilters and combinatorial number theory, see Hindman [91].

A recent book on general/set theoretic topology is Todorčević [176].

Acknowledgements

We are indebted to Henno Brandsma, Ryszard Engelking, Klaas Pieter Hart, Michael van Hartskamp and Gerke Nieuwland for a critical reading of (parts of) the manuscript and for many helpful comments.

We are grateful to Walter Purkert for supplying us with some useful information on Hausdorff.

Bibliography

[1] J.M. Aarts and T. Nishiura, *Dimension and Extensions*, North-Holland, Amsterdam (1993).
[2] J.F. Adams, *On the groups $J(X)$, IV*, Topology **5** (1966), 21–71.
[3] S. Agryros, *On non-separable Banach spaces*, Trans. Amer. Math. Soc. **270** (1982), 193–216.
[4] P. Alexandroff, *On local properties of closed sets*, Ann. of Math. **36** (1935), 1–35.
[5] P.S. Alexandrov, *Einige Problemstellungen in der mengentheoretischen Topologie*, Mat. Sb. **1** (1936), 619–634.
[6] R.D. Anderson, *The Hilbert cube as a product of dendrons*, Notices Amer. Math. Soc. **11** (1964), 572.
[7] R.D. Anderson, *Hilbert space is homeomorphic to the countable infinite product of lines*, Bull. Amer. Math. Soc. **72** (1966), 515–519.
[8] R.D. Anderson, *On topological infinite deficiency*, Michigan Math. J. **14** (1967), 365–383.
[9] R.D. Anderson, *Topological properties of the Hilbert cube and the infinite product of open intervals*, Trans. Amer. Math. Soc. **126** (1967), 200–216.
[10] R.D. Anderson and R.H. Bing, *A complete elementary proof that Hilbert space is homeomorphic to the countable infinite product of lines*, Bull. Amer. Math. Soc. **74** (1968), 771–792.
[11] R.D. Anderson and T.A. Chapman, *Extending homeomorphisms to Hilbert cube manifolds*, Pacific J. Math. **38** (1971), 281–293.
[12] A.V. Arhangel'skiĭ, *Topological Function Spaces*, Kluwer Academic, Dordrecht/Boston/London (1992).
[13] A.V. Arhangel'skiĭ and V.I. Ponomarev, *Fundamentals of General Topology*, Reidel, Dordrecht/Boston/Lancaster (1984).
[14] C. Arzelà, *Funzioni di linee, Nota del prof. Cesare Arzelà, presentata dal Corrispondente V. Volterra*, Atti della Reale Accademia dei Lincei, Serie quarta **5** (1889), 342–348.
[15] C.E. Aull and R. Lowen (eds), *Handbook of the History of General Topology, Vol. 1*, Kluwer Academic, Dordrecht/Boston/London (1997).
[16] B. Balcar and P. Štěpánek, *Teorie Množin*, Academia, Praha (1986).
[17] Z.T. Balogh, *A small Dowker space in ZFC*, Proc. Amer. Math. Soc. **124** (1996), 2555–2560.
[18] Z.T. Balogh, *A normal screenable nonparacompact space in ZFC* (1997), to appear.
[19] S. Banach, *Théorie des Opérations Linéaires*, PWN, Warszawa (1932).
[20] B. Barit, *Small extensions of small homeomorphisms*, Notices Amer. Math. Soc. **16** (1969), 295.
[21] V. Bergelson, H. Furstenberg, N. Hindman and Y. Katznelson, *An algebraic proof of van der Waerden's Theorem*, Enseign. Math. **35** (1989), 209–215.
[22] C. Bessaga, *Every infinite-dimensional Hilbert space is diffeomorphic with its unit sphere*, Bull. Polon. Acad. Sci. Sér. Math. Astronom. Phys. **14** (1966), 27–31.
[23] C. Bessaga, *Negligible sets in linear topological spaces*, Bull. Polon. Acad. Sci. Sér. Math. Astronom. Phys. **15** (1968), 397–399.
[24] C. Bessaga and A. Pełczyński, *A topological proof that every separable Banach space is homeomorphic to a countable product of lines*, Bull. Polon. Acad. Sci. Sér. Math. Astronom. Phys. **8** (1960), 487–493.
[25] C. Bessaga and A. Pełczyński, *Some remarks on homeomorphisms of Banach spaces*, Bull. Polon. Acad. Sci. Sér. Math. Astronom. Phys. **8** (1960), 757–760.
[26] C. Bessaga and A. Pełczyński, *Some remarks on homeomorphisms of F-spaces*, Bull. Polon. Acad. Sci. Sér. Math. Astronom. Phys. **10** (1962), 265–270.

[27] C. Bessaga and A. Pełczyński, *Selected Topics in Infinite-Dimensional Topology*, PWN, Warszawa (1975).
[28] M. Bestvina and J. Mogilski, *Characterizing certain incomplete infinite-dimensional absolute retracts*, Michigan J. Math. **33** (1986), 291–313.
[29] R.H. Bing, *Metrization of topological spaces*, Canad. J. Math. **3** (1951), 175–186.
[30] R.H. Bing, *A homeomorphism between the 3-sphere and the sum of two solid horned spheres*, Ann. of Math. **56** (1952), 354–362.
[31] K. Borsuk, *Sur les prolongements des transformations continus*, Fund. Math. **28** (1936), 99–110.
[32] K. Borsuk, *Theory of Retracts*, PWN, Warszawa (1967).
[33] K. Borsuk, *Theory of Shape*, PWN, Warszawa (1975).
[34] N. Bourbaki, *Topologie Générale*, 2^{e} ed., Hermann, Paris (1958).
[35] L.E.J. Brouwer, *Zur Analysis Situs*, Math. Ann. **68** (1910), 422–434; Reprinted in H. Freudenthal (ed.), *L.E.J. Brouwer Collected Works, Vol. 2, Geometry, Analysis, Topology and Mechanics*, North-Holland, Amsterdam (1976).
[36] L.E.J. Brouwer, *Beweis der Invarianz der Dimensionenzahl*, Math. Ann. **70** (1911), 161–165; Reprinted in H. Freudenthal (ed.), *L.E.J. Brouwer Collected Works, Vol. 2, Geometry, Analysis, Topology and Mechanics*, North-Holland, Amsterdam (1976).
[37] L.E.J. Brouwer, *Über Abbildung von Mannigfaltigkeiten*, Math. Ann. **71** (1912), 97–115.
[38] G. Cantor, *Über die Ausdehnung eines Satzes aus der Theorie der trigonometrischen Reihen*, Math. Ann. **5** (1872), 123–132; Reprinted in *Georg Cantor, Gesammelte Abhandlungen mathematischen und philosophischen Inhalts*, Reprint of the original 1932 edition, Springer, Berlin (1980).
[39] G. Cantor, *Ein Beitrag zur Mannigfaltigkeitslehre*, Crelles J. Math. **84** (1878), 242–258.
[40] G. Cantor, *Gesammelte Abhandlungen mathematischen und philosophischen Inhalts*, Springer, Berlin (1980), Reprint of the original 1932 edition.
[41] R. Cauty, *Un espace métrique linéaire qui n'est pas un rétracte absolu*, Fund. Math. **146** (1994), 11–22.
[42] R. Cauty, T. Dobrowolski and W. Marciszweski, *A contribution to the topological classification of the spaces* $C_p(X)$, Fund. Math. **142** (1993), 269–301.
[43] T.A. Chapman, *All Hilbert cube manifolds are triangulable*, unpublished manuscript.
[44] T.A. Chapman, *Topological invariance of Whitehead torsion*, Amer. J. Math. **96** (1974), 488–497.
[45] T.A. Chapman, *Lectures on Hilbert Cube Manifolds*, CBMS 28, Amer. Math. Soc., Providence, RI, USA (1975).
[46] A. Chigogidze, *Inverse Spectra*, North-Holland, Amsterdam (1996).
[47] E.W. Chittenden, *On the equivalence of écart and voisinage*, Trans. Amer. Math. Soc. **18** (1917), 161–166.
[48] M. Cohen, *A Course in Simple-Homotopy Theory*, Springer, Berlin (1970).
[49] P. Cohen, *The independence of the Continuum Hypothesis I, II*, Proc. Nat. Acad. Sci. USA **50, 51** (1964), 1143–1148 and 105–110.
[50] W.W. Comfort, *Ultrafilters: some old and some new results*, Bull. Amer. Math. Soc. **83** (1977), 417–455.
[51] W.W. Comfort and S. Negrepontis, *The Theory of Ultrafilters*, Grundlehren der Mathematischen Wissenschaften, Vol. 211, Springer, Berlin (1974).
[52] D.W. Curtis and R.M. Schori, *Hyperspaces of Peano continua are Hilbert cubes*, Fund. Math. **101** (1978), 19–38.
[53] J.W. Dauben, *George Cantor, His Mathematics and Philosophy of the Infinite*, Harvard Univ. Press, Cambridge, MA (1979).
[54] R.J. Daverman, *Decompositions of Manifolds*, Academic Press, New York (1986).
[55] R.J. Daverman and J.J. Walsh, *Examples of cell-like maps that are not shape equivalences*, Michigan J. Math. **30** (1983), 17–30.
[56] K.J. Devlin and H. Johnsbråten, *The Souslin Problem*, Springer, Berlin (1974).
[57] J. Dieudonné, *Une généralisation des espaces compacts*, J. Math. Pures Appl. **23** (1944), 65–76.
[58] T. Dobrowolski and J. Mogilski, *Problems on topological classification of incomplete metric spaces*, Open Problems in Topology, J. van Mill and G.M. Reed, eds., North-Holland, Amsterdam (1990), 409–429.
[59] E.K. van Douwen, J.D. Monk and M. Rubin, *Some questions about Boolean algebras*, Algebra Universalis **11** (1980), 220–243.
[60] A. Dow, P. Simon and J.E. Vaughan, *Strong homology and the proper forcing axiom*, Proc. Amer. Math. Soc. **106** (1989), 821–828.
[61] A. Dow and J. Vermeer, *Not all σ-complete Boolean algebras are quotients of complete Boolean algebras*, Proc. Amer. Math. Soc. **116** (1992), 1175–1177.
[62] H. Dowker, *On countably paracompact spaces*, Canad. J. Math. **3** (1951), 219–224.

[63] A.N. Dranišnikov, *On a problem of P.S. Alexandrov*, Mat. Sb. **135** (1988), 551–557.
[64] J. Dugundji, *An extension of Tietze's theorem*, Pacific J. Math. **1** (1951), 353–367.
[65] J. Dugundji, *Topology*, Allyn and Bacon, Boston (1970).
[66] J. Dydak and J.J. Walsh, *Infinite-dimensional dimensional compacta having cohomological dimension two: an application of the Sullivan conjecture*, Topology **32** (1993), 93–104.
[67] P. Ecklof, *Whitehead's problem is undecidable*, Amer. Math. Monthly **83** (1976), 775–788.
[68] H. Freudenthal (ed.), *L.E.J. Brouwer Collected Works, Vol. 2, Geometry, Analysis, Topology and Mechanics*, North-Holland, Amsterdam (1976).
[69] I. Grattan-Guinness (ed.), *From the Calculus to Set Theory 1630–1910*, Duckworth, London (1980).
[70] R. Ellis, *Lectures on Topological Dynamics*, Benjamin, New York (1969).
[71] R. Engelking, *General Topology*, Heldermann, Berlin (1989).
[72] R. Engelking, *Theory of Dimensions: Finite and Infinite*, Heldermann, Berlin (1995).
[73] M. Fort, *Homogeneity of infinite products of manifolds with boundary*, Pacific J. Math. **12** (1962), 879–884.
[74] M. Fréchet, *Généralisation d'un théorème de Weierstrass*, Comptes Rendus Hebdomadaires des Scéances de l'Académie des Sciences **139** (1904), 848–850.
[75] M. Fréchet, *La notion d'écart dans le calcul fonctionnel*, Comptes Rendus Hebdomadaires des Scéances de l'Académie des Sciences **140** (1905), 772–774.
[76] M. Fréchet, *Sur les fonctions d'une infinité de variables*, Comptes Rendus Hebdomadaires des Scéances de l'Académie des Sciences **140** (1905), 567–568.
[77] M. Fréchet, *Sur les fonctions limites et les opérations fonctionelles*, Comptes Rendus Hebdomadaires des Scéances de l'Académie des Sciences **140** (1905), 27–29.
[78] M. Fréchet, *Les Espaces Abstraits*, Hermann, Paris (1928).
[79] D.H. Fremlin, *Consequences of Martin's Axiom*, Cambridge Univ. Press, Cambridge, UK (1984).
[80] L. Gillman and M. Jerison, *Rings of Continuous Functions*, Van Nostrand, Princeton, NJ (1960).
[81] K. Gödel, *The Consistency of the Continuum Hypothesis*, Princeton Univ. Press, Princeton, NJ, USA (1940).
[82] J.V. Grabiner, *The Origins of Cauchy's Rigorous Calculus*, MIT Press, Cambridge, MA (1981).
[83] M.F. Hallett, *Towards a Theory of Mathematical Research Programmes*, British J. Philos. Sci. **30** (1979), 1–25 and 135–159.
[84] F. Hausdorff, *Grundzüge der Mengenlehre*, Chelsea, New York (1965), Reprint of the 1914 Berlin edition.
[85] W.E. Haver, *Mappings between ANR's that are fine homotopy equivalences*, Pacific J. Math. **58** (1975), 457–461.
[86] R. Haydon, *On dual L^1-spaces and injective bidual Banach spaces*, Israel J. Math. **31** (1978), 142–152.
[87] E. Heine, *Die elemente der Functionenlehre*, J. Crelle **74** (1872), 172–188.
[88] D. Hilbert, *Grundlagen der Geometrie*, Teubner, Leipzig/Berlin (1913).
[89] D. Hilbert, *Über das Unendliche*, Math. Ann. **68** (1926), 161–191.
[90] N. Hindman, *Finite sums from sequences within cells of a partition of N*, J. Combin. Theory Ser. A **17** (1974), 1–11.
[91] N. Hindman, *Ultrafilters and combinatorial number theory*, Number Theory Carbondale, M. Nathanson, ed., Springer, Berlin (1979), 119–184.
[92] S.-T. Hu, *Elements of General Topology*, Holden-Day, San Francisco/London/Amsterdam (1965).
[93] S.-T. Hu, *Theory of Retracts*, Wayne State Univ. Press, Detroit (1965).
[94] W. Hurewicz and H. Wallman, *Dimension Theory*, Van Nostrand, Princeton, NJ (1941).
[95] M. Hušek and J. van Mill (eds), *Recent Progress in General Topology*, North-Holland, Amsterdam (1992).
[96] T. Jech, *Nonprovability of Souslin's hypothesis*, Comm. Math. Univ. Carolinae **8** (1967), 291–305.
[97] R.B. Jensen, *Souslin's problem is incompatible with $V = L$*, Notices Amer. Math. Soc. **15** (1968), 935.
[98] D.M. Johnson, *The problem of the invariance of dimension, part I*, Arch. Hist. Exact Sci. **20** (1979), 97–188.
[99] D.M. Johnson, *The problem of the invariance of dimension, part II*, Arch. Hist. Exact Sci. **25** (1981), 85–267.
[100] I. Juhász, *Cardinal Functions in Topology – Ten Years Later*, Mathematisch Centrum, Amsterdam (1980).
[101] M.I. Kadec, *On topological equivalence of separable Banach spaces*, Dokl. Akad. Nauk SSSR **167** (1966), 23–25 (in Russian); English translation: Soviet Math. Dokl. **7** (1966), 319–322.
[102] M. Katětov, *On real-valued functions in topological spaces*, Fund. Math. **38** (1951), 85–91.
[103] A.S. Kechris, *Classical Descriptive Set Theory*, Springer, Berlin (1995).

[104] O.H. Keller, *Die Homoiomorphie der kompakten konvexen Mengen in Hilbertschen Raum*, Math. Ann. **105** (1931), 748–758.

[105] V. Klee, *Some topological properties of convex sets*, Trans. Amer. Math. Soc. **78** (1955), 30–45.

[106] V. Klee, *A note on topological properties of normed linear spaces*, Proc. Amer. Math. Soc. **7** (1956), 673–674.

[107] T. Koetsier, *Lakatos' Philosophy of Mathematics, A Historical Approach*, North-Holland, Amsterdam (1991).

[108] T. Koetsier and J. van Mill, *General topology, in particular, dimension theory in the Netherlands: the decisive influence of Brouwer's intuitionism*, Handbook of the History of General Topology, C.E. Aull and R. Lowen, eds, Kluwer Academic, Dordrecht (1997), 135–180.

[109] M. Kojman and S. Shelah, *A ZFC Dowker space in $\aleph_{\omega+1}$: An application of pcf theory to topology*, Preprint (1995).

[110] S. Koppelberg, *Homomorphic images of σ-complete Boolean algebras*, Proc. Amer. Math. Soc. **51** (1975), 171–175.

[111] G. Kozlowski, *Images of ANRs*, unpublished manuscript.

[112] G. Kozlowski and P. Zenor, *A differentiable, perfectly normal, nonmetrizable manifold*, Topology Proc. **4** (1979), 453–461.

[113] K. Kunen, *Products of S-spaces* (1975), unpublished manuscript.

[114] K. Kunen, *Set Theory. An Introduction to Independence Proofs*, Stud. Logic Found. Math. vol. 102, North-Holland, Amsterdam (1980).

[115] K. Kunen, *A compact L-space under CH*, Topology Appl. **12** (1981), 283–287.

[116] K. Kunen and J.E. Vaughan (eds), *Handbook of Set-Theoretic Topology*, North-Holland, Amsterdam (1984).

[117] K. Kuratowski, *Sur quelques problèmes topologiques concernant le prolongement des fonctions continus*, Compositio Math. **2** (1951), 186–191.

[118] K. Kuratowski, *Topology*, Vol. I, Academic Press, New York (1966).

[119] K. Kuratowski, *Topology*, Vol. II, Academic Press, New York (1968).

[120] K. Kuratowski and A. Mostowski, *Set Theory*, PWN–North-Holland, Warszawa/Amsterdam (1976).

[121] A. Louveau, *Charactérisation des sous-espaces compacts de $\beta\mathbb{N}$*, Bull. Sci. Math. **97** (1973), 259–263.

[122] C.B. De Lyra, *On spaces of the same homotopy type as polyhedra*, Bol. Soc. Mat. São Paulo **12** (1957), 43–62.

[123] J.H. Manheim, *The Genesis of Point Set Topology*, Pergamon Press, Oxford (1964).

[124] S. Mardešić and A. Prasolov, *Strong homology is not additive*, Trans. Amer. Math. Soc. **12** (1988), 725–744.

[125] S. Mardešić and J. Segal, *Shape Theory*, North-Holland, Amsterdam (1982).

[126] S. Mazur, *Une remarque sur l'homéomorphie des champs Fonctionels*, Studia Math. **1** (1929), 83–85.

[127] S. Mercourakis and S. Negrepontis, *Banach spaces and topology II*, Recent Progress in General Topology, M. Hušek and J. van Mill, eds, North-Holland, Amsterdam (1992), 493–536.

[128] E. Michael, *Continuous selections I*, Ann. of Math. **63** (1956), 361–382.

[129] E. Michael, *Continuous selections II*, Ann. of Math. **64** (1956), 562–580.

[130] E. Michael, *Continuous selections III*, Ann. of Math. **65** (1957), 556–575.

[131] J. van Mill, *Infinite-Dimensional Topology: Prerequisites and Introduction*, North-Holland, Amsterdam (1989).

[132] J. van Mill and G.M. Reed (eds), *Open Problems in Topology*, North-Holland, Amsterdam (1990).

[133] R.J. Miller, *Mapping cylinder neighborhoods of some ANRs*, Ann. of Math. **106** (1977), 1–18.

[134] J. Milnor, *On axiomatic homology theory*, Pacific J. Math. **12** (1962), 337–341.

[135] A.F. Monna, *Dirichlet's Principle. A Mathematical Comedy of Errors and its Influence on the Development of Analysis*, Oosthoek, Scheltema and Holkema, Utrecht (1975).

[136] G.H. Moore, *Zermelo's Axiom of Choice, Its Origins, Development, and Influence*, Springer, Berlin (1982).

[137] K. Morita, *On generalizations of Borsuk's homotopy extension theorem*, Fund. Math. **88** (1975), 1–6.

[138] K. Morita and J. Nagata, *Topics in General Topology*, North-Holland, Amsterdam (1989).

[139] S.B. Nadler, *Hyperspaces of Sets*, Marcel Dekker, New York/Basel (1978).

[140] J.I. Nagata, *On a necessary and sufficient condition of metrizability*, J. Inst. Polytech. Osaka City Univ. Ser. A Math. **1** (1950), 93–100.

[141] J.I. Nagata, *Modern Dimension Theory*, Heldermann, Berlin (1983).

[142] J.I. Nagata, *Modern General Topology*, North-Holland, Amsterdam (1985).

[143] S. Negrepontis, *Banach spaces and topology*, Handbook of Set-Theoretic Topology, K. Kunen and J.E. Vaughan, eds, North-Holland, Amsterdam (1984), 1045–1142.
[144] A. Pełczyński, *On Banach spaces containing $L^1(\mu)$*, Studia Math. **30** (1968), 231–246.
[145] W. Purkert and H.J. Ilgauds, *Georg Cantor*, Birkhäuser, Basel/Boston/Stuttgart (1987).
[146] C. Reid, *Hilbert*, Springer, Berlin/New York (1970).
[147] F. Riesz, *Die Genesis des Raumbegriffes*, Math. Naturwiss. Berichte Ungarn **24** (1907), 309–353; References are to Riesz, *Werke*, 110–154.
[148] F. Riesz, *Stetigkeit und abstrakte Mengenlehre*, Atti del IV Congresso Internazionale dei Matematici, Vol. 2, Rome (1908); References are to Riesz, *Werke*, 18–24, 155–161.
[149] M.E. Rudin, *Countable paracompactness and Souslin's problem*, Canad. J. Math. **7** (1955), 543–547.
[150] M.E. Rudin, *An new proof that metric spaces are paracompact*, Proc. Amer. Math. Soc. **20** (1969), 603.
[151] M.E. Rudin, *A normal space X for which $X \times \mathbb{I}$ is not normal*, Fund. Math. **78** (1971), 179–186.
[152] M.E. Rudin, *Lectures on Set Theoretic Topology*, CBMS 23, Amer. Math. Soc., Providence, RI, USA (1975).
[153] M.E. Rudin, *The undecidability of the existence of a perfectly normal nonmetrizable manifold*, Houston J. Math. **5** (1979), 249–252.
[154] M.E. Rudin and P. Zenor, *A perfectly normal nonmetrizable manifold*, Houston J. Math. **2** (1976), 129–134.
[155] A. Schoenflies, *Die Entwicklung der Lehre von den Mannigfaltigkeiten*, Teil 2, Teubner, Leipzig (1908).
[156] E. Scholz, *Logische Ordnungen im Chaos: Hausdorffs frühe Beiträge zur Mengenlehre*, Felix Hausdorff zum Gedächtnis, E. Brieskorn, ed., Westdeutscher Verlag, Opladen (1996), 107–134.
[157] R.M. Schori and J.E. West, *The hyperspace of the closed interval is a Hilbert cube*, Trans. Amer. Math. Soc. **213** (1975), 217–235.
[158] S. Shelah, *Infinite Abelian groups – Whitehead problem and some constructions*, Israel J. Math. **18** (1974), 243–256.
[159] S. Shelah, *A compactness theorem for singular cardinals, free algebras, Whitehead problem and transversals*, Israel J. Math. **21** (1975), 319–349.
[160] S. Shelah, *A Banach space with few operators*, Israel J. Math. **30** (1978), 181–191.
[161] S. Shelah, *Proper Forcing*, Lecture Notes in Mathematics vol. 940, Springer, Berlin (1982).
[162] S. Shelah, *Lifting problem of the measure algebra*, Israel J. Math. **45** (1983), 90–96.
[163] R. Siegmund-Schultze, *Functional analysis*, Companion Encyclopedia of the History and Philosophy of the Mathematical Sciences, Vol. I, Routledge (1994), 375–384.
[164] J.M. Smirnov, *A necessary and sufficient condition for metrizability of a topological space*, Dokl. Akad. Nauk SSSR **7** (1951), 197–200.
[165] R.M. Solovay and S. Tennenbaum, *Iterated Cohen extensions and Souslin's problem*, Ann. of Math. **94** (1971), 201–245.
[166] M. Souslin, *On problème 3*, Fund. Math. **1** (1920), 223.
[167] M. Starbird, *The Borsuk homotopy extension theorem without the binormality condition*, Fund. Math. **87** (1975), 207–211.
[168] A. Stone, *Paracompactness and product spaces*, Bull. Amer. Math. Soc. **54** (1948), 977–982.
[169] M. Talagrand, *Separabilité vague dans l'espace des measures sur un compact*, Israel J. Math. **37** (1980), 171–180.
[170] A.E. Taylor, *A study of Maurice Fréchet: I. His early work on point set theory and the theory of functionals*, Arch. Hist. Exact Sci. **27** (1982), 233–295.
[171] J.L. Taylor, *A counterexample in shape theory*, Bull. Amer. Math. Soc. **81** (1975), 629–632.
[172] S. Tennenbaum, *Souslin's problem*, Proc. Nat. Acad. USA **59** (1968), 60–63.
[173] H. Toda, *On unstable homotopy of spheres and classical groups*, Proc. Nat. Acad. Sci. **46** (1960), 1102–1105.
[174] S. Todorčević, *Trees and linearly ordered sets*, Handbook of Set-Theoretic Topology, K. Kunen and J.E. Vaughan, eds, North-Holland, Amsterdam (1984), 235–293.
[175] S. Todorčević, *Partition problems in topology*, Contemp. Math. vol. 84, Amer. Math. Soc., Providence, RI (1988).
[176] S. Todorčević, *Topics in Topology*, Springer, Berlin (1997).
[177] H. Toruńczyk, *Concerning locally homotopy negligible sets and characterizations of ℓ_2-manifolds*, Fund. Math. **101** (1978), 93–110.

[178] H. Toruńczyk, *On CE-images of the Hilbert cube and characterizations of Q-manifolds*, Fund. Math. **106** (1980), 31–40.

[179] H. Toruńczyk, *Characterizing Hilbert space topology*, Fund. Math. **111** (1981), 247–262.

[180] B. van der Waerden, *Beweis einer Baudetschen Vermutung*, Nieuw Arch. Wisk. **19** (1927), 212–216.

[181] A.D. Wallace, *A note on mobs. I*, An. Acad. Brasil. Ciênc. **24** (1952), 239–334.

[182] A.D. Wallace, *A note on mobs. II*, An. Acad. Brasil. Ciênc. **25** (1953), 335–336.

[183] J.J. Walsh, *Infinite-dimensional compacta containing no n-dimensional* ($n \geqslant 1$) *subsets*, Topology **18** (1979), 91–95.

[184] J.J. Walsh, *Dimension, cohomological dimension and cell-like mappings*, Shape Theory and Geometric Topology Conf., Dubrovnik, S. Mardešić and J. Segal, eds, Lecture Notes in Math. vol. 870, Springer, Berlin (1981), 105–118.

[185] J.E. West, *Infinite products which are Hilbert cubes*, Trans. Amer. Math. Soc. **150** (1970), 1–25.

[186] J.E. West, *Mapping Hilbert cube manifolds to ANR's: a solution to a conjecture of Borsuk*, Ann. of Math. **106** (1977), 1–18.

[187] J.E. West, *Problems in infinite-dimensional topology*, Open Problems in Topology, J. van Mill and G.M. Reed, eds, North-Holland, Amsterdam (1990), 523–597.

[188] H. Weyl, *Die Idee der Riemannschen Fläche*, Teubner, Leipzig/Berlin (1913).

[189] R.L. Wilder, *Topology of Manifolds*, Amer. Math. Soc., Providence, RI (1949).

[190] M. Wojdysławski, *Rétractes absolus et hyperespaces des continus*, Fund. Math. **32** (1939), 184–192.

[191] R.Y.T. Wong, *A wild Cantor set in the Hilbert cube*, Pacific J. Math. **24** (1968), 189–193.

This Page Intentionally Left Blank

CHAPTER 9

Absolute Neighborhood Retracts and Shape Theory

Sibe Mardešić

Department of Mathematics, University of Zagreb, Bijenička cesta 30, 10 000 Zagreb, Croatia
E-mail: smardes@math.hr

Absolute neighborhood retracts (ANR's) and spaces having the homotopy type of ANR's, like polyhedra and CW-complexes, form the natural environment for homotopy theory. Homotopy-like properties of more general spaces (shape properties) are studied in shape theory. This is done by approximating arbitrary spaces by ANR's. More precisely, one replaces spaces by suitable systems of ANR's and one develops a homotopy theory of systems. This approach links the theory of retracts to the theory of shape. It is, therefore, natural to consider the history of both of these areas of topology in one article. A further justification for this is the circumstance that both theories owe their fundamental ideas to one mathematician, Karol Borsuk. We found it convenient to organize the article in two sections, devoted to retracts and to shape, respectively.

1. Theory of retracts

The problem of extending a continuous mapping $f: A \to Y$ from a closed subset A of a space X to all of X, or at least to some neighborhood U of A in X, is very often encountered in topology. Karol Borsuk realized that the particular case, when $Y = X$ and f is the inclusion $i: A \to X$, deserves special attention. In this case, any extension of i is called a *retraction* (*neighborhood retraction*). If retractions exist, A is called a *retract* (*neighborhood retract*) of X. In his Ph.D. thesis "O retrakcjach i zbiorach związanych" ("On retractions and related sets"), defended in 1930 at the University of Warsaw, Borsuk introduced and studied these basic notions as well as the topologically invariant notion of *absolute retract* (abbreviated as AR). He thus laid the foundations of the *theory of retracts*. The very suggestive term *retract* was proposed by Stefan Mazurkiewicz (1888–1945), who was Borsuk's Ph.D. supervisor. The term *absolute retract* was suggested by Borsuk's colleague Nachman Aronszajn, also a student of Mazurkiewicz.

It appears that the original of Borsuk's thesis has been lost in the turmoils of the Second World War. However, its main results were published in [27]. *Absolute neighborhood retracts* (abbreviated as ANR) were introduced in [28]. In the beginning Borsuk only considered separable metric spaces, especially metric compacta. Other early contributions to

HISTORY OF TOPOLOGY
Edited by I.M. James

K. Borsuk was born in Warsaw in 1905 and studied mathematics at the University of Warsaw. He spent the period 1931–1932 on postdoctoral studies with leading European topologists of that time (Karl Menger in Wien, Heinz Hopf in Zürich and Leopold Vietoris in Innsbruck). His habilitation at the University of Warsaw took place in 1934 and he became Professor in 1946. After retirement in 1975, he continued with activities at the Mathematical Institute of the Polish Academy of Science. On four different occasions, Borsuk spent a Winter semester in US (Princeton, Berkeley, Madison, New Brunswick), which contributed to the quick spreading of his theory of retracts and later, the theory of shape. Borsuk died in Warsaw in 1982. N. Aronszajn was born in 1907 in Warsaw, where he went to school and university, obtaining his Ph.D. in 1930. He then worked in Paris and Cambridge until 1948, when he emmigrated to US. There he spent most of his career at the University of Kansas in Lawrence. He died in Corvallis, Oregon in 1980.

the theory of retracts, due to K. Kuratowski [154] and R.H. Fox [107], also refer to these classes of spaces. Gradually, the theory was extended, first to arbitrary metric spaces by C.H. Dowker [79] and J. Dugundji [85], then to more general classes of spaces $\mathcal{C}$, closed under homeomorphic images and closed subsets, by S.-T. Hu [126], O. Hanner [121] and E.A. Michael [184].

An AR (ANR) for the class $\mathcal{C}$ is a space Y from $\mathcal{C}$, such that, whenever Y is a closed subset of a space X from $\mathcal{C}$, then Y is a retract (neighborhood retract) of X. A space Y is an *absolute extensor* (*absolute neighborhood extensor*) for the class $\mathcal{C}$, abbreviated as AE($\mathcal{C}$) (ANE($\mathcal{C}$)), provided, for every closed subset A of a space X from $\mathcal{C}$, every mapping $f: A \to Y$ extends to all of X (to some neighborhood U of A in X). It is not required that Y belongs to $\mathcal{C}$. Clearly, if Y is from $\mathcal{C}$ and is an absolute extensor for $\mathcal{C}$, then Y is also an absolute retract for $\mathcal{C}$. The terminology AE and ANE was introduced in [184]. Gradually it became clear that the class $\mathcal{C}$ of metric spaces gives the most satisfactory theory. Hence, if we speak of ANR's and do not specify $\mathcal{C}$, we mean ANR's for metric spaces. A rather detailed and reliable study of the spaces ANR($\mathcal{C}$) and ANE($\mathcal{C}$), for various classes $\mathcal{C}$, has been carried out in Hu's monograph [128].

Borsuk's work on the theory of retracts had its precedents. The most important among these is the *Tietze–Urysohn extension theorem*. It was first proved, for metric spaces by H. Tietze [225]. Then P.S. Uryson proved his famous lemma: If A and B are closed disjoint subsets of a normal space X, there exists a mapping $f: X \to I$ to the real line segment $I = [0, 1]$ such that $f|A = 0$ and $f|B = 1$ [232]. In the case of metric spaces, the assertion of *Urysohn's lemma* is an elementary fact, which was used in Tietze's argument. Replacing this fact by its generalization enabled Uryson to obtain the extension theorem

Ernest A. Michael, Professor at the University of Washington in Seattle, was born in Zürich in 1925. He obtained the Ph.D. in 1951 from the University of Chicago. Sze-Tsen Hu, Professor at the University of California in Los Angeles, was born in Huchow, China in 1914. He obtained the B.Sc. from the University of Nanking, China and the D.Sc. in 1959 from the University of Manchester.

> Heinrich Tietze (1880–1964) was born in Schleinz, Austria. He studied in Wien, München and Göttingen and obtained his Ph.D. in 1904 in Wien, where he became Privatdozent in 1908. From 1910 to 1919 he was professor at the Technical University in Brno and it is during that period that he obtained his extension theorem. He spent the rest of his career at the universities of Erlangen and München. Pavel Samuilovich Uryson (1898–1924) was born in Odessa. He was a student of D.F. Egorov (1869–1931) and N.N. Luzin (1883–1950) in Moscow, where he obtained his Ph.D. in 1921. Urysohn was one of the most promising Russian mathematicians of his generation, when he lost his life at the age of 25 in a tragic accident, while swimming in the rough seas of French Bretagne. His collected papers fill up two volumes.

for normal spaces. In present terminology the theorem asserts that $I = [0, 1]$ and the real line $\mathbb{R}$ are AE's for normal spaces. Recently, J. Mioduszewski drew attention to the fact that the argument used by Uryson in constructing the mapping $f : X \to I$ appeared a year earlier (in a different context) in the only paper by W.S. Bogomolowa, a student of Luzin [25].

An important question raised in the early days of the theory of retracts was to determine whether an absolute retract Y for a class $\mathcal{C}$ is necessarily an absolute extensor for $\mathcal{C}$. This is true for many important classes $\mathcal{C}$. For separable metric spaces it was proved in [154] and for arbitrary metric spaces in [85]. To obtain this result, one first embeds Y in a normed vector space L, in such a way that it is a closed subset of its convex hull K. For L one can take the space of bounded mappings $f : Y \to \mathbb{R}$, which is even a Banach space [155, 248]. Then one applies the *Dugundji extension theorem* [85], an important generalization of the Tietze–Urysohn theorem. It asserts that every convex set in a normed vector space (more generally, in a locally convex vector space) is an absolute extensor for metric spaces. This result was made possible only after A.H. Stone proved that metric spaces are paracompact [223]. For separable metric spaces Dugundji's extension theorem was already known to Polish topologists. Note that paracompactness of these spaces is an elementary fact, because separable metric spaces are Lindelöf, hence, also paracompact. Dugundji's theorem was later generalized to *stratifiable spaces* [26], a class of spaces, introduced in [56], which includes both metric spaces and CW-complexes.

An important result in the theory of retracts was J.H.C. Whitehead's theorem that the adjunction space of a mapping $f : A \to Y$, where $A \subseteq X$, X and Y are compact ANR's, is again a compact ANR [244]. Another important result was obtained by Hanner. He considered *local* ANE's, i.e. spaces which admit an open covering formed by ANE's, and proved that for metric (more generally, for paracompact) spaces, every local ANE is an ANE [121]. This theorem implies, e.g., that (metric) manifolds are ANR's.

In introducing (compact) ANR's Borsuk wanted to generalize compact polyhedra in a way which excludes the pathology often present in arbitrary metric compacta. For example, compact AR's have the fixed-point property [27], but there exist acyclic (locally connected) continua in $\mathbb{R}^3$ which do not have this property [29]. Generalizing a sum theorem from [9], Borsuk proved that the union $X = A_1 \cup A_2$ of two compact ANR's is an ANR, provided $A_1 \cap A_2$ is an ANR [28]. This implies that every compact polyhedron is indeed an ANR. In the same paper he showed that in the class of finite-dimensional compacta, ANR's are characterized by *local contractibility*. For an infinite-dimensional compactum X, local contractibility alone is not sufficient to ensure that X be an ANR [31].

Kazimierz Kuratowski (1896–1980), one of the founders of the Polish topology school, was born and died in Warsaw. He obtained his Ph.D. from the University of Warsaw in 1921. He first worked at the Technical University in Lwów. Since 1934 he was Professor at the University of Warsaw. During the Nazi occupation of Poland, both Kuratowski and Borsuk lectured at the underground university in Warsaw. Clifford Hugh Dowker was born in 1912 in a rural area of Western Ontario. He obtained his B.A. and M.A. in Canada and his Ph.D. in 1938 in Princeton, where he came to study under Solomon Lefschetz (1884–1972). In 1950, during the period of McCarthyism, Dowker moved to England and eventually became Professor at Birckbeck College in London, where he worked until his retirement in 1979. He died in London in 1982. James Dugundji (1919–1985) was born in New York in a family of Greek immigrants. He obtained his B.A. degree from New York University in 1940. The same year he started his graduate studies at the University of North Carolina at Chapel Hill as a student of Witold Hurewicz (1904–1956). After spending four years of war in the US. Air Force, in 1946 he entered the Massachusetts Institute of Technology, where Hurewicz became Professor in 1945. Under him Dugundji obtained his Ph.D. in 1948. The same year he started teaching at the University of Southern California in Los Angeles, where he became Professor in 1958. The Swedish topologist Olof Hanner was born in Stockholm in 1922 and obtained his Ph.D. from the University of Stockholm in 1952 with a thesis which consisted of three of his papers on ANR's. He became interested in ANR's during a visit to the Institute for Advanced Study in Princeton in 1949/50, where he came in touch with the work of Ralph Hartzler Fox (1913–1973) and Lefschetz.

The important property LC^n (*local connectedness up to dimension* n) was introduced in 1930 in Lefschetz's book [158] (see p. 91) and studied further in [159]. Generalizing Borsuk's work, Kuratowski proved that an n-dimensional separable metric space X is an ANR if and only if it is LC^n [154]. The proof uses the fundamental concepts of *nerve of an open covering* and *canonical mapping*, whose origins can be traced back to the work of Alexandroff [2, 3] and Kuratowski [153], respectively. The Kuratowski theorem was later generalized to arbitrary metric spaces by several authors [151, 145, 87].

In 1973 W.E. Haver proved that a locally contractible metric space X, which is the union of a countable collection of finite-dimensional compacta, is an ANR [123]. This result had important consequences in the study of the space PLH(M) of piecewise linear homeomorphisms of a compact PL-manifold M. A.V. Chernavskiĭ proved in 1969 that the space H(M) of homeomorphisms of a compact manifold M is locally contractible [68]. A simplified proof of Chernavskiĭ's result was obtained by R.D. Edwards and R.C. Kirby [104]. It follows from this proof that, for a compact PL-manifold M, the space PLH(M) is also locally contractible. On the other hand, R. Geoghegan showed that, for a compact polyhedron P, PLH(P) is the union of a countable collection of compact finite-dimensional sets [113]. Consequently, Haver's result applies and yields the conclusion that, for a compact PL-manifold M, PLH(M) is an ANR. For compact topological manifolds M, the question if H(M) is an ANR, is still open. The analogous question for Q-manifolds was answered in the affirmative, independently by S. Ferry [105] and by H. Toruńczyk [229].

In general, the geometric realization of an infinite *simplicial complex* K can be endowed with the *weak topology* (also called CW-*topology*) or the *metric topology* [161]. The resulting spaces will be denoted by $|K|$ and $|K|_m$, respectively. We refer to spaces $|K|$ as *polyhedra*. A polyhedron $|K|$ is metrizable if and only if the complex K is locally finite.

S. Lefschetz was born in Moscow and educated in Paris. After working for some years in industry, he turned to mathematics (following an industrial accident in which he lost both hands). In 1925 he joined the Mathematics Department in Princeton, where he became a leading topologist, together with O. Veblen (1880–1960) and J.W. Alexander (1888–1971). After retiring from Princeton University, he continued his activities at Brown University and in Mexico. John Henry Constantine Whitehead (1904–1960), another leading topologist, was born in India and educated in Oxford. He continued his studies in Princeton under Veblen and there obtained his Ph.D. in 1931. He became Professor in Oxford in 1945. He died in Princeton, where he was spending a year's leave.

In this case the weak topology and the metric topology coincide. Polyhedra are special cases of CW-*complexes* (CW-*spaces*), introduced by Whitehead in [246]. It was shown by Dugundji [86] that CW-complexes are paracompact spaces and ANE's for metric spaces. For polyhedra, the latter assertion was proved independently by Y. Kodama [144]. If a polyhedron $|K|$ is locally compact, then the complex K is locally finite and therefore, $|K| = |K|_m$ must be an ANR.

For an arbitrary simplicial complex K, the space $|K|_m$ is an ANR. To prove this important fact, one first proves the assertion in the special case of *full* simplicial complexes, i.e. complexes where every finite set of vertices spans a simplex. This is easily done by applying Dugundji's extension theorem. In the general case, one needs the fact that, for every subcomplex $L \subseteq K$, $|L|_m$ is a neighborhood retract of $|K|_m$. The standard argument consists of showing that, in the first barycentric subdivision K' of K, the star of the carrier of L' is an open set in the carrier of K', which retracts to the carrier of L'. However, to apply this argument, one needs to know that $|K|_m = |K'|_m$. This was proved by Lefschetz in [161], a monograph devoted entirely to local n-connectedness and retraction.

The French topologist Robert Cauty studied closely the relationship between polyhedra and CW-complexes. In particular, in [50] he characterized spaces which embed into polyhedra as closed subsets. All CW-complexes satisfy his criterion. Moreover, if a CW-complex X is embedded as a closed subset of a polyhedron P, then there exists an open neighborhood U of X in P which retracts to X. It is well known that every open subset of a polyhedron is itself a polyhedron. Therefore, CW-complexes are retracts of polyhedra. Twenty years later, Cauty showed that an open subset of a CW-complex need not be a CW-complex [53]. He thus corrected an error, appearing occasionally in the literature.

Cauty showed that there exist CW-complexes which are not ANR's for paracompact (hereditarily paracompact) spaces [49]. An example, due to E. van Douwen and R. Pol [78] shows that, there exist a regular countable space X (hence, a Lindelöf space), a closed subset $A \subseteq X$ and a mapping f of A to a 1-dimensional polyhedron $|K|$, which does not extend to any neighborhood of A. Consequently, $|K|$ is a not an ANE for paracompact spaces. On the other hand, for a simplicial complex K with no infinite simplices, $|K|_m$ is an ANE even for collectionwise normal spaces [52]. This shows that the extension properties of complexes depend essentially on the choice of the topology.

Cauty proved that every CW-complex is an ANR for stratifiable spaces [51]. This was achieved using *topological convexity* (abbreviated as TC) and *local topological convexity* (abbreviated as TLC). A space X is TLC provided there exists a neighborhood U of the diagonal Δ in $X \times X$ and there exists a mapping $\phi : U \times I \to X$ such that $\phi(x, y, 0) = x$,

R. Cauty was born in 1946. He studied in Paris and belonged to M. Zisman's Algebraic Topology Seminar. He obtained his doctorat d'état in 1972. Since in Paris there was not much interest in General Topology, Cauty learned the subject by himself, beginning with Kuratowski's *Topologie*. ANR's and complexes, being the meeting ground of General and Algebraic Topology, constituted the natural topic of his research.

$\phi(x, y, 1) = y$, for all $(x, y) \in U$ and $\phi(x, x, t) = x$, for all $x \in X$, $t \in I$. In addition, one requires that every point $x \in X$ admits a basis of neighborhoods V such that $V \times V \subseteq U$ and $\phi(V \times V \times I) \subseteq V$. Property TC is obtained by requiring that $U = X \times X$. Clearly, locally convex topological vector spaces have property TC. There exist compact ANR's which are not TLC-spaces [42] (also see [33], Ch. VI.4).

A weaker notion, called *equiconnectedness* (*local equiconnectedness*) was already considered by Fox [108] and J.-P. Serre [212], who used the abbreviations UC (ULC). These properties are obtained from properties TC (TLC) by omitting the additional condition $\phi(V \times V \times I) \subseteq V$. It is easy to see that every AR (ANR) is a UC-space (ULC-space). Finite-dimensional metric ULC-spaces are ANR's [87]. For infinite-dimensional metric ULC-spaces, one finds in [88, 125] additional conditions, which make these spaces ANR's. The question whether every metric ULC-space is an ANR, remained open for a long time. Only recently, a counterexample was obtained by Cauty, who exhibited a metric linear space (hence, a UC-space), which is not an AR [55]. Cauty's example depends essentially on the existence of dimension-raising cell-like mappings of compacta [83].

In the literature there are many results characterizing ANR's. Here we mention a classical criterion, based on realizations of simplicial complexes K with respect to a covering $\mathcal{U}$. A *full realization* of K is a mapping $g : |K| \to X$ of the geometric realization of K (CW-topology) such that every (closed) simplex $\sigma \in K$ maps into some member U of $\mathcal{U}$. A *partial realization* is a mapping $f : |L| \to X$, defined on the carrier of some subcomplex L of K such that, for every $\sigma \in K$, the set $f(|L| \cap \sigma)$ is contained in some member U of $\mathcal{U}$. A metric space X is an ANR if and only if every open covering $\mathcal{U}$ of X admits a refinement $\mathcal{V}$ such that, for every subcomplex $L \subseteq K$, which contains all the vertices of K, every partial realization $f : |L| \to X$ with respect to $\mathcal{V}$ admits an extension to a full realization $g : |K| \to X$ with respect to $\mathcal{U}$. This was proved in [160], for compact metric spaces and in [87], for arbitrary metric spaces. The problem of finding convenient characterizations of infinite-dimensional ANR's still deserves attention.

A very useful theorem on ANR's asserts that sufficiently near mappings into an ANR must be homotopic. More precisely, if $\mathcal{U}$ is an open covering of an ANR Y, then there exists an open covering $\mathcal{V}$ such that any two $\mathcal{V}$-near mappings $\phi, \psi : X \to Y$ are $\mathcal{U}$-homotopic, i.e. are connected by a homotopy $H : X \times I \to Y$ with paths $H(x \times I)$, $x \in X$, contained in members of $\mathcal{U}$ [85, 120]. ANR's can be characterized as metrizable spaces Y having the property that, for every open covering $\mathcal{U}$, Y is $\mathcal{U}$-homotopy dominated by some polyhedron P, i.e. there exist mappings $f : Y \to P$, $g : P \to Y$ such that gf and id are $\mathcal{U}$-homotopic [87, 120]. Necessity of the condition is a consequence of the fact that every covering $\mathcal{V}$ of an ANR Y admits a polyhedron P and admits mappings f, g such that gf and id are $\mathcal{V}$-near mappings. Spaces Y having this property are called *approximate polyhedra* [173]. That every ANR Y is an approximate polyhedron is a consequence of the *bridge theorem,* which asserts that, for every mapping $f : X \to Y$ of a space into an ANR and for every open covering $\mathcal{V}$ of Y, there exists a normal covering $\mathcal{U}$ of X and a mapping $g : |N(\mathcal{U})| \to Y$ of the geometric realization of the nerve $N(\mathcal{U})$, such that, for any canonical mapping $p : X \to |N(\mathcal{U})|$, the mappings f and gp are $\mathcal{V}$-near [127].

It was John Milnor who in 1959 renewed the interest of topologists in the class of spaces having the homotopy type of CW-complexes [187]. ANR's belong to this class, because every ANR X has the *homotopy type* of the *geometric realization* $|S(X)|$ of its singular complex $S(X)$. For an arbitrary space X, there is a canonical mapping $j_X : |S(X)| \to X$, which is a weak homotopy equivalence, i.e. it induces isomorphisms of homotopy groups. By a well-known theorem, a weak homotopy equivalence between CW-complexes is a homotopy equivalence [129, 245]. This theorem readily extends to spaces homotopy dominated by CW-complexes and, therefore, applies to j_X, whenever X is an ANR. For every space X, $|S(X)|$ is triangulable and thus, every ANR has the homotopy type of a polyhedron. Actually, the geometric realization $|K|$ of any simplicial set K is a polyhedron. The proof given in [13] and reproduced in [169] contained an error, which was corrected in the Ph.D. thesis of Rudolf Fritsch, a student of Dieter Puppe [110–112].

Conversely, every polyhedron P has the homotopy type of an ANR. Indeed, if K is a simplicial complex such that $P = |K|$, then the identity mapping $|K| \to |K|_m$ is a homotopy equivalence [80]. However, $|K|_m$ is an ANR. A recent result of Cauty characterizes ANR's as metric spaces all of whose open subsets have the homotopy types of ANR's [54].

One of the most useful results on ANR's is Borsuk's *homotopy extension theorem* [30]. A pair of spaces (X, A) is said to have the *homotopy extension property* (abbreviated as HEP) with respect to a space Y, provided every mapping $f : (X \times 0) \cup (A \times I) \to Y$ admits an extension $F : X \times I \to Y$. Borsuk's theorem asserts that every pair, where X is a metric space and A is closed, has HEP with respect to any ANR Y. Among many generalizations of this theorem, especially interesting was the result of Dowker, which asserts that pairs, where $X \times I$ is normal and A is closed, have HEP with respect to separable Čech complete ANR's, in particular, with respect to compact ANR's [81]. This result naturally led to the question, does normality of X imply normality of $X \times I$? This proved to be a very challenging problem, which generated much research in general topology. It was finally solved in the negative by Mary Ellen Rudin [206].

After the development of the theory of fibrations [212, 130], it became clear that, for a pair (X, A), HEP with respect to all spaces Y, viewed as a property of the inclusion $A \to X$, is a notion dual to the notion of fibration, hence, it is referred to as a *cofibration*. Cofibration pairs (X, A) are also called *neighborhood deformation pairs* and play an important role in homotopy theory.

One of the central problems of geometric topology in the last decades has been the *recognition problem* for topological manifolds: Find a list of topological properties which characterize manifolds among topological spaces. The properties should be easy to check, hence, they should not use notions like homeomorphisms. In 1978 James W. Cannon solved the famous *double suspension problem*, by showing that the double suspension of a homology 3-sphere is homeomorphic to the 5-sphere S^5 [47]. This work led him to state the following conjecture, which would solve the recognition problem. *Cannon's conjecture*: A topological space X is an n-manifold (separable metric), $n \geqslant 5$, if and only if it is a *homology n-manifold* having the *disjoint disc property*. By definition, homology n-manifolds are finite-dimensional separable locally compact ANR's X, whose local homology groups (integer coefficients) coincide with those of $\mathbb{R}^n$, i.e. $H_m(X, X \backslash \{x\}; \mathbb{Z}) \approx$

$H_m(\mathbb{R}^n, \mathbb{R}^n \backslash \{0\}; \mathbb{Z})$, for all m. A metric space (X, d) has the disjoint disc property provided, for any two mappings $f_1, f_2 : B^2 \to X$ of the 2-ball B^2 and any $\varepsilon > 0$, there exist two mappings $g_1, g_2 : B^2 \to X$, such that $d(f_i, g_i) < \varepsilon$, $i = 1, 2$, and the images $g_1(B^2)$ and $g_2(B^2)$ are disjoint. Note that ANR's play an important role in this conjecture.

A major step towards proving Cannon's conjecture was the *cell-like approximation theorem* of R.D. Edwards, which considerably strengthened earlier work on cell-like mappings between manifolds [214]. Edwards announced his result in 1977 and published an outline of the proof in [98]. A detailed proof, for $n \geqslant 6$, appeared in Daverman's monograph [77]. The Edwards theorem asserts that a cell-like mapping $f : M \to X$ from an n-manifold M to a finite-dimensional space X is a near-homeomorphism, i.e. can be approximated by homeomorphisms, if and only if it has the *disjoint disc property*. Consequently, Cannon's conjecture is true provided X is the image of an n-manifold M under a cell-like mapping f. Edwards theorem is the crown of years of efforts of many geometric topologists. An essential ingredient in the proof is R.H. Bing's *shrinking criterion*, which gives necessary and sufficient conditions in order that a proper mapping $f : X \to Y$ be approximable by homeomorphisms. The criterion requires that for every pair of open coverings $\mathcal{U}$ of X and $\mathcal{V}$ of Y, there exists a homeomorphism $h : X \to X$ having the following properties:

(i) The mappings $fh, f : X \to Y$ are $\mathcal{V}$-near.

(ii) For every $y \in Y$, there exists a U in $\mathcal{U}$ such that $h(f^{-1}(y)) \subseteq U$ [182].

In view of the cell-like approximation theorem, to complete the proof of Cannon's conjecture, it would have been sufficient to show that every homology n-manifold X, $n \geqslant 5$, is *resolvable*, i.e. it is the cell-like image of an n-manifold. Frank Quinn discovered an integer-valued obstruction $i(X) \equiv 1 (\mathrm{mod}\, 8)$ and showed that the above question has a positive answer if and only if $i(X) = 1$ [203, 204]. For a while it was not known if there actually exist homology manifolds with $i(X) \neq 1$. The existence of such homology manifolds is a major recent achievement in topology, due to J. Bryant, S. Ferry, W. Mio and S. Weinberger [44].

A mapping $f : X \to Y$ is *cell-like* provided it is proper (counter-images of compact sets are compact) and all the fibers $f^{-1}(y)$, $y \in Y$, are *cell-like spaces*, i.e. have the shape of a point. Cell-like spaces and mappings were studied before the advent of shape theory. Note that a space X is cell-like if and only if every mapping $f : X \to P$ to an ANR P is homotopic to a constant mapping. A metric space X is cell-like if and only if for every embedding in an ANR M the following UV^∞ *property* holds: For every neighborhood U of X in M, there exists a neighborhood V of X such that $V \subseteq U$ and the inclusion $i : V \to U$ is nullhomotopic. A special case of cell-likeness is *cellularity* of sets in an n-manifold M, a notion introduced by Morton Brown in connection with the Schoenflies problem [43]. A subset X of an n-manifold M is cellular in M if there exists a sequence (B_i^n) of n-dimensional balls in M such that $B_{i+1}^n \subseteq \operatorname{Int} B_i^n$, for all i, and $X = \bigcap_i B_i^n$. Cellularity was studied extensively by D.R. McMillan, Jr. [183]. For a survey on cell-like mappings see [157]. Mappings between ANR's with AR-fibers as well as mappings satisfying the corresponding LC^n and n-contractibility conditions were studied already in [217].

Problems encountered in the research concerning infinite-dimensional manifolds, especially manifolds modelled on the Hilbert space l_2 and the Hilbert cube Q, were similar to problems encountered in the research concerning n-manifolds and progress in one area often stimulated progress in the other one. In many cases the infinite-dimensional problems turned out to be more accessible than the corresponding finite-dimensional problems and

R.H. Bing (1914–1986) was a student of the legendary topology teacher Robert Lee Moore (1882–1974) at the University of Texas in Austin. Bing obtained his Ph.D. in Austin in 1945. He did pioneering work concerning decomposition spaces and homeomorphisms in 3-dimensional manifolds [22]. The first systematic study of homology manifolds is due to another student of Moore, Raymond Louis Wilder (1896–1982) [247].

the solution of the former preceded the solution of the latter. The center of this research was the group around Richard Davis Anderson, Professor at the University of Louisiana in Baton Rouge.

R.D. Anderson was born in Hamden, Connecticut in 1922. He was a student of R.L. Moore at Austin, Texas, where he obtained his Ph.D. in 1948. One can associate with the Anderson group T.A. Chapman, D.W. Curtis, S. Ferry, R. Geoghegan, D.W. Henderson, R.M. Schori, J.E. West, R.Y.T. Wong.

The direct product of an n-manifold by the Hilbert space l_2 is obviously an l_2-manifold. In 1960 V. Klee asked the converse. Is every l_2-manifold homeomorphic to the product of an n-manifold with l_2? In 1961 in a surprising article Borsuk answered this question in the negative [32]. He also posed the following intriguing problems: Is it true that the cartesian product of a compact polyhedron (ANR) by Q is a Q-manifold? Is it true that every Q-manifold is homeomorphic to the product of a compact polyhedron by Q?

A very special case of the first problem, contributed by Borsuk to the Scottish book in 1938, asked whether the product of a triod with Q is homeomorphic to Q. It was answered affirmatively by Anderson in 1964. The first problem for (locally compact) polyhedra was answered affirmatively in 1970 by West [240]. In 1973 Chapman developed a procedure to perform surgery on infinite-dimensional manifolds, which enabled him to establish an infinite-dimensional version of the handle-straightening theorem of R.C. Kirby and L.C. Siebenmann [143]. This result was an essential ingredient in the proof of two important theorems of Chapman. The first one was the *triangulation theorem*, which answered affirmatively the second of the Borsuk problems [64]. The second one was an unexpected proof of the topological invariance of the Whitehead torsion, i.e. proof of the assertion that homeomorphisms between compact polyhedra are simple homotopy equivalences. The solution of this more than 20 years old finite-dimensional problem of Whitehead was a great achievement of infinite-dimensional topology. More precisely, Chapman proved that a mapping between compact polyhedra $f : X \to Y$ is a simple homotopy equivalence if $f \times \mathrm{id} : X \times Q \to Y \times Q$ is homotopic to a homeomorphism [65]. The converse implication was proved earlier by West [240]. Chapman also succeeded in extending the simple homotopy theory from compact polyhedra and CW-complexes to compact ANR's [67].

In 1973 Toruńczyk proved that the direct product of a compact AR with the Hilbert space l_2 is homeomorphic to l_2 [227]. Generalizations to products of ANR's with normed vector spaces were obtained in [228]. Finally, in 1975 R.D. Edwards proved that the product of a locally compact ANR with Q is a Q-manifold (see [66]). Combining Edwards'

ANR theorem with Chapman's triangulation theorem, one immediately concludes that every compact ANR has the homotopy type of a compact polyhedron P, which answers a classical problem stated by Borsuk at the International Congress of Mathematicians held in Amsterdam in 1954.

This problem was first solved by West. He proved that every locally compact ANR X is *resolvable*, i.e. it is the image of a Q-manifold M under a cell-like mapping $f : M \to X$ [241, 242]. Since cell-like mappings between locally compact ANR's are (fine) homotopy equivalences [124], X has the homotopy type of M. If X is compact, M is also compact and, by the triangulation theorem, M has the homotopy type of a compact polyhedron. The fact that $X \times [0, 1)$ is resolvable is often referred to as Miller's theorem. Actually, R.T. Miller proved the analogous assertion for finite-dimensional ANR's and finite-dimensional manifolds [186], but the arguments were applicable to the infinite-dimensional case as well. Note the difference of behavior between Q-manifolds and n-manifolds, exemplified by the resolvability of ANR's and the lack of resolvability of homology n-manifolds (which are finite-dimensional ANR's).

In Warsaw Toruńczyk proved a remarkable characterization of Q-manifolds as locally compact ANR's having the *disjoint n-cube property*, for all n [230]. He discovered this property independently of Cannon's discovery of the disjoint disc property [47]. Actually, his result preceded Cannon's work by a few months (see p. 291 of [114]). The preprints were widely disseminated already in the beginning of 1977. However, the paper appeared only in 1980, because of the long waiting time in Fundamenta Mathematicae at that time. The strategy of Toruńczyk's proof consisted in showing that the projection $X \times Q \to X$ (under the assumptions of the theorem) fulfills Bing's shrinking criterion, which yielded a homeomorphism $X \times Q \approx X$. However, by the Edwards ANR theorem, $X \times Q$ is a Q-manifold. Alternative proofs of Toruńczyk's theorem were obtained by Edwards [96] and later by J.J. Walsh [238]. These proofs use neither the West resolution theorem nor the Edwards ANR theorem. Instead they use Miller's theorem and the scheme used in proving the characterization theorem for finite-dimensional manifolds. Toruńczyk's characterization theorem for Q-manifolds implies the Edwards ANR theorem and many other results on Q-manifolds. In 1981 Toruńczyk characterized l_2-manifolds as ANR's having the discrete-cells property [231]. An alternative proof was given in [21]. Toruńczyk also considered the characterization of nonseparable Hilbert space manifolds and solved an old problem by proving that the weight of an infinite-dimensional Fréchet space determines its topological type.

There exist elementary examples of cell-like mappings $f : X \to Y$ between metric compacta, which are not homotopy equivalences. A much deeper fact is the existence of cell-like mappings which are not shape equivalences. The first such example was described by J.L. Taylor [224], who used sophisticated algebraic topology [1, 226]. In this example X is not an ANR and $Y = Q$. There exist similar examples, where $X = Q$ and Y is not an ANR [140]. At this point it was natural to ask whether the cell-like image of a compact finite-dimensional ANR must always be an ANR? It follows from a result of George Kozlowski [149] that this is equivalent to the following question. Must a cell-like image Y of a compact finite-dimensional ANR X be finite-dimensional? This problem proved to be very difficult and for a number of years defied the efforts of many topologists.

Finally, the problem was answered negatively. First it was proved that the following two problems are equivalent: (i) Does there exist a finite-dimensional metric compactum, which admits an infinite-dimensional cell-like image? (ii) Does there exist an infinite-dimensional

Thomas A. Chapman (born in 1940 in Mt. Hope, West Virginia) obtained his Ph.D. in 1970 at Louisiana State University from Anderson. Robert Duncan Edwards (born in 1942 in Freeport, New York) obtained his Ph.D. in 1969 from the University of Michigan under James Kister. Ross Geoghegan (born in 1943 in Dublin, Ireland) obtained his Ph.D. in 1968 at Cornell University from David Wilson Henderson. James Earl West (born in 1944 in Grinnell, Iowa) obtained his Ph.D. in 1967 at Louisiana State University from Anderson. Henryk Toruńczyk (born in 1945 in Warsaw) obtained his Ph.D. in 1971 in Warsaw from Czesław Bessaga. Steven Charles Ferry (born in 1947 in Takoma Park, Maryland) obtained his Ph.D. in 1973 from Morton Brown at the University of Michigan. John Joseph Walsh (born in 1948 in Helena, Montana) obtained his Ph.D. in 1973 at the State University of New York in Binghamton from Louis McAuley.

metric compactum X with finite (integral) cohomological dimension $\dim_{\mathbb{Z}} X < \infty$? The latter was a more than 50 years old unsolved problem of P.S. Aleksandrov. The equivalence of the two questions was announced in 1978 by R.D. Edwards in an abstract in the Notices of the American Mathematical Society [97]. In 1981 J.J. Walsh published a proof in [237] with acknowledgement to Edwards. A construction described in this proof proved to be very useful in cohomological dimension theory and is usually referred to as the Edwards–Walsh complex. In 1988 Aleksandr Nikolaevich Dranishnikov in Moscow [83, 84] (a student of E.V. Shchepin born in 1958) solved the Aleksandrov problem by producing an infinite-dimensional metric compactum X having $\dim_{\mathbb{Z}} X = 3$. He used the Edwards–Walsh complex and some sophisticated computations in reduced complex K-theory with mod p coefficients [4, 45]. It was then easy to obtain a cell-like mapping $f : S^7 \to Y$ with $\dim Y = \infty$.

An important strengthening of cell-like mappings are the *hereditary shape equivalences*, i.e. proper mappings $f : X \to Y$, which have the property that, for every closed subset $B \subseteq Y$, the restriction of f to $A = f^{-1}(B)$ is a shape equivalence $f|A : A \to B$. It was proved by Kozlowski [149] that the image of a compact ANR under a hereditary shape equivalence is always an ANR. Kozlowski's influential paper was never published. According to its author, the referee (Trans. Amer. Math. Soc.) required too many changes.

Research in the theory of retracts was also going on in Moscow, especially in Smirnov's seminar. Yu.M. Smirnov, a well-known general topologist, started his seminar in 1953. In the beginning it was devoted to general and infinite-dimensional topology. Later it included the theory of retracts and shape. Yu.T. Lisitsa, a member of Smirnov's seminar, successfully applied factorization techniques to problems concerning the extension of mappings. In particular, he obtained extension theorems for mappings into LC^∞-spaces, which are analogues of Dugundji's theorems for mappings into LC^n-spaces. Moreover, he showed that ANR's for metric spaces are always ANR's for the class of M-paracompact spaces, i.e. Hausdorff spaces, which admit perfect mappings onto metric spaces [166]. S.A. Bogatyĭ, another member of the seminar, studied various types of approximate retracts, especially from the point of view of shape theory [24]. Smirnov and his group devoted a number of papers to equivariant theory of retracts [218, 219, 6, 7, 170], a topic initially studied by J.W. Jaworowski [133, 134]. E.V. Shchepin obtained the surprising result that an ANR for the class of compact Hausdorff spaces must be either infinite-dimensional or metrizable [213]. The proof uses results on uncountable inverse systems of compacta, which he developed in his Ph.D. thesis.

Yuriĭ Mikhailovich Smirnov, professor at Moscow State University, was born in Kaluga in 1921. He began his studies at Moscow State University in 1939. He first belonged to the seminar of A.N. Kolmogorov (1903–1987). He became Alexandrov's student when, on Kolmogorov's recommendation, he was assigned to Aleksandrov to help him write his papers (Aleksandrov had a very poor eyesight). Smirnov's studies were interrupted by the second world war, which he spent in the navy. Returning from the war to the University, he defended his candidate's thesis in 1951 and his D.Sc. thesis in 1958. Yuriĭ Trofimovich Lisitsa was born near Bershad' in Ukraine in 1947. He defended his candidate's thesis in 1973 at Moscow State University. Eugeniĭ Vitalevich Shchepin was born in Moscow in 1951. He was the last student of Aleksandrov. At Moscow State University he defended the candidate's thesis in 1977 and the D.Sc. thesis in 1979.

Recent advances in cohomological dimension theory led to the formation of a new area of topology, called *extension theory*. According to a classical theorem on the (covering) dimension, $\dim X \leqslant n$ if and only if every mapping $f: A \to S^n$, defined on a closed subset A of X, extends to a mapping $\tilde{f}: X \to S^n$. Similarly, for the cohomological dimension with coefficients in G, one has $\dim_G X \leqslant n$ provided every mapping $f: A \to K(G,n)$ into the Eilenberg–MacLane complex $K(G,n)$ extends to a mapping $\tilde{f}: X \to K(G,n)$. More generally, in extension theory one considers the problem of extending mappings into metric simplicial complexes and CW-complexes. This unifies and generalizes the theories of covering and cohomological dimensions [92].

2. Theory of shape

It is generally considered that shape theory was founded in 1968, when Borsuk published his well-known paper on the homotopy properties of compacta [34]. Borsuk's starting point was the observation that many theorems in homotopy theory are valid only for spaces with good local behavior, e.g., manifolds, CW-complexes, ANR's, but fail when applied to spaces like metric compacta. A simple example of this phenomenon is the already mentioned Whitehead theorem that a weak homotopy equivalence between connected CW-complexes is a homotopy equivalence.

An example showing the failure of Whitehead's theorem for metric compacta is provided by the mapping $f: X \to Y$, where X is the *Warsaw circle* and $Y = \{*\}$ is a point. The Warsaw circle, an object popular in shape theory, is the planar continuum obtained from the closure of the graph of the function $\sin(1/t)$, $t \in (0, 1/\pi]$, by identifying the points $(0, 1)$ and $(1/\pi, 0)$. The mapping f is a weak homotopy equivalence, because all the homotopy groups of the Warsaw circle vanish. Nevertheless, f is not a homotopy equivalence.

To overcome such difficulties, caused by local irregularities of spaces, Borsuk considered metric compacta embedded in the Hilbert cube Q (more generally, in a fixed absolute retract). Instead of mappings $f: X \to Y$ between such compacta, he considered *fundamental sequences* $(f_n): X \to Y$, i.e. sequences of mappings $f_n: Q \to Q$, $n = 1, 2, \dots$, such that, for every neighborhood V of Y in Q, there exist a neighborhood U of X in Q and an integer m such that $f_n(U) \subseteq V$, for $n \geqslant m$. Moreover, the restrictions $f_n|U$ and $f_{n'}|U$ are homotopic in V, for $n, n' \geqslant m$. Fundamental sequences compose by composing their components, i.e. $(g_n)(f_n) = (g_n f_n)$. Two fundamental sequences (f_n), (f_n') are considered *homotopic* provided every V admits a U and an m such that $f_n|U \simeq f_n'|U$ in V, whenever $n \geqslant m$. Homotopy of fundamental sequences is an equivalence relation and the homotopy classes $[(f_n)]$ compose by composing their representatives, i.e. $[(g_n)][(f_n)] = [(g_n)(f_n)]$. In this way one obtains a category, whose objects are compacta in Q and the morphisms are homotopy classes of fundamental sequences. Since arbitrary metric compacta embed in Q, one readily extends this category to an equivalent category Sh(CM), whose objects are all metric compacta. This is *Borsuk's shape category.*

Every mapping $f : X \to Y$ induces a fundamental sequence, whose homotopy class depends only on the homotopy class of f. In this way one obtains a functor S: Ho(CM)$\to$ Sh(CM) from the homotopy category of metric compacta to Borsuk's shape category, called the *shape functor*. Compacta X, Y of the same homotopy type have the same shape, $\mathrm{sh}(X) = \mathrm{sh}(Y)$, i.e. are isomorphic objects of Sh(CM). Borsuk showed that, for a compact ANR Y, shape morphisms $F : X \to Y$ are in one-to-one correspondence with the homotopy classes of mappings $X \to Y$. Therefore, for compact ANR's, shape coincides with homotopy type. The Warsaw circle and the circle S^1 are examples of metric continua which have different homotopy types, but the same shape.

Borsuk's work on shape theory also had its precedents. These include cell-like spaces and cell-like mappings, i.e. *property* UV^∞, as well as its finite analogue, the *property* UV^n. They also include the Vietoris and the Čech homology (cohomology) groups [234, 57]. D.E. Christie's Ph.D. thesis, written in Princeton under Lefschetz's supervision, contains the beginnings of ordinary and strong shape theories [70]. In particular, Christie's homotopy groups coincide with Borsuk's shape groups. The 1-dimensional shape group was discovered even before [148]. The Brasilian topologist Elon L. Lima, a student of Edwin H. Spanier (1921–1996), generalized the Spanier–Whitehead duality to compact subsets of the sphere, by introducing a stable shape category [162]. However, in his paper no attempt was made to develop the shape category. Lima's work was "discovered" by the shape-theorists with considerable delay.

Undoubtedly, many topologists became aware of Borsuk's work on shape theory after he presented his ideas and results in Baton Rouge, Louisiana, in 1967, during a symposium on infinite dimensional topology (the proceedings were published only in 1972) and in Hercegnovi (former Yugoslavia) in 1968, during an international conference on topology. At the second of these events Borsuk used for the first time the suggestive term *shape* [35].

Shortly after Borsuk's talks and seminal papers on shape theory [34–38], an avalanche of articles on this new branch of topology appeared. By 1980 the literature on shape theory already consisted of about 400 papers. Around the world, groups of shape theorists were formed. Three specialized conferences, organized in Dubrovnik in 1976, 1981 and 1986 (Volumes 870 and 1283 of the Springer Lecture Notes in Mathematics) also contributed to the quick growth of shape theory.

In the initial period Warsaw was the center of activities in shape theory and the seat of the Borsuk group, which included J. Dydak, S. Godlewski, W. Holsztyński, A. Kadlof, J. Krasinkiewicz, Krystyna Kuperberg, P. Minc, Maria Moszyńska, S. Nowak, Hanna Patkowska, S. Spież, M. Strok, A. Trybulec.

In the US the first contributions to shape theory were made by Jack Segal, Professor at the University of Washington in Seattle (born in Philadelphia in 1934, Ph.D. in 1960 at the University of Georgia from M.K. Fort, Jr.), R.H. Fox (a well-known specialist in knot theory) and T.A. Chapman, Professor at the University of Kentucky. They were quickly joined by Billy Joe Ball (born in 1925, died in Austin, Texas in 1996) and R.B. Sher (born in Flint, Michigan in 1939) (Athens, Georgia), J.E. Keesling (born in 1942) and Philip Bacon (born in Chicago in 1929, died in Gainesville 1991) (Gainesville, Florida), R. Geoghegan and D.A. Edwards (Binghamton, New York), H.M. Hastings (Hempstead, New York), R.C. Lacher (Tallahassee, Florida), L.R. Rubin (Norman, Oklahoma), T.B. Rushing (born in Marshville, N. Carolina in 1941, died in Salt Lake City in 1998) (Salt Lake City, Utah), J.B. Quigley (Bloomington, Indiana), D.S. Coram and P. F. Duvall (Stillwater, Oklahoma), L.S. Husch (Knoxville, Tennessee), S. Ferry (Lexington, Kentucky), F.W. Cathey and G. Kozlowski (Seattle, Washington), G.A. Venema (Grand Rapids, Michigan) and many others.

In Moscow, since 1924, Aleksandrov conducted a seminar on topological spaces and dimension theory. Smirnov was a member of Aleksandrov's seminar and from 1953 to 1987 had his own seminar. Since 1970 the name of the seminar was *Seminar for shape theory and retracts*. Among the participants interested in shape theory and related areas were V.V. Agaronian, S. Antonian, S.A. Bogatyĭ, A.I. Bykov, A.Ch. Chigogidze, V.A. Kalinin, S.S. Kotanov, B.T. Levshenko, Yu.T. Lisitsa, I.S. Rubanov, A.P. Shostak, E.G. Sklyarenko, G. Skordev. Smirnov's group especially studied FANR's and related spaces as well as equivariant shape theory. In the Soviet Union contributions to shape theory were also made in Tbilisi, Georgia, by Z.R. Miminoshvili, a student of L.D. Mdzinarishvili (who in his turn was a student of G.S. Chogoshvili (1914–1998), the leading topologist in Georgia). Research in shape theory and related areas was also done in Novosibirsk by V.I. Kuz'minov, I.A. Shvedov, M.A. Batanin.

In Japan contributions to shape theory came from Kiiti Morita (1915–1995), the founder of general topology in Japan (dimension theory, product spaces) and from the group around Yukihiro Kodama at the University of Tsukuba. Kodama's group included H. Fukaishi, H. Hosokawa, H. Kato, K. Kawamura, A. Koyama, J. Ono, K. Sakai, K. Tsuda, T. Watanabe, T. Yagasaki, K. Yokoi.

The shape group in Zagreb (earlier Yugoslavia, now Croatia) was led by Sibe Mardešić (born in 1927 in Bergedorf near Hamburg, Germany). It included Z. Čerin, Q. Haxhibeqiri, K. Horvatić, I. Ivanšić, Vlasta Matijević, N. Šekutkovski, Š. Ungar, N. Uglešić. In Germany, shape theorists were led by Friedrich Wilhelm Bauer, Professor in Frankfurt a.M (born in Berlin in 1932). His group included B. Günther, P. Mrozik, H. Thiemann. In Great Britain the first contributions to shape theory were made by Timothy Porter, Professor at the University of Wales in Bangor (born in Abergavenny, Gwent in 1947). Further contributions were made by Allan Calder from Birckbeck College in London. In France shape

K. Morita was born in Hamamatsu-shi, Shizuoka. He studied at Tokyo Higher Normal School and Tokyo University of Science and Literature. He defended his Ph.D. thesis in 1950 at the University of Osaka. However, he was essentially a self-taught topologist. He was Professor at the Tokyo University of Education, which later became the University of Tsukuba. Morita also worked in algebra (module and ring theory). Y. Kodama was born in Tsuruoka in 1929. He obtained his B.Sci. from Tokyo University of Literature and Science in 1950 and his Ph.D. from Tokyo University of Education in 1960, under Morita. For his work in topology he was primarily inspired by studying the papers of Aleksandrov and Borsuk. He was Professor at Tsukuba University until his retirement in 1993.

S. Mardešić obtained his Ph.D. in 1957 from the University of Zagreb. He is essentially a self-taught topologist, influenced primarily by the work of Aleksandrov and Borsuk. F.W. Bauer obtained his Ph.D. in 1955 in Frankfurt a.M. In his work he was primarily influenced by W. Franz, P.S. Aleksandrov and S. MacLane, and considers himself a member of the Aleksandrov school. T. Porter obtained his Ph.D. from the University of Sussex in 1972. J.-M. Cordier obtained his doctorat d'état from University Paris 7 in 1987. J.M.R. Sanjurjo obtained his Ph.D. in Madrid in 1979 under the supervision of J.M. Montesinos. Being a knot-theorist, Montesinos came in touch with shape theory through Fox.

theory began with Jean-Marc Cordier and Dominique Bourn from the University of Picardie in Amiens. The Spanish shape group was led by José M.R. Sanjurjo, Professor at the Complutense University in Madrid (born in Madrid in 1951). His group included A. Giraldo, V.F. Laguna, M.A. Morón, F.R. Ruiz del Portal. Some shape theory was also done in Belgium (R.W. Kieboom), Canada (L. Demers), Italy (E. Giuli, L. Stramaccia, A. Tozzi), Mexico (Mónica Clapp, R. Jimenez, L. Montejano, Sylvia de Neymet), Romania (I. Pop), Switzerland (H. Kleisli, C. Weber).

Jack Segal spent the academic year 1969/70 in Zagreb. The result of this visit was joint work with Mardešić, generalizing Borsuk's shape theory to compact Hausdorff spaces [178, 179]. The new description of shape was based on a systematic use of inverse systems. Every compact Hausdorff space X can be represented as the inverse limit of a cofinite inverse system $\boldsymbol{X} = (X_\lambda, p_{\lambda\lambda'}, \Lambda)$ of compact polyhedra (or compact ANR's). Shape morphisms $F: X \to Y$ are given by *homotopy classes* of *homotopy mappings* $(f, f_\mu): \boldsymbol{X} \to \boldsymbol{Y} = (Y_\mu, p_{\mu\mu'}, M)$. The latter consist of an increasing function $f: M \to \Lambda$ and a family of mappings $f_\mu : X_{f(\mu)} \to Y_\mu$ such that, for $\mu_0 \leqslant \mu_1$, the following diagram commutes up to homotopy

$$\begin{array}{ccc} X_{f(\mu_0)} & \xleftarrow{\;p_{f(\mu_0)f(\mu_1)}\;} & X_{f(\mu_1)} \\ {\scriptstyle f_{\mu_0}}\downarrow & & \downarrow{\scriptstyle f_{\mu_1}} \\ Y_{\mu_0} & \xleftarrow[\;q_{\mu_0\mu_1}\;]{} & Y_{\mu_1} \end{array} . \tag{1}$$

Two homotopy mappings (f', f'_μ), (f'', f''_μ) are considered *homotopic* if there exists an increasing function $f \geqslant f', f''$ such that $f'_\mu p_{f'(\mu)f(\mu)} \simeq f''_\mu p_{f''(\mu)f(\mu)}$. Equivalence with the Borsuk approach was proved using inverse systems which consist of a decreasing sequence of compact ANR-neighborhoods of X in Q and of inclusion mappings.

While Borsuk's approach was rather geometric, the inverse system approach was more categorical and led quickly to further generalizations. In 1972 Fox generalized Borsuk's approach in a different direction, i.e. to arbitrary metric spaces X [109]. He embedded X as a closed subset in a suitable absolute retract L and used inclusion systems of ANR-neighborhoods of X in L. Both generalizations were unified in the papers by Mardešić [171] and K. Morita [191], where the general shape category Sh(Top) of arbitrary topological spaces was defined. Morita allows his systems $\boldsymbol{X}$ to be *homotopy systems*, i.e. the usual conditions on bonding mappings $p_{\lambda\lambda'}: X_{\lambda'} \to X_\lambda$ and projections $p_\lambda : X \to X_\lambda$ are replaced by homotopy conditions $p_{\lambda\lambda'}p_{\lambda'\lambda''} \simeq p_{\lambda\lambda''}$, $p_{\lambda\lambda'}p_{\lambda'} \simeq p_\lambda$, $\lambda \leqslant \lambda' \leqslant \lambda''$. Moreover, some theorems from [179] now became conditions (M1), (M2), which are part of the definition of a system being *associated* with a space:

(M1) For every mapping $f : X \to P$ to a polyhedron (or ANR) P, there exist a $\lambda \in \Lambda$ and a mapping $f_\lambda : X_\lambda \to P$ such that $f_\lambda p_\lambda \simeq f$.

(M2) For every $\lambda \in \Lambda$ and mappings $f_\lambda, f'_\lambda : X_\lambda \to P$ such that $f_\lambda p_\lambda \simeq f'_\lambda p_\lambda$, there exists an index $\lambda' \geqslant \lambda$ such that $f_\lambda p_{\lambda\lambda'} \simeq f'_\lambda p_{\lambda\lambda'}$.

Morita proved that the *Čech system,* formed by the nerves of all normal coverings of X, is a homotopy system associated with X [192]. In the terminology first used in algebraic geometry [116], shape morphisms are given by morphisms $\boldsymbol{X} \to \boldsymbol{Y}$ from the category pro-Ho(Top), where Ho(Top) denotes the homotopy category of topological spaces.

One of the first successful applications of shape theory is Fox's theory of *overlays,* a modification of covering spaces [109]. The classical theorem of covering space theory asserts that n-fold covering spaces of a connected arcwise locally connected and semilocally 1-connected space X are in a one-to-one correspondence with the classes of homomorphisms of the fundamental group $\pi_1(X)$ into the symmetric group Σ_n, where two homomorphisms ϕ, ψ belong to the same class provided there exists an inner automorphism $\theta : \Sigma_n \to \Sigma_n$ such that $\phi = \theta\psi$. Fox's shape theoretic version of the theorem, refers to overlays of arbitrary metric spaces X (embedded in some ANR). However, the fundamental group $\pi_1(X)$ has to be replaced by the *fundamental pro-group* $\pi_1(\boldsymbol{X}, *)$, the inverse system of fundamental groups of ANR-neighborhoods of X.

Further significant successes of shape theory were the shape-theoretic versions of the theorems of Whitehead, Hurewicz and Smale. The statements of these results also use pro-groups, i.e. inverse systems of groups. Application of the singular homology functor $H_m(\,.\,; G)$ to $\boldsymbol{X}$ yields an inverse system of Abelian groups $H_m(\boldsymbol{X}; G) = (H_m(X_\lambda; G), p_{\lambda\lambda'*}, \Lambda)$, called the m-th-*homology pro-group* of $\boldsymbol{X}$. Similarly, for systems of pointed spaces $(\boldsymbol{X}, *)$, one defines the m-th-*homotopy pro-group* $\pi_m(\boldsymbol{X}, *)$. If $\boldsymbol{X}$ and $(\boldsymbol{X}, *)$ are systems of ANR's associated with the space X and $(X, *)$, respectively, then these pro-groups do not depend on the choice of the associated systems. Moreover, they are shape invariants of X and $(X, *)$, respectively. The inverse limit $\check{H}_m(X; G) = \lim H_m(\boldsymbol{X}; G)$ is the *Čech homology group.* The *shape groups* $\check{\pi}_m(X, *) = \lim \pi_m(\boldsymbol{X}, *)$, were first defined in [70]. One should keep in mind that the Čech groups and the shape groups give less information about the space than the corresponding homology and homotopy pro-groups.

The most general version of the Whitehead theorem in shape theory is due to K. Morita [190]. It asserts that a morphism of pointed shape $F : (X, *) \to (Y, *)$ between finite-dimensional topological spaces is a *shape equivalence,* i.e. an isomorphism of *pointed shape* if and only if it induces isomorphisms of all homotopy pro-groups $F_\# : \pi_m(\boldsymbol{X}, *) \to \pi_m(\boldsymbol{Y}, *)$. In contrast to the classical Whitehead theorem, there are no restrictions on the local behavior of the spaces involved. Morita's result was preceded by less general versions of the theorem, obtained by Moszyńska [193] and Mardešić [172]. The restriction to finite dimensions cannot be omitted. A counterexample was obtained in [82], using a metric continuum defined by D.S. Kahn [135]. For every odd prime p, one considers the CW-complex X_0, obtained by attaching a $(2p+1)$-cell to S^{2p} by a mapping of degree p. One defines X_{n+1} as the $(2p-2)$-fold suspension $\Sigma^{2p-2}(X_n)$, $n \geqslant 0$. A particular mapping $f_1 : X_1 \to X_0$ is chosen. For $n > 1$, one defines mappings $f_n : X_n \to X_{n-1}$ by putting $f_n = \Sigma^{2p-2}(f_{n-1})$. The Kahn continuum is the limit of the inverse sequence defined

Stanisław Spież (born in Kalisz, Poland in 1944), Sławomir Nowak (born in Sosnowiec, Poland in 1946) and Jerzy Dydak (born in Brzozów, Poland in 1951) obtained their Ph.D. degrees from the University of Warsaw in 1973, 1973 and 1975, respectively. They were Borsuk's students. Dydak moved to US in 1982.

by the spaces X_n and by the mappings f_n. The crucial property that all the compositions $f_i \circ \cdots \circ f_j$, $i < j$, are essential mappings depends on deep results in homotopy theory [1, 226].

In the Whitehead theorem mentioned above the restriction to finite-dimensional spaces can be replaced by the weaker restriction to spaces of finite *shape dimension* sd (also called *fundamental dimension* and denoted by Fd). This is a numerical shape invariant introduced by Borsuk [36]. An extensive study of this notion was carried out by Polish topologists S. Nowak [196] and S. Spież [220, 221].

The shape-theoretic Hurewicz theorem involves homology pro-groups. One assumes that X is a $(n-1)$-*shape connected* space, $n \geqslant 2$, i.e. its homotopy pro-groups $\pi_m(\boldsymbol{X}, *)$ vanish, for $m \leqslant n-1$. One concludes that the corresponding homology pro-groups $H_m(\boldsymbol{X}; \mathbb{Z})$ vanish and there exists a natural isomorphism $\phi_n : \pi_n(\boldsymbol{X}, *) \to H_n(\boldsymbol{X}; \mathbb{Z})$ of the n-th-pro-groups. The general result is due to Morita [190]. Earlier versions involving shape groups were obtained by M. Artin and B. Mazur [10] and K. Kuperberg [152]. A Hurewicz theorem involving Steenrod homology is due to Y. Kodama and A. Koyama [146] and to Yu.T. Lisitsa [168].

The classical Smale theorem is the homotopy version of a theorem of Vietoris concerning cell-like mappings of compacta [217]. The shape-theoretic Smale theorem was proved by J. Dydak [89, 91] and asserts that, for metric compacta, every cell-like mapping induces isomorphisms of homotopy pro-groups $f_\# : \pi_n(\mathbf{X}, *) \to \pi_n(\boldsymbol{Y}, *)$, for all n and all basepoints. Consequently, if sd X, sd $Y < \infty$, the Whitehead theorem applies and f is a shape equivalence.

Among the most important contributions of Borsuk to shape theory is the introduction of two shape invariant classes of metric compacta, the *fundamental absolute neighborhood retracts* FANR's [36] and *movable compacta* [37]. X is an FANR provided, for any compact metric space Y containing X, there exist a closed neighborhood U of X in Y and a *shape retraction* $R : U \to X$, i.e. a shape morphism which is a left shape inverse of the inclusion mapping $i : X \to U$, $RS[i] = \mathrm{id}_X$. Clearly, every compact ANR is an FANR. Many results from the theory of retracts have their analogues in the theory of shape. For example, if X is *shape dominated* by X' (i.e. there exist shape morphisms $f : X \to X'$ and $g : X' \to X$ such that $gf = \mathrm{id}_X$) and X' is an FANR, then X is also an FANR. This implies that FANR's coincide with metric compacta which are shape dominated by compact polyhedra.

A compact space X, embedded in the Hilbert cube Q, is movable provided every neighborhood U of X in Q admits a neighborhood U' of X such that, for any neighborhood $U'' \subseteq U$ of X, there exists a homotopy $H : U' \times I \to U$ with $H(x, 0) = x$, $H(x, 1) \in U''$, for all $x \in U'$. In other words, sufficiently small neighborhoods of X can be deformed arbitrarily close to X. Borsuk proved that this remarkable property is a shape invariant. In a subsequent paper, he characterized FANR's by a similar property, called *strong movability* [38]. From its definition it is clear that FANR's are always movable. In fact, Borsuk introduced movability as a tool needed to detect that some compacta, e.g., the solenoids, are not FANR's. Borsuk also introduced the notion of *n-movability* and proved that LC^{n-1} compacta are always n-movable [39]. A compactum $X \subseteq Q$ is n-movable provided every neighborhood U of X in Q admits a neighborhood U' of X in Q such that, for any neighborhood $U'' \subseteq U$ of X, any compactum K of dimension $\dim K \leqslant n$ and any mapping $f : K \to U'$, there exists a mapping $g : K \to U''$, such that f and g are homotopic in U. Clearly, if a compactum X is n-movable and $\dim X \leqslant n$, then X is movable. The notion of n-movability was the beginning of *n-shape theory*, which was especially developed in the papers of A.Ch. Chigogidze [69]. The n-shape theory is an important tool in the theory of n-dimensional Menger manifolds, developed by M. Bestvina [19].

Further studies revealed the importance of *pointed* FANR's and *pointed movability*. For example, the union of two pointed FANR's, whose intersection is a pointed FANR, is again a pointed FANR [94]. The main protagonists of this research were D.A. Edwards, R. Geoghegan, H.M. Hastings, A. Heller and J. Dydak. It was shown in [99, 101] that connected pointed FANR's coincide with *stable continua*, i.e. continua having the shape of a polyhedron. In general one cannot achieve that this polyhedron be compact. This is because there exist noncompact polyhedra P, which are homotopy dominated by compact polyhedra, but do not have the homotopy type of a compact polyhedron [236]. Edwards and Geoghegan [100] defined a Wall obstruction $\sigma(X)$ for FANR's X and they showed that X has the shape of a compact polyhedron if and only if $\sigma(X) = 0$. Since $\sigma(X)$ is an element of the reduced projective class group $\tilde{K}^0(\check{\pi}_1(X, *))$ of the first shape group $\check{\pi}_1(X, *)$, this result linked shape theory to K-theory.

The question whether every FANR is a pointed FANR eluded the efforts of shape theorists for several years. Finally, in 1982, Hastings and Heller proved that this is always the case [122]. The crucial step in their proof is a purely homotopy theoretic result. This is the theorem that on a finite-dimensional polyhedron X every homotopy idempotent $f : X \to X$ splits, i.e. $f^2 \simeq f$ implies the existence of a space Y and of maps $u : Y \to X$, $v : X \to Y$, such that $vu \simeq 1_Y$, $uv \simeq f$. The proof uses nontrivial combinatorial group theory as well as the spectral sequence of a covering mapping. More precisely, it uses a particular group G and a particular homomorphism $\phi : G \to G$, which induces an unsplittable homotopy idempotent $f : K(G, 1) \to K(G, 1)$ of Eilenberg–Mac Lane complexes. It also uses the fact that the construction is universal in the sense that whenever $f' : X \to X$ is an unsplit homotopy idempotent, then there is an injection $G \to \pi_1(X)$, which is equivariant with respect to $f_\#$ and $f'_\#$. The group G itself has been considered before by R.J. Thompson (unpublished). Parts of the argument were discovered independently by P. Minc, by J. Dydak [90] and by P. Freyd and A. Heller (unpublished). The question whether movable continua are always pointed movable is still open.

For movable spaces various shape-theoretic results assume simpler form. For example, if $f : (X, *) \to (Y, *)$ is a pointed shape morphism between pointed movable metric continua, which induces isomorphisms of shape groups $f_\# : \check{\pi}_k(X, *) \to \check{\pi}_k(Y, *)$, for all k and if the spaces X, Y are finite-dimensional, then f is a pointed shape equivalence. This is a consequence of the shape-theoretic Whitehead theorem and the fact that such an f induces isomorphisms of homotopy pro-groups $\pi_k(\boldsymbol{X}, *) \to \pi_k(\boldsymbol{Y}, *)$ [141, 91].

In 1972 a new direction in shape theory was inaugurated by Chapman. He applied methods of infinite-dimensional topology to the study of shape of metric compacta [62]. More precisely, he considered compacta X which are *Z-embedded* in the Hilbert cube Q, i.e. have the property that there exist mappings $f : Q \to Q$, which are arbitrarily close to the identity but their image $f(Q)$ misses X. This condition, introduced by R.D. Anderson [5], implies tameness and unknottedness of compacta and proved to be fundamental in the development of the theory of Q-manifolds [66]. Chapman's *complement theorem* asserts that two compacta X, Y, embedded in Q as Z-sets, have the same shape if and only if their complements $Q\backslash X$, $Q\backslash Y$ are homeomorphic. Chapman also exhibited an isomorphism of categories $T : \mathcal{WP} \to \mathcal{S}$. The domain of T is the *weak proper homotopy category* of complements $M = Q\backslash X$ of Z-sets X of Q. Morphisms of $\mathcal{WP}$ are equivalence classes of proper mappings $f : M \to N = Q\backslash Y$. Two such mappings $f, g : M \to N$ are considered equivalent provided every compact set $B \subseteq N$ admits a compact set $A \subseteq M$ and a homotopy $H : M \times I \to N$ such that H connects f to g and $H((M\backslash A) \times I) \subseteq N\backslash B$. The codomain of T is the restriction of the shape category Sh(CM) to Z-sets X of Q. On objects $M = Q\backslash X$ of $\mathcal{WP}$ one has $T(M) = Q\backslash M = X$.

Subsequently, Chapman published a second paper, which contained a finite-dimensional complement theorem, i.e. a theorem where the ambient space was the Euclidean space [63]. This paper had a strong geometric flavor and immediately attracted the attention of a number of specialists in geometric topology, in particular in PL-topology, who produced a series of finite-dimensional complement theorems. In most of these theorems one assumes that X and Y are "nicely" embedded in the Euclidean space $\mathbb{R}^n$ and satisfy the appropriate dimensional conditions. The conclusion is that X and Y have the same shape if and only if their complements $R^n\backslash X$, $R^n\backslash Y$ are homeomorphic. The most general of the results obtained is the complement theorem from [132]. It assumes that X and Y are shape r-connected, $\operatorname{sd} X = \operatorname{sd} Y = k$, $n - k \geqslant 4$ and $n \geqslant \max\{5, 2k + 2 - r\}$. The "niceness" condition is the *inessential loops condition* ILC, introduced by G.A. Venema [233]. A compactum $X \subseteq \mathbb{R}^n$ satisfies ILC provided every open neighborhood U of X in $\mathbb{R}^n$ admits an open neighborhood V of X in U, such that each loop in $V\backslash X$, which is null-homotopic in V, is also null-homotopic in $U\backslash X$. This condition was preceded by McMillan's *cellularity criterion* CC [183]. Complement theorems in more general ambient spaces and different categories were studied extensively by P. Mrozik [195]

A compact metric space X *embeds up to shape* in a space Y provided Y contains a metric compactum X' such that $\operatorname{sh}(X) = \operatorname{sh}(X')$. L.S. Husch and I. Ivanšić obtained several interesting results concerning this notion. In particular, they showed that every r-shape connected and pointed $(r + 1)$-movable compactum X with $\operatorname{sd}(X) = k$, $k \geqslant 3$, embeds up to shape in $\mathbb{R}^{2k-r}$ [131].

Based on Quillen's homotopical algebra [202], Edwards and Hastings introduced a homotopy category of inverse systems, denoted by Ho(pro-Top). It is obtained from the category pro-Top by localization at level homotopy equivalences. Using this category instead of pro-Ho(Top), they defined a strong shape category SSh(CM) of compact metric spaces. Strong shape has distinct advantages over shape, e.g., Edwards and Hastings showed that the analogue of Chapman's category isomorphism theorem assumes a more natural form. It asserts the existence of an isomorphism $T : \mathcal{P} \to \mathcal{SS}$, between *proper homotopy category* $\mathcal{P}$ of complements $M = Q\backslash X$ of Z-sets X of Q and the restriction $\mathcal{SS}$ of the strong shape category SSh(CM) to Z-sets of Q [102, 147]. The strong shape category for metric compacta was first defined by J.B. Quigley, a student of J. Jaworowski at the University of Indiana in Bloomington [201].

Through efforts of various authors over several years, in particular, Porter [199], Bauer [15, 16], Calder and Hastings [46], Miminoshvili [189], Cathey and Segal [48], Lisitsa [167], Lisica and Mardešić [163, 164], Dydak and Nowak [93], Günther [117], David A. Edwards was born in 1946 in New York. He obtained his Ph.D. in 1971 from Columbia University in New York. Harold M. Hastings was born in 1946 in Dayton, Ohio. He obtained his Ph.D. in 1972 from Princeton University.

a *strong shape category* for topological spaces SSh(Top) was defined and so was a strong shape functor $\overline{S}$: Ho(Top) $\to$ SSh(Top). It is related to the shape functor S by a factorization $S = \overline{E}\,\overline{S}$, where $\overline{E}$: SSh(Top) $\to$ Sh(Top) is a functor which forgets part of the richer structure of strong shape.

In defining the strong shape category for arbitrary spaces, one needed a method of associating with any given space X a system of polyhedra (or ANR's) in the category Top. One way of doing this is provided by the Vietoris system [199, 118]. Another approach, used by Bauer, rigidifies a construction from [171] and associates with X a 2-category P_X. Its objects are mappings into polyhedra $g : X \to P$ and its 1-morphisms $g_1 \to g_2$ are given by a mapping $r : P_1 \to P_2$ and a homotopy ω, which connects rg_1 with g_2. The 2-morphisms are defined by homotopies of order 2. This approach was generalized to homotopies of arbitrarily high order (expressed in simplicial terms) by Günther [117].

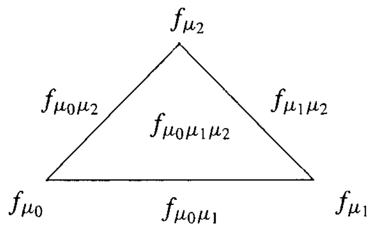
Another method is based on the notion of *resolution* of a space X [173] (more generally, on *strong expansions* [93, 117, 174]). A resolution $\mathbf{p} : X \to \mathbf{X}$ is a morphism of pro-Top which satisfies a stronger version of Morita's conditions.

(R1) Given a polyhedron P and an open covering $\mathcal{V}$ of P, any mapping $f : X \to P$ admits a $\lambda \in \Lambda$ and a mapping $h : X_\lambda \to P$ such that the mappings hp_λ and f are $\mathcal{V}$-near.

(R2) There exists an open covering $\mathcal{V}'$ of P, such that whenever, for a $\lambda \in \Lambda$ and two mappings $h, h' : X_\lambda \to P$, the mappings $hp_\lambda, h'p_\lambda$ are $\mathcal{V}'$-near, then there exists a $\lambda' \geqslant \lambda$ such that the mappings $hp_{\lambda\lambda'}, h'p_{\lambda\lambda'}$ are $\mathcal{V}$-near.

To define a strong shape morphism $F : X \to Y$, it suffices to choose (cofinite) polyhedral resolutions $\mathbf{p} : X \to \mathbf{Y}$, $\mathbf{q} : Y \to \mathbf{Y}$ and a morphism $\mathbf{X} \to \mathbf{Y}$ of Ho(pro-Top).

It is an important fact that the category Ho(pro-Top) is equivalent to the coherent homotopy category CH(Top), which can be viewed as a concrete realization of the former category [163, 164]. Its morphisms are coherent homotopy classes of coherent mappings $\mathbf{f} : \mathbf{X} \to \mathbf{Y}$. The latter consist of an increasing function $f : M \to \Lambda$ and of mappings $f_{\mu_0} : X_{f(\mu_0)} \to Y_{\mu_0}$, which make diagram (1) commutative up to a homotopy $f_{\mu_0\mu_1} : X_{f(\mu_1)} \times I \to Y_{\mu_0}$, which is also part of the structure of $\mathbf{f}$. For three indices $\mu_0 \leqslant \mu_1 \leqslant \mu_2$, one has homotopies $f_{\mu_0\mu_1\mu_2} : X_{f(\mu_2)} \times \Delta^2 \to Y_{\mu_0}$, where Δ^2 is the standard 2-simplex. One requires that, on the faces of Δ^2, $f_{\mu_0\mu_1\mu_2}$ is given by the mappings $f_{\mu_1\mu_2}, f_{\mu_0\mu_2}, f_{\mu_0\mu_1}$ as indicated on the following figure.



Analogous requirements are imposed on higher homotopies $f_{\mu_0 \dots \mu_n} : X_{f(\mu_n)} \times \Delta^n \to Y_{\mu_0}$, for all increasing sequences $\mu_0 \leqslant \cdots \leqslant \mu_n$ and all n. There are other, more sophisticated descriptions of coherent categories, due to J.M. Boardman and R.M. Vogt [23, 235], Cordier and Porter [72, 73], N. Šekutkovski [211], Batanin [14], but they all yield categories equivalent to CH(Top).

An important circle of ideas, related to strong shape, refers to *strong* or *Steenrod homology*. It was originally defined only for metric compacta [222]. Over the years, especially in former USSR, much work was done on strong homology of general spaces [215, 216]. The relation of strong homology to singular and Čech homology is similar to the relation of strong shape to homotopy and ordinary shape. For pairs of spaces (X, A), where A is normally embedded in X (e.g., if A is closed and X is paracompact), all the Eilenberg–Steenrod axioms are fulfilled. From the point of view of shape theory, the most important property of strong homology is its invariance with respect to strong shape [163, 165]. In contrast to Čech homology, Čech cohomology has a long record of successful applications. The explication lies in the fact that direct limit is an exact functor, while inverse limit is not, i.e. in general, the derived functors $\lim^n$ of lim are nontrivial. The higher limits $\lim^n H_m(X; \mathbb{Z})$ of the homology pro-groups play an important role in strong homology of spaces. Actually, there exist paracompact spaces X with $\lim^n H_m(X; \mathbb{Z}) \neq 0$, for n arbitrarily high [175]. However, if X is compact, $\lim^n H_m(X; \mathbb{Z}) = 0$, for $n \geqslant 2$ [156, 176].

Using a suitable *approximate homotopy lifting property*, D.S. Coram and P.F. Duvall have introduced *approximate fibrations* as mappings $f : X \to Y$ between ANR's, which generalize cell-like mappings and share many homotopy-theoretic properties with fibrations [71]. This class of mappings proved very useful in the study of mappings between manifolds. For mappings between metric compacta, approximate fibrations had to be replaced by *shape fibrations* [177, 250]. The definition of a shape fibration between arbitrary spaces required the notion of resolution of a mapping [173]. A very useful generalization of the latter notion was introduced by T. Watanabe, who introduced *approximate resolutions* of mappings [239]. Subsequently, a more general theory was developed in [181].

Appropriate variations of the basic ideas of shape led to new types of shape theories. In particular, there is *fibered shape* [138, 249], *equivariant shape* [8, 61], *stable shape* [197, 18], *proper shape* [12, 11], *uniform shape* [209, 188].

Generally, one expects to find applications of shape theory in problems concerning global properties of spaces having irregular local behavior. Such spaces naturally appear in many areas of mathematics. A typical example is provided by the fibers of a mapping as in the case of cell-like mappings. Other examples are given by remainders of compactifications, by sets of fixed points, by attractors of dynamical systems and by spectra of operators. In the latter, strong extraordinary homology plays and important role [137, 136, 17, 76].

Keesling has devoted a series of papers to the study of the remainder $\beta X \setminus X$ of a locally compact space X in its Čech–Stone compactification [142]. In this research he used his earlier results concerning the Čech cohomology groups of movable spaces. Recently, shape theory found applications also in the field of geometric group theory. More precisely, the boundary ∂G of a (discrete) group G is defined as a Z – set of a finite-dimensional compact AR $\tilde{X}$, such that the following two axioms hold: (i) $X = \tilde{X} \setminus Z$ admits a covering space action of G with compact quotient; (ii) The collection of translates of a compact set in X forms a null-sequence in $\tilde{X}$, i.e. for every open covering $\mathcal{U}$ of $\tilde{X}$ all but finitely many translates are $\mathcal{U}$-small. The boundary ∂G is determined up to shape, i.e. if Z_1 and Z_2 satisfy the above axioms, then $\mathrm{sh}(Z_1) = \mathrm{sh}(Z_2)$ [20].

Shape theory also led to new developments in the fixed point theorems. For every compact ANR X and every mapping $f : X \to X$, the Lefschetz number $\Lambda(f)$ is a well-defined integer. If $\Lambda(f) \neq 0$, then f has a fixed point. This well-known theorem is not true for arbitrary metric compacta, because for acyclic continua $\Lambda(f) = 1$ and they need not have the fixed point property. Nevertheless, Borsuk proved that, for an arbitrary metric compactum X, $\Lambda(f) \neq 0$ implies the existence of fixed points, provided f belongs to a certain class of mappings, called *nearly extendible mappings* [40]. Another new result asserts that the space 2^X of nonempty compacta and the space $C(X)$ of nonempty continua in a locally connected Hausdorff continuum X have the fixed point property [210]. This was known before only for Peano continua X.

As an example of application of shape theory in dynamical systems we state the following result. A finite-dimensional metric compactum embeds in a (differentiable) manifold M as an attractor of a (smooth) dynamical system on M if and only if it has the shape of a compact polyhedron [119, 208]. Another application of shape concerns the definition of the Conley index for continuous and discrete dynamical systems [205].

Shape theory has also applications in the theory of continua. For example, joinable continua were characterized as pointed 1-movable continua [150]. H. Kato successfully applied shape theory to the study of Whitney mappings of hyperspaces 2^X and $C(X)$ [139].

There are many situations, where shape itself does not apply, but its methods do. Typical examples are properties at infinity of locally compact spaces (see [58]) and proper homotopy (see [185, 200, 59]). Ideas of shape theory had a bearing on homology of groups [115]. The abstract aspects of shape led to *categorical shape theory* [75] and opened further possibilities of application, e.g., in pattern recognition [198, 74].

The approach to shape by inverse systems of polyhedra or ANR's is not the only one. A different approach, recently inaugurated by Sanjurjo [207] and further developed by Čerin [60] is based on the idea of replacing mappings by multivalued mappings, which map points into sufficiently small sets.

Acknowledgement

In writing this paper I consulted many colleagues who provided valuable information and suggestions. It is a pleasure to express to all of them my sincere gratitude.

Bibliography

[1] J.F. Adams, *On the groups $J(X)$ – IV*, Topology **5** (1966), 21–71.

[2] P.S. Alexandroff, *Simpliziale Approximationen in der allgemeinen Topologie*, Math. Ann. **96** (1926), 489–511.

[3] P.S. Alexandroff, *Une définition des nombres de Betti pour un ensemble fermé quelconque*, C. R. Acad. Sci. Paris **184** (1927), 317–320.

[4] D.W. Anderson and L. Hodgkin, *The K-theory of Eilenberg–MacLane complexes*, Topology **7** (1968), 317–329.

[5] R.D. Anderson, *On topological infinite deficiency*, Michigan J. Math. **14** (1967), 365–383.

[6] S.A. Antonian, *An equivariant theory of retracts*, Aspects of Topology, London Math. Soc. Lecture Notes vol. 93, Cambridge Univ. Press, Cambridge (1985), 251–269.

[7] S.A. Antonian, *Retraction properties of the space of orbits*, Mat. Sb. **137 (179)** (3) (1988), 300–318 (in Russian).
[8] S.A. Antonian and S. Mardešić, *Equivariant shape*, Fund. Math. **127** (1987), 213–224.
[9] N. Aronszajn and K. Borsuk, *Sur la somme et le produit combinatoire des rétractes absolus*, Fund. Math. **18** (1932), 193–197.
[10] M. Artin and B. Mazur, *Etale homotopy*, Lecture Notes in Math. vol. 100, Springer, Berlin (1969).
[11] B.J. Ball, *Alternative approaches to proper shape theory*, Studies in Topology, Charlotte Conference, 1974, Academic Press, New York (1975), 1–27.
[12] B.J. Ball and R.B. Sher, *A theory of proper shape for locally compact metric spaces*, Fund Math. **86** (1974), 163–192.
[13] M.G. Barratt, *Simplicial and semisimplicial complexes*, Mimeographed Lecture Notes, Princeton University, Princeton (1956).
[14] M.A. Batanin, *Coherent categories with respect to monads and coherent prohomotopy theory*, Cahiers Topol. Géom. Diff. Categor. **34** (4) (1993), 279–304.
[15] F.W. Bauer, *A shape theory with singular homology*, Pacific J. Math. **64** (1976), 25–65.
[16] F.W. Bauer, *Duality in manifolds*, Annali Mat. Pura Appl. **136** (4) (1984), 241–302.
[17] F.W. Bauer, *Extensions of generalized homology theories*, Pacific. J. Math. **128** (1987), 25–61.
[18] F.W. Bauer, *A strong shape theory admitting an S-dual*, Topology Appl. **62** (1995), 207–232.
[19] M. Bestvina, *Characterizing k-dimensional universal Menger compacta*, Mem. Amer. Math. Soc. **71** (380) (1988), 1–110.
[20] M. Bestvina, *Local homology properties of boundaries of groups*, Mich. Math. J. **43** (1) (1996), 123–139.
[21] M. Bestvina, P. Bowers, J. Mogilski and J. Walsh, *Characterization of Hilbert space manifolds revisited*, Topology Appl. **24** (1986), 53–69.
[22] R.H. Bing, *The geometric topology of 3-manifolds*, Amer. Math. Soc. Colloq. Publ. vol. 40, Providence, RI (1983).
[23] J.M. Boardman and R.M. Vogt, *Homotopy invariant algebraic structures on topological spaces*, Lecture Notes in Math. vol. 347, Springer, Berlin (1973).
[24] S.A. Bogatyĭ, *Approximate and fundamental retracts*, Mat. Sbornik **93** (135) (1974), 90–102 (in Russian). (Math. USSR Sbornik **22** (1974), 91–103.)
[25] W.S. Bogomolowa, *On a class of functions everywhere asymptotically continuous*, Mat. Sb. **32** (1924), 152–171 (in Russian).
[26] C.J.R. Borges, *On stratifiable spaces*, Pacific J. Math. **17** (1966), 1–16.
[27] K. Borsuk, *Sur les rétracts*, Fund. Math. **17** (1931), 152–170.
[28] K. Borsuk, *Über eine Klasse von zusammenhängenden Räumen*, Fund. Math. **19** (1932), 220–241.
[29] K. Borsuk, *Sur un continu acyclic qui se laisse transformer topologiquement en lui même sans points invariants*, Fund. Math. **24** (1935), 51–58.
[30] K. Borsuk, *Sur les prolongements des transformations continues*, Fund. Math. **28** (1937), 99–110.
[31] K. Borsuk, *Sur un espace compact localement contractile qui n'est pas un rétracte absolu de voisinage*, Fund. Math. **35** (1948), 175–180.
[32] K. Borsuk, *On a problem of V. Klee concerning the Hilbert manifolds*, Colloq. Math. **8** (1961), 239–242.
[33] K. Borsuk, *Theory of Retracts*, Polish Scientific Publishers, Warszawa (1967).
[34] K. Borsuk, *Concerning homotopy properties of compacta*, Fund. Math. **62** (1968), 223–254.
[35] K. Borsuk, *Concerning the notion of the shape of compacta*, Proc. Internat. Symp. Topology and its App., Herceg-Novi 1968, Đ. Kurepa, ed., Savez Društava Mat. Fiz. Astronom., Beograd (1969), 98–104.
[36] K. Borsuk, *Fundamental retracts and extensions of fundamental sequences*, Fund. Math. **64** (1969), 55–85.
[37] K. Borsuk, *On movable compacta*, Fund. Math. **66** (1969), 137–146.
[38] K. Borsuk, *A note on the theory of shape of compacta*, Fund. Math. **67** (1970), 265–278.
[39] K. Borsuk, *On the n-movability*, Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys. **20** (1972), 859–864.
[40] K. Borsuk, *On the Lefschetz–Hopf fixed point theorem for nearly extendable maps*, Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys. **23** (1975), 1273–1279.
[41] K. Borsuk, *Theory of Shape*, Polish Scientific Publishers, Warszawa (1975).
[42] K. Borsuk and S. Mazurkiewicz, *Sur les rétractes absolus indécomposables*, C. R. Acad. Sci. Paris **199** (1934), 110–112.
[43] M. Brown, *A proof of the generalized Schoenflies theorem*, Bull. Amer. Math. Soc. **66** (1960), 74–76.

[44] J. Bryant, S. Ferry, W. Mio and S. Weinberger, *Topology of homology manifolds*, Ann. of Math. **143** (2) (1996), 435–467.
[45] V.M. Bukhshtaber and A.S. Mishchenko, *K-theory for the category of infinite cellular complexes*, Izv. Akad. Nauk SSSR, Ser. Mat. **32** (3) (1968), 560–604 (in Russian). (Mathematics of the USSR Izvjestija **2** (3) (1969), 515–556.)
[46] A. Calder and H.M. Hastings, *Realizing strong shape equivalences*, J. Pure Appl. Algebra **20** (1981), 129–156.
[47] J.W. Cannon, *The recognition problem: What is a topological manifold?*, Bull. Amer. Mat. Soc. **84** (1978), 832–866.
[48] F.W. Cathey and J. Segal, *Strong shape theory and resolutions*, Topology Appl. **15** (1983), 119–130.
[49] R. Cauty, *Sur le prolongement des fonctions continues a valeurs dans les CW-complexes*, C. R. Acad. Sci. Paris, Ser. A **273** (1971), 1208–1211.
[50] R. Cauty, *Sur les sous-espaces des complexes simpliciaux*, Bull. Soc. Math. France **100** (1972), 129–155.
[51] R. Cauty, *Convexité topologique et prolongement des fonctions continues*, Compositio Math. **27** (1973), 233–273.
[52] R. Cauty, *Sur le prolongement des fonctions continues dans les complexes simpliciaux infinis*, Fund. Math. **134** (1990), 221–245.
[53] R. Cauty, *Sur les ouverts des CW-complexes et les fibrés de Serre*, Colloq. Math. **63** (1992), 1–7.
[54] R. Cauty, *Une caractérisation des rétractes absolus de voisinage*, Fund. Math. **144** (1994), 11–22.
[55] R. Cauty, *Un espace métrique linéaire qui n'est pas un rétract absolu*, Fund. Math. **146** (1994), 85–99.
[56] J.G. Ceder, *Some generalizations of metric spaces*, Pacific J. Math. **11** (1961), 105–126.
[57] E. Čech, *Théorie générale de l'homologie dans un espace quelconque*, Fund. Math. **19** (1932), 149–183.
[58] Z. Čerin, *Homotopy properties of locally compact spaces at infinity – calmness and smoothness*, Pacific J. Math. **79** (1979), 69–91.
[59] Z. Čerin, *On various relative proper homotopy groups*, Tsukuba J. Math. **4** (1980), 177–202.
[60] Z. Čerin, *Shape via multi-nets*, Tsukuba J. Math. **19** (1995), 245–268.
[61] Z. Čerin, *Equivariant shape theory*, Math. Proc. Cambridge Phil. Soc. **117** (1995), 303–320.
[62] T.A. Chapman, *On some applications of infinite-dimensional manifolds to the theory of shape*, Fund. Math. **76** (1972), 181–193.
[63] T.A. Chapman, *Shapes of finite-dimensional compacta*, Fund. Math. **76** (1972), 261–276.
[64] T.A. Chapman, *Compact Hilbert cube manifolds and the invariance of Whitehead torsion*, Bull. Amer. Math. Soc. **79** (1973), 52–56.
[65] T.A. Chapman, *Topological invariance of Whitehead torsion*, Amer. J. Math. **96** (1974), 488–497.
[66] T.A. Chapman, *Lectures on Hilbert cube manifolds*, CBMS Reg. Series in Math. vol. 28, Amer. Math. Soc., Providence, RI (1976).
[67] T.A. Chapman, Simple homotopy theory for ANR's, General Topology Appl. **7** (1977), 165–174.
[68] A.V. Chernavskiĭ, *Local contractibility of the homeomorphism group of a manifold*, Mat. Sb. **79** (121) (1969), 307–356 (in Russian).
[69] A.Ch. Chigogidze, *Theory of n-shape*, Uspehi Mat. Nauk **44** (5) (1989), 117–140 (in Russian).
[70] D.E. Christie, *Net homotopy for compacta*, Trans. Amer. Math. Soc. **56** (1944), 275–308.
[71] D.S. Coram and P.F. Duvall, *Approximate fibrations*, Rocky Mountain J. Math. **7** (1977), 275–288.
[72] J.M. Cordier, *Sur la notion de diagramme homotopiquement cohérent*, Cahiers Topol. Géom. Diff. **23** (1982), 93–112.
[73] J.-M. Cordier and T. Porter, *Vogt's theorem on categories of homotopy coherent diagrams*, Math. Proc. Cambridge Phil. Soc. **100** (1986), 65–90.
[74] J.-M. Cordier and T. Porter, *Pattern recognition and categorical shape theory*, Pattern Recognition Letters **7** (1988), 73–76.
[75] J.-M. Cordier and T. Porter, *Shape Theory – Categorical Methods of Approximation*, Ellis Horwood, Chichester (1989).
[76] M. Dadarlat, *Shape theory and asymptotic morphisms for C^*-algebras*, Duke Math. J. **73** (1993), 687–711.
[77] R.J. Daverman, *Decompositions of Manifolds*, Academic Press, New York (1986).
[78] E.K. van Douwen and R. Pol, *Countable spaces without extension properties*, Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys. **25** (1977), 987–991.
[79] C.H. Dowker, *Retracts of metric spaces*, Bull. Amer. Math. Soc. **54** (1948), 645.
[80] C.H. Dowker, *Topology of metric complexes*, Amer. J. Math. **74** (1952), 555–577.
[81] C.H. Dowker, *Homotopy extension theorems*, Proc. London Math. Soc. **6** (1956), 100–116.

[82] J. Draper and J.E. Keesling, *An example concerning the Whitehead theorem in shape theory*, Fund. Math. **92** (1976), 255–259.

[83] A.N. Dranishnikov, *On a problem of P.S. Aleksandrov*, Mat. Sb. **135** (1988), 551–557 (in Russian).

[84] A.N. Dranishnikov, *Homological dimension theory*, Uspehi Mat. Nauk **43** (4) (1988), 11–55 (in Russian).

[85] J. Dugundji, *An extension of Tietze's theorem*, Pacific J. Math. **1** (1951), 353–367.

[86] J. Dugundji, *Note on CW polytopes*, Portugaliae Math. **11** (1952), 7–10.

[87] J. Dugundji, *Absolute neighborhood retracts and local connectedness of arbitrary metric spaces*, Compositio Math. **13** (1958), 229–246.

[88] J. Dugundji, *Locally equiconnected spaces and absolute neighborhood retracts*, Fund. Math. **57** (1965), 187–193.

[89] J. Dydak, *Some remarks on the shape of decomposition spaces*, Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys. **23** (1975), 561–563.

[90] J. Dydak, *A simple proof that pointed FANR-spaces are regular fundamental retracts of ANR's*, Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys. **25** (1977), 55–62.

[91] J. Dydak, *The Whitehead and the Smale theorems in shape theory*, Dissertationes Math. **156** (1979), 1–55.

[92] J. Dydak, *Extension theory: The interface between set-theoretic and algebraic topology*, Topology Appl. **74** (1996), 225–258.

[93] J. Dydak and S. Nowak, *Strong shape for topological spaces*, Trans. Amer. Math. Soc. **323** (1991), 765–796.

[94] J. Dydak, S. Nowak and S. Strok, *On the union of two FANR-sets*, Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys. **24** (1976), 485–489.

[95] J. Dydak and J. Segal, *Shape theory. An introduction*, Lecture Notes in Math. vol. 688, Springer, Berlin (1978).

[96] R.D. Edwards, *Characterizing infinite dimensional manifolds (after Henryk Toruńczyk)*, Séminaire Bourbaki 1978/79, exposés 525–542, Lecture Notes in Math. vol. 770, Berlin (1980), 278–302.

[97] R.D. Edwards, *A theorem and a question related to cohomological dimension and cell-like maps*, Notices Amer. Math. Soc. **25** (1978), A-259.

[98] R.D. Edwards, *The topology of manifolds and cell-like maps*, Proc. ICM, Helsinki 1978, Acad. Sci. Fennica, Helsinki (1980), 11–127.

[99] D.A. Edwards and R. Geoghegan, *The stability problem in shape and a Whitehead theorem in pro-homotopy*, Trans. Amer. Math. Soc. **214** (1975), 261–277.

[100] D.A. Edwards and R. Geoghegan, *Shapes of complexes, ends of manifolds, homotopy limits and the Wall obstruction*, Ann. Math. **101** (1975), 521–535. *Correction* **104** (1976), 389.

[101] D.A. Edwards and R. Geoghegan, *Stability theorems in shape and pro-homotopy*, Trans. Amer. Math. Soc. **222** (1976), 389–403.

[102] D.A. Edwards and H.M. Hastings, *Čech and Steenrod homotopy theories with applications to geometric topology*, Lecture Notes in Math. vol. 542, Springer, Berlin (1976).

[103] D.A. Edwards and H.M. Hastings, *Čech theory: its past, present and future*, Rocky Mountain J. Math. **10** (1980), 429–468.

[104] R.D. Edwards and R.C. Kirby, *Deformations of spaces of imbeddings*, Ann. Math. **93** (1971), 63–88.

[105] S. Ferry, *The homeomorphism group of a compact Hilbert cube manifold is an ANR*, Ann. Math. **106** (1977), 101–119.

[106] V.V. Fedorchuk and A.Ch. Chigogidze, *Absolute Retracts and Infinite-Dimensional Manifolds*, Nauka, Moskva (1992) (in Russian).

[107] R.H. Fox, *A characterization of absolute neighborhood retracts*, Bull. Amer. Math. Soc. **48** (1942), 271–275.

[108] R.H. Fox, *On fiber spaces II*, Bull. Amer. Math. Soc. **49** (1943), 733–735.

[109] R.H. Fox, *On shape*, Fund. Math. **74** (1972), 47–71.

[110] R. Fritsch, *Zur Unterteilung semisimplizialer Mengen. I*, Math. Z. **108**, 329–367; II, Math. Z. **109** (1969), 131–152.

[111] R. Fritsch and R.A. Piccinini, *Cellular Structures in Topology*, Cambridge Univ. Press, Cambridge (1990).

[112] R. Fritsch and D. Puppe, *Die Homöomorphie der geometrischen Realisierung einer semisimplizialen Menge und ihrer Normalunterteilung*, Arch. Math. **18** (1967), 508–512.

[113] R. Geoghegan, *On spaces of homeomorphisms, embeddings and functions II*, Proc. London Math. Soc. **27** (1973), 463–483.

[114] R. Geoghegan, *Open problems in infinite-dimensional topology*, Topology Proc. **4** (1979), 287–338.

[115] R. Geoghegan, *The shape of a group – connections between shape theory and the homology of groups*, Geometric and Algebraic Topology, Banach Center Publ. vol. 18, PWN, Warsaw (1986), 271–280.

[116] A. Grothendieck, *Technique de descentes et théorèmes d'existence en géométrie algébrique II*, Séminaire Bourbaki, **12** ème année, 1959/60, exposé 190–195.

[117] B. Günther, *The use of semisimplicial complexes in strong shape theory*, Glasnik Mat. **27** (47) (1992), 101–144.

[118] B. Günther, *The Vietoris system in strong shape and strong homology*, Fund. Math. **141** (1992), 147–168.

[119] B. Günther and J. Segal, *Every attractor of a flow on a manifold has the shape of a finite polyhedron*, Proc. Amer. Math. Soc. **119** (1993), 321–329.

[120] O. Hanner, *Some theorems on absolute neighborhood retracts*, Arkiv Mat. **1** (1951), 389–408.

[121] O. Hanner, *Retraction and extension of mappings of metric and nonmetric spaces*, Arkiv Mat. **2** (1952), 315–360.

[122] H.M. Hastings and A. Heller, *Homotopy idempotents on finite-dimensional complexes split*, Proc. Amer. Math. Soc. **85** (1982), 619–622.

[123] W.E. Haver, *Locally contractible spaces that are absolute neighborhood retracts*, Proc. Amer. Math. Soc. **40** (1973), 280–284.

[124] W.E. Haver, *Mapping between ANR's that are fine homotopy equivalences*, Pacific J. Math. **58** (1975), 457–461.

[125] C.J. Himmelberg, *Some theorems on equiconnected and locally equiconnected spaces*, Trans. Amer. Math. Soc. **115** (1965), 43–53.

[126] S.-T. Hu, *A new generalization of Borsuk's theory of retracts*, Indag. Math. **9** (1947), 465–469.

[127] S.-T. Hu, Mappings of a normal space into an absolute neighborhood retract, Trans. Amer. Math. Soc. **64** (1948), 336–358.

[128] S.-T. Hu, *Theory of Retracts*, Wayne State Univ. Press, Detroit (1965).

[129] W. Hurewicz, *Beiträge zur Theorie der Deformation II*, Proc. Kon. Akad. Wet. Amst. **38** (1935), 521–528.

[130] W. Hurewicz, *On the concept of fiber space*, Proc. Nat. Acad. Sci. USA **27** (1955), 956–961.

[131] L.S. Husch and I. Ivanšić, *Embedding compacta up to shape*, Shape Theory and Geometric Topology, Lecture Notes in Math. vol. 870, Springer, Berlin (1981), 119–134.

[132] I. Ivanšić, R.B. Sher and G.A. Venema, *Complement theorems beyond the trivial range*, Illinois J. Math. **25** (1981), 209–220.

[133] J.W. Jaworowski, *Equivariant extensions of maps*, Pacific J. Math. **45** (1973), 229–244.

[134] J.W. Jaworowski, *Extensions of G-maps and euclidean G-retracts*, Math. Z. **146** (1976), 143–148.

[135] D.S. Kahn, *An example in Čech cohomology*, Proc. Amer. Math. Soc. **16** (1965), 584.

[136] D.S. Kahn, J. Kaminker and C. Schochet, *Generalized homology theories on compact metric spaces*, Michigan Math. J. **24** (1977), 203–224.

[137] J. Kaminker and C. Schochet, *K-theory and Steenrod homology; Applications to the Brown–Douglas–Fillmore theory of operator algebras*, Trans. Amer. Math. Soc. **227** (1977), 63–107.

[138] H. Kato, *Fiber shape categories*, Tsukuba J. Math. **5** (1981), 247–265.

[139] H. Kato, *Shape properties of Whitney maps for hyperspaces*, Trans. Amer. Math. Soc. **297** (1986), 529–546.

[140] J.E. Keesling, *A nonmovable trivial-shape decomposition of the Hilbert cube*, Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys. **23** (1975), 997–998.

[141] J.E. Keesling, *On the Whitehead theorem in shape theory*, Fund. Math. **92** (1976), 247–253.

[142] J.E. Keesling, *Shape theory and the Stone–Čech compactification*, Proc. Intern. Conf. on Geom. Top., Warszawa, 1978, Polish. Sci. Publ., Warszawa (1980), 235–240.

[143] R.C. Kirby and L.C. Siebenmann, *On the triangulation of manifolds and the Hauptvermutung*, Bull. Amer. Math. Soc. **75** (1969), 742–749.

[144] Y. Kodama, *Note on an absolute neighborhood extensor for metric spaces*, J. Math. Soc. Japan **8** (1956), 206–215.

[145] Y. Kodama, *On LC^n metric spaces*, Proc. Japan Acad. **33** (1957), 79–83.

[146] Y. Kodama and A. Koyama, *Hurewicz isomorphism theorem for Steenrod homology*, Proc. Amer. Math. Soc. **74** (1979), 363–367.

[147] Y. Kodama and J. Ono, *On fine shape theory*, Fund. Math. **105** (1979), 29–39.

[148] A. Komatu, *Bemerkungen über die Fundamentalgruppe eines Kompaktums*, Proc. Imp. Acad. Tokyo **13** (1937), 56–58.

[149] G. Kozlowski, *Images of ANR's*, Mimeographed Notes, Univ. of Washington, Seattle (1974).

[150] J. Krasinkiewicz and P. Minc, *Generalized paths and 1-movability*, Fund. Math. **104** (1979), 141–153.
[151] A.H. Kruse, *Introduction to the Theory of Block Assemblages and Related Topics in Topology*, Univ. of Kansas, Lawrence (1956).
[152] K. Kuperberg, *An isomorphism theorem of Hurewicz type in Borsuk's theory of shape*, Fund. Math. **77** (1972), 21–32.
[153] K. Kuratowski, *Sur un théorème fondamental concernant le nerf d'un système d'ensembles*, Fund. Math. **20** (1933), 191–196.
[154] K. Kuratowski, *Sur les espaces localement connexes et péaniens en dimension n*, Fund. Math. **24** (1935), 269–287.
[155] K. Kuratowski, *Quelques problèmes concernant les espaces métriques nonséparables*, Fund. Math. **25** (1935), 534–545.
[156] V. Kuz'minov, *Derived functors of inverse limits and extension classes*, Sibirski Mat. Ž. **12** (2) (1971), 384–396 (in Russian).
[157] R.C. Lacher, *Cell-like mappings and their generalizations*, Bull. Amer. Math. Soc. **83** (1977), 495–552.
[158] S. Lefschetz, *Topology*, Amer. Math. Soc. Coll. Publ. vol. 12, New York (1930).
[159] S. Lefschetz, *On locally connected and related sets*, Ann. Math. **35** (1934), 118–129.
[160] S. Lefschetz, *Locally connected and related sets II*, Duke Math. J. **2** (1936), 435–442.
[161] S. Lefschetz, *Topics in Topology*, Princeton Univ. Press, Princeton, NJ (1942).
[162] E.L. Lima, *The Spanier–Whitehead duality in new homotopy categories*, Summa Brasiliensis Math. **4** (3) (1959), 91–148.
[163] Ju.T. Lisica and S. Mardešić, *Steenrod–Sitnikov homology for arbitrary spaces*, Bull. Amer. Math. Soc. **9** (1983), 207–210.
[164] Ju.T. Lisica and S. Mardešić, *Coherent prohomotopy and strong shape theory*, Glasnik Mat. **19** (39) (1984), 335–399.
[165] Ju.T. Lisica and S. Mardešić, *Strong homology of inverse systems of spaces, II*, Topology Appl. **19** (1985), 45–64.
[166] Yu.T. Lisitsa, *Extension of continuous mappings and the factorization theorem*, Sibirski Mat. Ž. **14** (1973), 128–139 (in Russian).
[167] Yu.T. Lisitsa, *Strong shape theory and the Steenrod–Sitnikov homology*, Sibirski Mat. Ž. **24** (1983), 81–99 (in Russian).
[168] Yu.T. Lisitsa, *The theorems of Hurewicz and Whitehead in strong shape theory*, Dokl. Akad. Nauk SSSR **283** (1) (1985), 31–35.
[169] A.T. Lundell and S. Weingram, *The Topology of CW-Complexes*, Van Nostrand Reinhold, New York (1969).
[170] M. Madirimov, *On the J. Jaworowski equivariant extension theorem*, Uspehi Mat. Nauk **39** (3) (1984), 237–239 (in Russian).
[171] S. Mardešić, *Shapes of topological spaces*, General Topology Appl. **3** (1973), 265–282.
[172] S. Mardešić, *On the Whitehead theorem in shape theory I*, Fund. Math. **91** (1976), 51–64.
[173] S. Mardešić, *Approximate polyhedra, resolutions of maps and shape fibrations*, Fund. Math. **114** (1981), 53–78.
[174] S. Mardešić, *Strong expansions and strong shape theory*, Topology Appl. **38** (1991), 275–291.
[175] S. Mardešić, *Nonvanishing derived limits in shape theory*, Topology **35** (1996), 521–532.
[176] S. Mardešić and A.V. Prasolov, *On strong homology of compact spaces*, Topology Appl. **82** (1998), 327–354.
[177] S. Mardešić and T.B. Rushing, *Shape fibrations I*, Topology Appl. **9** (1978), 193–215.
[178] S. Mardešić and J. Segal, *Movable compacta and ANR-systems*, Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys. **18** (1970), 649–654.
[179] S. Mardešić and J. Segal, *Shapes of compacta and ANR-systems*, Fund. Math. **72** (1971), 41–59.
[180] S. Mardešić and J. Segal, *Shape Theory. The Inverse System Approach*, North-Holland, Amsterdam (1982).
[181] S. Mardešić and T. Watanabe, *Approximate resolutions of spaces and mappings*, Glasnik Mat. **24** (1989), 587–637.
[182] A. Marin and Y.M. Visetti, *A general proof of Bing's shrinkability criterion*, Proc. Amer. Math. Soc. **53** (1975), 501–507.
[183] D.R. McMillan, Jr., *A criterion for cellularity in a manifold*, Ann. of Math. **79** (1964), 327–337.
[184] E. Michael, *Some extension theorems for continuous functions*, Pacific J. Math. **3** (1953), 789–804.

[185] M.L. Mihalik, *Ends of fundamental groups in shape and proper homotopy*, Pacific J. Math. **90** (1980), 431–458.
[186] R.T. Miller, *Mapping cylinder neighborhoods of some ANR's*, Ann. of Math. **103** (1976), 417–427.
[187] J. Milnor, *On spaces having the homotopy type of a CW-complex*, Trans. Amer. Math. Soc. **90** (1959), 272–280.
[188] T. Miyata, *Uniform shape theory*, Glasnik Mat. **29** (49) (1994), 123–168.
[189] Z. Miminoshvili, *On a strong spectral shape theory*, Trudy Tbilissk. Mat. Inst. Akad Nauk Gruzin. SSR **68** (1982), 79–102 (in Russian).
[190] K. Morita, *The Hurewicz and the Whitehead theorems in shape theory*, Science Reports Tokyo Kyoiku Daigaku, Sec. A **12** (1974), 246–258.
[191] K. Morita, *On generalizations of Borsuk's homotopy extension theorem*, Fund. Math. **88** (1975), 1–6.
[192] K. Morita, *On shapes of topological spaces*, Fund. Math. **86** (1975), 251–259.
[193] M. Moszyńska, *The Whitehead theorem in the theory of shapes*, Fund. Math. **80** (1973), 221–263.
[194] P. Mrozik, *Chapman's complement theorem in shape theory: A version for the infinite poroduct of lines*, Arch. Math. **42**, 564–567.
[195] P. Mrozik, *Hereditary shape equivalences and complement theorems*, Topology Appl. **22** (1986), 61–65.
[196] S. Nowak, *Algebraic theory of the fundamental dimension*, Dissertationes Math. **187** (1981), 1–59.
[197] S. Nowak, *On the relationship between shape properties of subcompacta of S^n and homotopy properties of their complements*, Fund. Math. **128** (1987), 47–60.
[198] M. Pavel, *Shape theory and pattern recognition*, Pattern Recognition **16** (1983), 349–356.
[199] T. Porter, *Čech homotopy I*, J. London Math. Soc. **6** (1973), 429–436.
[200] T. Porter, *Proper homotopy theory*, Handbook of Algebraic Topology, I.M. James, ed., Elsevier, Amsterdam (1995), 127–167.
[201] J.B. Quigley, *An exact sequence from the n-th to $(n-1)$-st fundamental group*, Fund. Math. **77** (1973), 195–210.
[202] D.G. Quillen, *Homotopical algebra*, Lecture Notes in Math. vol. 43, Springer, Berlin (1967).
[203] F.S. Quinn, *Resolutions of homology manifolds and the topological characterization of manifolds*, Invent. Math. **72** (1983), 267–284; *Corrigendum*: Invent. Math. **85** (1986), 653.
[204] F.S. Quinn, *An obstruction to the resolution of homology manifolds*, Michigan Math. J. **34** (1987), 285–291.
[205] J.W. Robbin and D. Salamon, *Dynamical systems, shape theory and the Conley index*, Ergod. Th. & Dynamic. Sys. **8** (1988), 375–393.
[206] M.E. Rudin, *A normal space X for which $X \times I$ is not normal*, Fund. Math. **73** (1971), 179–186.
[207] J. Sanjurjo, *An intrinsic description of shape*, Trans. Amer. Math. Soc. **329** (1992), 625–636.
[208] J. Sanjurjo, *On the structure of uniform attractors*, J. Math. Anal. Appl. **152** (1995), 519–528.
[209] J. Segal, S. Spież and B. Günther, *Strong shape of uniform spaces*, Topology Appl. **49** (1993), 237–249.
[210] J. Segal and T. Watanabe, *Cosmic approximate limits and fixed points*, Trans. Amer. Math. Soc. **333** (1992), 1–61.
[211] N. Šekutkovski, *Category of coherent inverse systems*, Glasnik Mat. **23** (1988), 373–396.
[212] J.-P. Serre, *Homologie singulière des espaces fibrés. Applications*, Ann. of Math. **54** (2) (1951), 425–505.
[213] E.V. Shchepin, *Finite-dimensional bicompact absolute neighborhood retracts are metrizable*, Dokl. Akad. Nauk SSSR **233** (3) (1977), 304–307 (in Russian).
[214] L.C. Siebenmann, *Approximating cellular maps by homeomorphisms*, Topology **11** (1972), 271–294.
[215] E.G. Sklyarenko, *Homology and cohomology of general spaces*, Itogi Nauki i Tehniki, Series – Contemporary Problems of Mathematics, Fundamental Directions, vol. 50, Akad. Nauk SSSR, Moscow (1989), 125–228 (in Russian).
[216] E.G. Sklyarenko, *Hyper (co) homology left exact covariant functors and homology theory of topological spaces*, Uspehi Mat. Nauk **50** (3) (1995), 109–146 (in Russian).
[217] S. Smale, *A Vietoris mapping theorem for homotopy*, Proc. Amer. Math. Soc. **8** (1957), 604–610.
[218] Yu.M. Smirnov, *Sets of H-fixed points are absolute extensors*, Mat. Sb. **27** (1) (1975), 85–92 (in Russian).
[219] Yu.M. Smirnov, *On equivariant embeddings of G-spaces*, Uspehi. Mat. Nauk **31** (5) (1976), 137–147 (in Russian).
[220] S. Spież, *An example of a continuum X with* $\mathrm{Fd}(X \times S^1) = \mathrm{Fd}(X) = 2$, Bull. Acad. Polon. Sci. Sér. Sci. Math. **27** (1979), 923–927.
[221] S. Spież, *On the fundamental dimension of the cartesian product of compacta with fundamental dimension 2*, Fund. Math. **116** (1983), 17–32.

[222] N.E. Steenrod, *Regular cycles of compact metric spaces*, Ann. Math. Soc. **41** (1940), No. 4, 833–851.
[223] A.H. Stone, *Paracompactness and product spaces*, Bull. Amer. Math. Soc. **54** (1948), 977–982.
[224] J.L. Taylor, *A counterexample in shape theory*, Bull. Amer. Math. Soc. **81** (1975), 629–632.
[225] H. Tietze, *Über Funktionen die auf einer abgeschlossenen Menge stetig sind*, J. Reine Angew. Math. **145** (1915), 9–14.
[226] H. Toda, *On unstable homotopy of spheres and classical groups*, Proc. Nat. Acad. Sci. **46** (1960), 1102–1105.
[227] H. Toruńczyk, *Compact absolute retracts as factors of the Hilbert space*, Fund. Math. **83** (1973), 75–84.
[228] H. Toruńczyk, *Absolute retracts as factors of normed linear spaces*, Fund. Math. **86** (1974), 53–67.
[229] H. Toruńczyk, *Homeomorphism groups of compact Hilbert cube manifolds that are manifolds*, Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys. **25** (1977), 401–408.
[230] H. Toruńczyk, *On CE-images of the Hilbert cube and characterization of Q-manifolds*, Fund. Math. **106** (1980), 31–40.
[231] H. Toruńczyk, *Characterizing Hilbert space topology*, Fund. Math. **111** (1981), 247–262.
[232] P. Urysohn, *Über die Mächtigkeit der zusammenhängenden Mengen*, Math. Ann. **94** (1925), 262–295.
[233] G.A. Venema, *Embeddings of compacta with shape dimension in the trivial range*, Proc. Amer. Math. Soc. **55** (1976), 443–448.
[234] L. Vietoris, *Über den höheren Zusammenhang kompakter Räume und eine Klasse von zusammenhangstreuen Abbildungen*, Math. Ann. **97** (1927), 454–472.
[235] R.M. Vogt, *Homotopy limits and colimits*, Math. Z. **134** (1973), 11–52.
[236] C.T.C. Wall, *Finiteness conditions for CW-complexes*, Ann. of Math. **81** (1965), 11–52.
[237] J.J. Walsh, *Dimension, cohomological dimension, and cell-like mappings*, Shape Theory and Geometric Topology Proc., Dubrovnik, 1981, Lecture Notes in Math. vol. 870, Springer, Berlin (1981), 105–118.
[238] J.J. Walsh, *Characterization of Hilbert cube manifolds: an alternate proof*, Geometric and Algebraic Topology, Banach Center Publ. vol. 18, Warsaw (1986), 153–160.
[239] T. Watanabe, *Approximative shape I*, Tsukuba J. Math. **11** (1987), 17–59.
[240] J.E. West, *Infinite products which are Hilbert cubes*, Trans. Amer. Math. Soc. **150** (1970), 1–25.
[241] J.E. West, *Compact ANR's have finite type*, Bull. Amer. Math. Soc. **81** (1975), 163–165.
[242] J.E. West, *Mapping Hilbert cube manifolds to ANR's: a solution of a conjecture of Borsuk*, Ann. Math. **106** (1977), 1–18.
[243] J.E. West, *Open problems in infinite dimensional topology*, Open Problems in Topology, J. van Mill and G.M. Reed, eds, North-Holland, Amsterdam (1990), 523–597.
[244] J.H.C. Whitehead, *Note on a theorem due to Borsuk*, Bull. Amer. Math. Soc. **54** (1948), 1125–1132.
[245] J.H.C. Whitehead, *On the homotopy type of ANR's*, Bull. Amer. Math. Soc. **54** (1948), 1133–1145.
[246] J.H.C. Whitehead, *Combinatorial homotopy I*, Bull. Amer. Math. Soc. **55** (1949), 213–245.
[247] R.L. Wilder, *Topology of Manifolds*, Amer. Math. Soc. Colloquium Publ. vol. 32, New York (1949).
[248] M. Wojdysławski, *Rétracts absolus et hyperespaces des continus*, Fund. Math. **32** (1939), 184–192.
[249] T. Yagasaki, *Fiber shape theory*, Tsukuba J. Math. **9** (1985), 261–277.
[250] T. Yagasaki, *Movability of maps and shape fibrations*, Glasnik Mat. **21** (1986), 153–177.

This Page Intentionally Left Blank

CHAPTER 10

Fixed Point Theory

Robert F. Brown

Department of Mathematics, University of California, Los Angeles, CA 90095-1555, USA
E-mail rfb@math.ucla.edu

1. Introduction

In a paper for the Symposium on the Mathematical Heritage of Henri Poincaré in 1980, Felix Browder [8] wrote that

> Among the most original and far-reaching of the contributions made by Henri Poincaré to mathematics was his introduction of the use of topological or 'qualitative' methods in the study of nonlinear problems in analysis .... The ideas introduced by Poincaré include the use of fixed point theorems, the continuation method, and the general concept of global analysis.

Fixed point theory was an integral part of topology at the very birth of the subject in the work of Poincaré in the 1880's. He showed that the solutions to certain important analytic problems could be studied by defining a set X and a function $f : X \to X$ in such a way that the solutions correspond to the *fixed points* of the function f, that is, to the points $x \in X$ such that $f(x) = x$.

In Section 2, we will examine the circle of ideas about fixed points that originated in Poincaré's work and that were subsequently developed, in a topological setting independent of Poincaré's analytic motivation, especially in the celebrated fixed point theorem of L.E.J. Brouwer. As topology became a well-established branch of mathematics in the 1920's, fixed point theory continued to be central to the subject. In particular, the use of homology theory in the fixed point theory of Solomon Lefschetz, that was further developed in that same period by Heinz Hopf, permitted a considerable refinement of Brouwer's discoveries, as we shall see in Section 3.

Lefschetz introduced what is now called the Lefschetz number of a map and proved that if the number is nonzero, then the map has a fixed point. Since the Lefschetz number is defined in terms of the homology homomorphism induced by the map, it is a homotopy invariant. Consequently, a nonzero Lefschetz number implies that each of the maps in the given homotopy class has at least one fixed point. Papers of Jakob Nielsen in the 1920's sought to do more than establish the existence of fixed points. He wished to find the least

HISTORY OF TOPOLOGY
Edited by I.M. James

number of fixed points among all the maps in a homotopy class. We will discuss Nielsen's theory, and what became known as the Nielsen number, in Section 4.

Although, as topology continued to develop in the 1930's and 1940's it came to focus less on fixed point theory, there were important further advances within fixed point theory itself. Kurt Reidemeister and his student Franz Wecken unified the Lefschetz and Nielsen numbers through the Reidemeister trace, which is the subject of Section 5. Wecken also considerably expanded and strengthened the theory that Nielsen had introduced. We will discuss that development in Section 6.

As the summary above indicates, we will be concerned here, for the most part, with the early history of fixed point theory, that is, from its origins in the 1880's until the 1940's. Furthermore, we will not attempt a survey even within that restricted time period, but we will instead concentrate on the principal topological themes of the subject. However, the theory of Nielsen and Wecken raised many questions about the least number of fixed points among the maps in a homotopy class that were only settled much later, primarily through the work of Boju Jiang in the 1980's. Thus, in Sections 6 and 7, we will examine how Jiang completed much of the program that was initiated by Nielsen more than 50 years earlier.

The 1980's was a period of great activity in fixed point theory, including many important developments in addition to the work of Jiang that we will discuss, and that level of activity has continued into the present decade. This is not the place to attempt a survey of that activity, but we have dedicated Section 8 to a brief sampling of some topics that were chosen to illustrate the diversity of topological contexts that now form a part of fixed point theory. In presenting these topics, we will concentrate on how they arose and will not follow their subsequent development, which is still continuing. Consequently, many of the most exciting discoveries of recent years will not be mentioned. It will be the task of some future history of fixed point theory to include the detailed discussion of them that they well merit.

Fixed point theory started in response to the needs of nonlinear analysis, a subject which has undergone a spectacular growth throughout this century, and the topic of such applications continues to be an important one. However, we will be concerned only with the purely topological aspects of fixed point theory. We have limited our discussion to a strictly topological setting: maps on connected finite polyhedra. Thus, by a *manifold* we mean one that is both compact and triangulated. Although many of the results we present remain true in more general settings, our focus on finite polyhedra will unify the presentation as well as exclude related but distinct areas of mathematics. In particular, we will not be concerned with the type of fixed point theory that is related to the Contraction Mapping Principle of Banach.

2. Brouwer's theorem

When a topologist sees the phrase "fixed point theory", the first thing likely to come to mind is the

BROUWER FIXED POINT THEOREM. *Let B^n be the unit ball in Euclidean n-space $\mathbb{R}^n$ and let $f: B^n \to B^n$ be a map, then f has a fixed point.*

To discover the family of mathematical ideas to which this result belongs, we consider the case $n = 1$. We have a map $f:[-1,1] \to [-1,1]$. Let $g:[-1,1] \to \mathbb{R}$ be defined by

$g(x) = x - f(x)$, then $g(-1) \leqslant 0$ and $g(1) \geqslant 0$. If $g(-1)$ or $g(1)$ are 0, then f certainly has a fixed point. Otherwise, we know that f has a fixed point because of the

INTERMEDIATE VALUE THEOREM. *If $g:[-1,1] \to \mathbb{R}$ is a map such that $g(-1) < 0$ and $g(1) > 0$, then $g(c) = 0$ for some c with $-1 < c < 1$.*

This theorem, first proved by Bernard Bolzano in 1817, is a direct consequence of a basic property of the real numbers: every subset that is bounded above has a least upper bound. Its connection to the $n = 1$ case of Brouwer's theorem, is very close indeed because, in fact, the results are "equivalent" in the sense that each directly implies the other. To show that Brouwer's theorem implies Bolzano's, we begin with a map $g:[-1,1] \to \mathbb{R}$ such that $g(-1) < 0$ and $g(1) > 0$. Define $f:[-1,1] \to [-1,1]$ by setting $f(x) = \rho(x - g(x))$, where $\rho(y) = -1$ for $y < -1$, $\rho(y) = 1$ for $y > 1$ and $\rho(y) = y$, otherwise. The Brouwer theorem tells us that $f(c) = c$ for some $c \in [-1,1]$. If $c = -1$, that would imply $g(1) - 1 \leqslant -1$ contrary to the hypothesis, and, similarly, $c \neq 1$. But $-1 < c < 1$ implies $c = f(c) = c - g(c)$ so $g(c) = 0$.

Just as the Brouwer theorem holds in all dimensions, there is a form of the Intermediate Value Theorem that is valid in all dimensions. It concerns a map $f: I^n \to \mathbb{R}^n$, where $I^n = [-1,1] \times [-1,1] \times \cdots \times [-1,1]$. Let

$$I_k(-) = \{x = (x_1, \ldots, x_n) \in I^n \mid x_k = -1\}$$

and let $I_k(+)$ be the same except $x_k = 1$. Write $f = (f_1, \ldots, f_n)$, where $f_k: I^n \to \mathbb{R}$. Denote the origin in $\mathbb{R}^n$ by $\mathbf{0}$.

n-DIMENSIONAL INTERMEDIATE VALUE THEOREM. *Let $f: I^n \to \mathbb{R}^n$ be a map such that $f_k(x) \leqslant 0$ for all $x \in I_k(-)$ and $f_k(x) \geqslant 0$ for all $x \in I_k(+)$, for $k = 1, \ldots, n$, then $f(c) = \mathbf{0}$ for some $c \in I^n$.*

Poincaré announced this result, for f differentiable, in 1883 [46] and published a proof three years later [47]. Since we have seen that the $n = 1$ case of the Brouwer fixed point theorem and Bolzano's intermediate value theorem are equivalent in the sense that each directly implies the other, it should not surprise us to learn that this n-dimensional intermediate value theorem is equivalent to the general Brouwer fixed point theorem in the same sense. But this equivalence was not established until C. Miranda did it in 1940 [43]. As a consequence, what we have here called the n-dimensional intermediate value theorem is usually known as Miranda's theorem or, more accurately, the Poincaré–Miranda theorem.

There is another result equivalent to the Brouwer fixed point theorem, that was published by P. Bohl before any appearance in print of Brouwer's theorem:

BOHL'S THEOREM. *Let $\mathbf{0}$ denote the origin in $\mathbb{R}^n$. There is no map $f: I^n \to \mathbb{R}^n - \mathbf{0}$ such that f is the identity on the boundary of I^n.*

We will see below that Bohl's theorem is also equivalent to the Brouwer theorem. Bohl's theorem was published in 1904 [1], with a proof that required that f be differentiable.

Brouwer published his fixed point theorem, for continuous functions on the 3-ball, in 1909 [5]. When the first proof for the n-ball, with f differentiable, appeared in print a year later, in an appendix by J. Hadamard to a text by Tannery [21], the theorem was called the "Brouwer Fixed Point Theorem", which suggests that the result was already famous by that time. It is not known in what year Brouwer made his discovery and, apparently, communicated it to other mathematicians in an informal manner. The first published proof of the general case, that is, for continuous functions on the n-ball, was by Brouwer himself in 1912 [6].

Poincaré based the proof of the generalized form of the intermediate value theorem on the Kronecker index, introduced in 1869 [36]. The purpose of Hadamard's appendix to Tannery's text [21] was to demonstrate the usefulness of the same tool (see [55]).

Brouwer's proof instead used the notion of topological degree, which he introduced as a replacement for the Kronecker index that would not require the map to be differentiable. To understand Brouwer's proof of his fixed point theorem, we need only consider maps $f: S^n \to S^n$, where S^n is the unit sphere in Euclidean space $\mathbb{R}^{n+1}$. Brouwer proved that f could be approximated by a simplicial map $\phi: S_1^n \to S_2^n$ with respect to triangulations of S^n we denote by S_1^n and S_2^n. For an n-simplex $s \in S_2^n$ that is well-behaved with respect to ϕ, he let p (respectively q) denote the number of n-simplices of S_1^n that ϕ maps to s preserving (respectively reversing) their orientation; and he defined the degree of f to be $p - q$. Thus, in contrast to Hadamard's analytic proof, Brouwer's approach can be described as combinatorial in the sense that the simplicial approximation to the given map relates two combinatorial structures, the triangulations of the sphere. The modern view of this subject, that relates topology to modern algebra, depends on the realization, beginning with Emmy Noether in the mid 1920's, that homology associates groups with spaces and homomorphisms with maps. That approach gives us the present definition of degree of a map of a sphere. We know that $H_n(S^n)$ is infinite cyclic and we choose a generator $[S^n]$. The homomorphism induced by $f: S^n \to S^n$ is determined by the image of a generator, so the degree d defined by $f_*([S^n]) = d[S^n]$ characterizes the homomorphism. It can then be proved that $d = p - q$ and therefore the definition is independent of the many choices Brouwer made in computing $p - q$, as Brouwer had shown by a lengthy combinatorial argument. Brouwer also proved that homotopic maps have the same degree which, as we shall now see, was the other crucial property he required to prove his fixed point theorem. (For a more detailed exposition of Brouwer's development of the degree, see [11].)

Brouwer made use of the degree to prove his theorem as follows. It is easy to see from Brouwer's definition that the degree of the antipodal map $a: S^n \to S^n$ defined by $a(x) = -x$ is nonzero, specifically $(-1)^{n+1}$. If a map $g: S^n \to S^n$ has no fixed points, then $(1-t)g(x) + ta(x) = 0$ has no solutions, so defining $H(x,t) = \pi((1-t)g(x) + ta(x))$, where $\pi(x) = x/|x|$, shows that g is homotopic to a and is therefore of the same degree. What Brouwer had thus established is worth stating as a formal result, which we call

BROUWER'S SPHERE THEOREM. *If $g: S^n \to S^n$ is a map of degree different from $(-1)^{n+1}$, then g has a fixed point.*

Now, to complete Brouwer's proof of his fixed point theorem let

$$\mathbb{R}_+^{n+1} = \{(x_1, \ldots, x_n, x_{n+1}) \in \mathbb{R}^{n+1}: x_{n+1} \geqslant 0\}$$

and define $\mathbb{R}_-^{n+1}$ similarly, and set $S_+^n = S^n \cap \mathbb{R}_+^{n+1}$. A map $f: B^n \to B^n$ gives $f' = hfh^{-1}: S_+^n \to S_+^n$, where $h: B^n \to S_+^n$ is a homeomorphism. Define $g: S^n \to S_+^n \subset S^n$

by setting it equal to f' on S^n_+ and for $x = (x_1, \ldots, x_n, x_{n+1}) \in S^n_-$, letting $g(x) = f'(x_1, \ldots, x_n, -x_{n+1})$. It is clear from Brouwer's definition that the degree of g must be zero so, by his Sphere theorem, it has a fixed point, which must lie in S^n_+. Thus f', and consequently f, has a fixed point.

In 1931, Karol Borsuk observed [2] that the Brouwer fixed point theorem is a consequence of (and it in turn implies) the

NO-RETRACTION THEOREM. *There is no map $r : B^n \to \partial B^n$ such that $r(x) = x$ for all x in ∂B^n, the boundary of B^n.*

This result has been particularly amenable to proof using a wide variety of mathematical tools, and thus the Brouwer fixed point theorem is the consequence of insights obtained from strikingly diverse points of view. But we will not examine this extensive literature, nor will we explore other mathematical statements that have been shown to be the equivalent of Brouwer's theorem. (A good source of information on these topics is [14].)

However, it is instructive to observe how easily the no-retraction theorem can be related to Bohl's theorem, that was published 27 years earlier (this result was apparently unknown to Borsuk). Actually, it is the negations of the no-retraction theorem and of Bohl's theorem that are so similar. On the one hand, a retraction of I^n to its boundary is certainly a map of I^n to $\mathbb{R}^n - \mathbf{0}$ that is the identity on the boundary of I^n. On the other hand, if there is a map of I^n to $\mathbb{R}^n - \mathbf{0}$ that is the identity on the boundary of I^n, then following that map by radial retraction of $\mathbb{R}^n - \mathbf{0}$ onto the boundary of I^n gives us a retraction of I^n onto its boundary.

We can quickly complete this circle of ideas by showing the equivalence of the no-retraction theorem and Brouwer's theorem or, again more easily, their negations. First of all, if a retraction of I^n to its boundary exists, then, identifying the boundary of I^n with S^{n-1} and following the retraction by the antipodal map a we used above produces a map of I^n without fixed points. Conversely, if there is a map f of I^n to itself without fixed points, send each point x in I^n to the point where the ray from $f(x)$ through x intersects the boundary of I^n. Once we convince ourselves that the function obtained in this manner is continuous, we see that it retracts I^n to its boundary.

In fairness to Borsuk, we should note that his purpose in using the Brouwer theorem was not just the no-retraction theorem, but rather a more subtle observation. Borsuk proved in [2] that if A is a retract of $\mathbb{R}^n$, then every component of $\mathbb{R}^n - A$ is unbounded. Thus, although the boundary of I^n cannot be such a retract of $\mathbb{R}^n$ and thus by restriction not a retract of I^n, many other subsets of $\mathbb{R}^n$ are excluded as retracts as well.

3. Lefschetz–Hopf theory

Lefschetz published three brief notes in the 1923 and 1925 volumes of the *Proceedings* of the U.S. National Academy of Sciences about the research that included his celebrated fixed point theorem. In the first of these [37], he announced that he had "new and far reaching methods" for investigating continuous maps of manifolds and, in particular, their fixed points.

Lefschetz had been studying intersections within manifolds. Given a closed oriented manifold W and oriented submanifolds M and N with $\dim M + \dim N = \dim W$ intersecting nicely in a finite number of points, each intersection point is assigned a sign, either positive or negative. The number of positive intersections minus the number of negative intersections Lefschetz called the *algebraic number of intersections* and denoted (MN). In a suitably restricted context, this number is the Kronecker index to which we made reference in the previous section. Thus, Lefschetz was concerned with extending the Kronecker index for differentiable functions of Euclidean spaces to obtain a tool for studying continuous functions on manifolds.

Lefschetz related the study of a map $f : X \to X$ to intersection theory by means of the graph. Within the manifold $X \times X$ the graph

$$\Gamma_f = \{(x, f(x)) \colon x \in X\}$$

is embedded as a submanifold. The graph may also be thought of as an n-cycle in the $2n$-manifold $X \times X$; this was important to Lefschetz because he used homology to investigate the algebraic number of intersections. Given another map $g : X \to X$, then its graph Γ_g is another such n-cycle. A point (x_1, x_2) is in the intersection of Γ_f and Γ_g if and only if

$$(x_1, x_2) = (x_1, f(x_1)) = (x_1, g(x_1))$$

so $x_1 \in X$ has the property that $f(x_1) = g(x_1)$. Such a point is called a *coincidence* of the maps f and g. Thus $(\Gamma_f \Gamma_g)$ gives an algebraic count of the number of coincidences. The coincidence question is related to fixed points by taking g to be the identity map, so that $(\Gamma_f \Gamma_g)$ is then concerned with fixed points. Although coincidences arise naturally in the context of the intersection of graphs, the study of solutions to the coincidence equation $f(x) = g(x)$ was certainly not a common topic in the topology of the time. However, it seems that Lefschetz's point of view was not unusual in algebraic geometry, a subject in which Lefschetz also had a very strong interest. The problem, whether one is concerned with coincidences or the more familiar fixed points, is to relate the algebraic number of intersections of the graphs to properties of the maps themselves, and to be able to calculate it.

In the second of the three announcements [38], Lefschetz takes up the rather technical matter of determining the algebraic intersection of cycles in a suitably general setting for his later purposes. Then, in the final announcement [39] he presents his main result: a formula for calculating $(\Gamma_f \Gamma_g)$, the algebraic number of intersections of the graphs of $f, g : X \to X$, from what we would now call the induced homomorphisms of rational homology. It is instructive to see the form of the main formula of [39]. Choosing a suitable basis for the homology of a closed orientable n-manifold X, the formula is

$$(\Gamma_f \Gamma_g) = \sum_{\mu=0}^{n} (-1)^{\mu} \sum_{i,j=1}^{R_\mu} \alpha_{\mu}^{ij} \beta_{n-\mu}^{ij},$$

where the R_μ are the Betti numbers of X, the α depend on the homomorphism induced by f and the β come from the Poincaré dual of the homomorphism induced by g (see below). A more thorough discussion of the Lefschetz formula from the point of view of the topology of that time can be found in [11].

Lefschetz published the detailed proofs of the results announced in the *Proceedings* in [40]. This paper is in two parts. In the first, Lefschetz develops what he once more characterizes as a "far reaching theory of the intersection of complexes on a manifold" in which the "intersection" is a cycle with the property that its homology class is independent of the constructions necessary to define it. He obtains the results described in [38], but in a somewhat simpler manner than he outlined in that announcement. In the case that the complexes intersect in a finite set of points, this intersection is seen to be a generalization of the Kronecker index. In the second part of the paper, Lefschetz applies the intersection theory to calculate $(\Gamma_f \Gamma_g)$ in the way he described it in [39]. This calculation holds only in the setting of oriented manifolds without boundary, but he states: "With suitable restrictions the formulas derived are susceptible of extension to a wider range of manifolds, but this will be reserved for a later occasion".

That "later occasion" was the publication just one year later of [41]. This paper is much more than an extension of [40] to manifolds with boundary. Lefschetz uses matrices to simplify substantially the formulas he had presented in [40], in particular by making use of the concept of the trace of a matrix. Here, in modern terminology, is what the main formula of [39] tells us, as it was presented in [41]. Let M be a closed, orientable n-manifold and let $f, g : M \to M$ be maps. For each k, where $0 \leqslant k \leqslant n$, define a linear transformation of the rational homology $H_k(M)$, that I'll denote (in a notation suggested by Lefschetz's) by $(PQ)_k$, to be the composition

$$(PQ)_k = D_k^{-1} \circ g^{n-k} \circ D_k \circ f_k,$$

where $D_k : H_k(M) \to H^{n-k}(M)$ is the Poincaré duality isomorphism and $f_k : H_k(M) \to H_k(M)$ and $g^{n-k} : H^{n-k}(M) \to H^{n-k}(M)$ are induced homomorphisms. Lefschetz's main formula is

$$(\Gamma_f \Gamma_g) = \sum_{k=0}^{n} (-1)^k \, trace(PQ)_k. \tag{$*$}$$

The right-hand side of $(*)$ is now called the *Lefschetz coincidence number* and denoted by $L(f, g)$. What Lefschetz obtained from the formula is the

LEFSCHETZ COINCIDENCE THEOREM. *Let M be a closed, orientable n-manifold and let $f, g : M \to M$ be maps. If $L(f, g) \neq 0$, then f and g have a coincidence, that is, $f(x) = g(x)$ for some $x \in M$.*

Lefschetz did not state this result explicitly because there was no need to do so; it was evident from the left-hand side of $(*)$. If $L(f, g) \neq 0$ but there were no coincidences, then Γ_f and Γ_g would be disjoint subsets of $M \times M$. In order to define $(\Gamma_f \Gamma_g)$ in general it is necessary to modify these subsets, but only by an arbitrarily small amount, so disjoint subsets could still be kept disjoint and thus there would be no points of intersection. Therefore $(\Gamma_f \Gamma_g)$, the difference between the number of positive and the number of negative intersection points, would be 0, contrary to the formula $(*)$.

In addition to this important simplification of [40], Lefschetz did, as promised, extend the theory to manifolds with boundary. For this purpose, he devoted part of [41] to obtaining duality theorems for compact orientable manifolds with boundary, to produce isomorphisms that replace the Poincaré duality isomorphism D_k in the definition of $(PQ)_k$.

Furthermore, he extended the coincidence setting by considering maps $f, g: M \to N$, where M and N are different manifolds of the same dimension.

Once $(*)$ was established, Lefschetz specialized it to obtain what he called his "fixed point formula". From the modern definition above, we immediately see that, taking g to be the identity map, $(PQ)_k$ reduces to the induced homomorphism f_k and the right-hand side of $(*)$ becomes what is now called the *Lefschetz number*, namely,

$$L(f) = \sum_{k=0}^{n} (-1)^k \, trace(f_k).$$

From $(*)$, now in the more general setting of [41], we have the

LEFSCHETZ FIXED POINT THEOREM. *Let M be a compact, orientable manifold, with or without boundary, and let $f: M \to M$ be a map. If $L(f) \neq 0$, then f has a fixed point.*

Even the closed manifold version of this result, already obtained in [40], is sufficient to prove the Brouwer fixed point theorem because, if $M = S^n$, then

$$L(f) = 1 + (-1)^n (\text{degree of } f)$$

which is zero if and only if the degree of f is $(-1)^{n+1}$, so Lefschetz's theorem implies Brouwer's sphere theorem in this case.

A particular case of $(*)$ that Lefschetz emphasized was the fixed point setting with the map f a *deformation* of M, that is, a map homotopic to the identity. In that case, $trace(f_k)$ is the dimension of the vector space $H_k(M)$ so $L(f)$ equals the Euler characteristic of M. Lefschetz was sufficiently impressed with this part of his theory to restate $(*)$ as a formal theorem, in this way: "For every M, with or without boundary, the number of signed fixed points of a deformation is the Euler characteristic".

Coincidence theory requires a duality isomorphism in order to define $L(f, g)$, but it is clear that, for the definition of the Lefschetz fixed point number $L(f)$, all that is needed is that the sum of the traces of the induced homomorphisms f_k be finite. A map $f: X \to X$ on a finite polyhedron X thus has a well-defined Lefschetz number. The generalization of Lefschetz's theorem to this setting was published by Hopf in 1929 [24].

Hopf first considers a simplicial map $f: X \to X$ on a finite polyhedron. On the simplicial k-chains $C_k(X)$ generated by the k-simplices of X, we have the chain map $\phi_k: C_k(X) \to C_k(X)$ induced by f. The fact that ϕ_k maps cycles Z_k to cycles and boundaries B_k to boundaries gives us the induced homomorphism of homology $f_k: H_k(X) \to H_k(X)$. Hopf then proved the

HOPF TRACE THEOREM. *Let $f: X \to X$ be a simplicial map of a finite polyhedron of dimension n, then*

$$\sum_{k=0}^{n} (-1)^k \, trace(f_k) = \sum_{k=0}^{n} (-1)^k \, trace(\phi_k).$$

It is not difficult to convince ourselves of the truth of Hopf's equation. The definition of f_k implies that

$$trace(f_k) = trace(\phi_k|Z_k) - trace(\phi_k|B_k).$$

On the other hand, we can express the trace of the matrix of ϕ_k in terms of traces of restrictions, that is

$$trace(\phi_k) = trace(\phi_k|Z_k) + trace(\phi_{k-1}|B_{k-1}).$$

Comparing these formulas, we see that, because of the alternating sign, we obtain the same sum in each case.

If f is the identity map, we have seen that $L(f)$ is the Euler characteristic. On the other hand, in this case $trace(\phi_k)$ counts the number of k-simplices in X, so the Hopf trace theorem reduces to the classic *Euler–Poincaré formula*

$$\text{Euler characteristic} = \sum_{k=0}^{n}(-1)^k(\text{number of } k\text{-simplices}).$$

The Hopf trace theorem allowed Hopf to express the Lefschetz number $L(f)$ in terms of the chain map ϕ. Hopf then proved the Lefschetz fixed point theorem for maps of finite polyhedra by observing that if a map $f : X \to X$ of a finite polyhedron has no fixed points then, using sufficiently fine triangulations of X, no simplex is mapped into a simplex that contains it, from which it follows that $trace(\phi_k) = 0$ for all k and therefore $L(f) = 0$.

In the last part of [24], Hopf studied the relationship between the Lefschetz number and the concept of degree that had been so central to Brouwer's earlier work, as we discussed in the previous section. Consider a map $f : X \to X$ on a finite dimensional polyhedron. Suppose f has finitely many fixed points and each fixed point p of f is contained in a maximal simplex, that is, a simplex that is not a face of a simplex of higher dimension. Denote by s_p the maximal simplex containing the fixed point p and the dimension of s_p by $k(p)$, which varies with p. The simplex s_p may be identified through a homeomorphism with $\mathbb{R}^{k(p)}$ in such a way as to identify p with the origin. For each such p, take the image B_p in s_p of an Euclidean ball centered at the origin small enough so that B_p is mapped into s_p by f. Denote the boundary of B_p by S_p. The map on S_p, defined by mapping x to $x - f(x)$ followed by radial retraction onto S_p (both steps make sense in terms of the Euclidean structure on s_p) is a map which we denote by $(i - f)_p : S_p \to S_p$.

LEFSCHETZ–HOPF THEOREM. *If $f : X \to X$ is a map on a finite polyhedron that has finitely many fixed points and each fixed point p of f is contained in a maximal simplex, then*

$$L(f) = \sum_{p} \textit{degree of } (i - f)_p,$$

where the sum is taken over all fixed points of f.

This theorem obviously extends the Lefschetz fixed point theorem to maps of finite polyhedra since it implies that an empty fixed point set makes the value of the Lefschetz number zero. Hopf proved the theorem only in the case that X is an n-dimensional polyhedron and each fixed point lies in an n-dimensional simplex. His proof then made use of a subdivision of X such that no simplex of dimension $k < n$ contributes to the trace of the chain map ϕ_k, that is, all diagonal entries of a matrix representation of ϕ_k are zero, so $trace(\phi_k) = 0$. Then he was able to identify the contribution to $trace(\phi_n)$ of an n-simplex containing a fixed point p with the degree of $(i - f)_p$ and the Hopf trace theorem completed the argument. Actually, instead of the maps we have called $(i - f)_p$, Hopf's definition of the maps whose degrees determine $L(f)$ used the difference $f(x) - x$, which would be $(f - i)_p$ in our notation, and that introduced a factor of $(-1)^n$ into the statement of the theorem.

The Lefschetz–Hopf theorem stated above seems to have been long known, but only as a "folk theorem" of fixed point theory with no proof available in print. A result of Albrecht Dold [13] published in 1965 concerning the fixed point index, an extension of the Lefschetz number concept, implies the Lefschetz–Hopf theorem. However, there is no statement of the theorem above in [13] and, in fact, the relationship between the fixed point index and the degree concept, in the case of a finite fixed point set, is not discussed. That relationship was made explicit in [9], which includes Dold's theorem, but there is no formal statement of the Lefschetz–Hopf theorem there either.

The reason the dimensions of the maximal simplices may be permitted to vary in the statement of the Lefschetz–Hopf theorem is made clear in Dold's work. Embedding X in a Euclidean space of some dimension N, there is a neighborhood U of X in $\mathbb{R}^N$ that can be retracted to X by a (deformation) retraction $r: U \to X$. Each fixed point p in a maximal simplex lies in an N-dimensional ball B_p that is mapped near p by the composition fr. Making use of the map $(i - fr)_p : S_p \to S_p$ of the boundary S_p of B_p puts us into a setting like the one Hopf considered. It is clear that $L(fr) = L(f)$ and a suspension argument readily establishes the fact that degree of $(i - fr)_p =$ degree of $(i - f)_p$.

Since Hopf's paper [24], there have been a number of generalizations and extensions of the Lefschetz fixed point theorem. We will not explore that topic here except to state one extension that we will need to refer to in a later section. Suppose X is a finite n-dimensional polyhedron, A is a subpolyhedron and $f: X \to X$ is a map such that $f(A) \subseteq A$, so we have a map of pairs $f: (X, A) \to (X, A)$, then f induces $f_k: H_k(X, A) \to H_k(X, A)$. In 1968, C. Bowszyc [3] defined, by analogy with the Lefschetz number, a number that we denote in this way

$$L(f; X, A) = \sum_{k=0}^{n} (-1)^k \, trace(f_k)$$

and he proved

BOWSZYC'S THEOREM. *If X is a finite polyhedron, A is a subpolyhedron, and $f: (X, A) \to (X, A)$ is a map of pairs such that $L(f; X, A) \neq 0$, then there is a fixed point of f in the closure of the set $X - A$.*

4. Nielsen theory

At the beginning of his first announcement [37], Lefschetz identified as a key problem in the study of maps: "Determination of the minimum number of fixed points and related questions". In the context of his work, the minimum number he was referring to concerned the homotopy class of a map. The Lefschetz number is a homotopy invariant *algebraic* count of the number of fixed points, but it does not in general offer information on the actual number of points kept fixed by the maps in the homotopy class, as Lefschetz pointed out.

The determination of the minimum number of fixed points for all maps homotopic to a given one was a question of great interest to Nielsen, starting with his doctoral dissertation of 1913. In [44], Nielsen considered the lifts of a map $f : T \to T$ to the universal covering space of the torus, that is, the plane. Choosing a basepoint in T, he identified the set of lifts with a fiber, so it is a set of points in the plane whose coordinates differ by integer amounts. He then defined an equivalence relation on the fiber as follows: two points are equivalent if the difference of their coordinates lies in the image of the linear transformation of the plane with matrix $I - F$, where I denotes the identity matrix and F is a matrix representing the homomorphism of the fundamental group of T induced by f. This relation partitions the lifts of f into $|det(I - F)|$ equivalence classes. The fixed points of inequivalent lifts project onto disjoint sets of fixed points of f (see below). Nielsen proved that if $I - F$ is nonsingular, then every lift has fixed points so, since F only depends on the homotopy class of f, he concluded that, if $I - F$ is nonsingular, then every map homotopic to f has at least $|det(I - F)|$ fixed points.

This lower bound had already been established in a paper of Brouwer [7], using methods that applied directly to the torus and did not relate the problem to the plane. Nielsen, on the other hand, used his equivalence relation on the lifts to the plane to partition the entire set of fixed points of the map f of the torus into disjoint subsets, namely, the projections of the fixed points of the equivalence classes of lifts. At the end of that section of his paper, Nielsen noted that although the bound was already known through the work of Brouwer, "the division of . . . fixed points into finitely many classes as above seems to be the simplest and most natural way of proving minimality" [44].

Nielsen put this "class" concept in a more general context in [45]. Now he was concerned with mappings of closed orientable surfaces of genus greater than one, so the universal covering space is the hyperbolic rather than the Euclidean plane. Given a selfmap f of such a surface, Nielsen still considered the lifts of f to the universal covering space, but he expressed the equivalence relation between them in different terms. Lifts $\tilde{f}_1$ and $\tilde{f}_2$ are *equivalent* if there is a covering transformation ω such that $\tilde{f}_2 = \omega \tilde{f}_1 \omega^{-1}$, that is, the lifts $\tilde{f}_1$ and $\tilde{f}_2$ are in the same conjugacy class with respect to the covering transformations. We will see below that this definition extended his definition for maps of the torus.

Expressed in this way, it is not difficult to understand how fixed points of equivalence classes of lifts of f relate to the set of fixed points of f itself. First of all, each fixed point p of f is the projection of a fixed point of some lift since we can choose any point $\tilde{p}$ in the covering space that projects to p and there is a unique lift of f that takes $\tilde{p}$ to itself. Furthermore, if $\tilde{f}_1$ fixes $\tilde{p}$ then $\omega \tilde{f}_1 \omega^{-1}$ fixes $\omega(\tilde{p})$, which is in the same fiber, so the fixed point sets of conjugate lifts project to the same subset of fixed points of f. Moreover, if lifts $\tilde{f}_1$ and $\tilde{f}_2$ have fixed points $\tilde{p}_1$ and $\tilde{p}_2$, respectively, that are in the same fiber, let ω be the covering transformation that takes $\tilde{p}_1$ to $\tilde{p}_2$, then it is not difficult to see that $\tilde{f}_1$ and $\tilde{f}_2$ are conjugate by ω. In conclusion, letting a *fixed points class* be the projection of the fixed points of a conjugacy class of lifts, Nielsen again partitioned the set of fixed points of f into disjoint subsets in a "natural way".

Suppose we choose a lift $\tilde{f}$ of f, then every lift can be expressed uniquely as $\alpha^{-1}\tilde{f}$ for some covering transformation α. Identifying the covering transformations with the elements of the fundamental group, the homomorphism f_π of the fundamental group induced by f is represented as follows: For δ a covering transformation, the composition of maps $\tilde{f}\delta$ is a lift of f which is of the form $\tilde{f}\delta = \alpha^{-1}\tilde{f}$, where $\alpha^{-1} = f_\pi(\delta)$ as elements of the fundamental group.

Lifts $\alpha^{-1}\tilde{f}$ and $\beta^{-1}\tilde{f}$ are conjugate by a covering transformation ω if and only if

$$\beta^{-1}\tilde{f} = \omega\alpha^{-1}\tilde{f}\omega^{-1} = \omega\alpha^{-1}f_\pi(\omega^{-1})\tilde{f},$$

that is, if and only if

$$\beta = f_\pi(\omega)\alpha\omega^{-1}.$$

(The form of this relationship depends on the choice of convention for the operation of the fundamental group on the universal covering space. We are following [22]; see Section 5 below.) If the fundamental group is Abelian, as it is for the torus, we can write that relationship additively and rearrange the terms as

$$\alpha - \beta = \omega - f_\pi(\omega) = (Id - f_\pi)(\omega).$$

To relate Nielsen's more general theory to his earlier analysis of the torus, consider a fiber in the plane, that is, a basepoint (x_0, y_0) and all integral translates of it, that in turn correspond to the covering transformations for the torus. Now representing the homomorphism $Id - f_\pi$ by a matrix $I - F$, we see from the equation above that lifts are equivalent if and only if, for the points in the fiber corresponding to the covering transformations that determine these lifts, the difference of their coordinates lies in the image of the linear transformation of the plane with matrix $I - F$, just as in [44].

In the torus setting, it was possible to show that all lifts have fixed points if $I - F$ is nonsingular. Thus the lower bound for the number of fixed points was equal to the number of equivalence classes of lifts. For maps of other surfaces, not all lifts have fixed points, so Nielsen required a tool to establish the existence of fixed points of lifts. Assuming that the set of fixed points is finite, Nielsen took advantage of the structure of the hyperbolic plane to define a "winding number" that measured the way that a map goes around a fixed point. The *index of a fixed point class* (assuming a finite set of fixed points) was then the sum of the winding numbers of the fixed points in it. We will require a definition that applies to more general contexts, so instead of winding numbers, we will next use some ideas from the slightly later paper of Hopf [24], that we discussed in the previous section, to defined the index of a fixed point class.

Let $f : X \to X$ be a map of a finite polyhedron. The definition of a fixed point class of f, as the projection of the fixed points of a conjugacy class of lifts, extends to this setting because the discussion above made no use of any property of a surface other than that it has a universal covering space. The fixed point classes $\mathbf{F}_i$ are disjoint compact subsets of X, so there are a finite number of them and there are disjoint open sets U_i containing them. By a technique in [24], the map f may be approximated arbitrarily closely by a simplicial map, call it f', such that each fixed point of f' lies in a maximal simplex of a triangulation of X. Making f' close enough to f ensures that f' has no fixed points outside of the U_i

and the fixed points in each U_i form a single fixed point class of f'. The *index of the fixed point class* $\mathbf{F}_i$ is defined to be the sum of the degrees of the $(i - f')_p$ for all fixed points p of f' in U_i (compare the Lefschetz–Hopf theorem of the previous section). The index is independent of the choice of the approximating simplicial map f'. Notice that the sum of the indices of all the fixed point classes is the sum of the degrees of the $(i - f')_p$ for all fixed points of f' and therefore the sum of the indices of all fixed point classes equals the Lefschetz number $L(f') = L(f)$ by the Lefschetz–Hopf theorem.

Nielsen stated in [45] that "it is to be expected" that the number of fixed point classes of nonzero index will be invariant under homotopy, though he was able to establish this fact only in special cases. A class of nonzero index is called an *essential* fixed point class and the number of essential fixed point classes is now called the *Nielsen number* of a map $f : X \to X$ of a finite polyhedron and denoted by $N(f)$.

Nielsen was correct: if f and g are homotopic maps, then $N(f) = N(g)$, as Wecken proved in 1941 [58]. There is a convenient way to view the behavior of fixed point classes under a homotopy, that was introduced by Kurt Scholz in 1974 [53]. Let

$$H = \{h_t\} : X \times I \to X$$

be a homotopy between maps f and g of a finite polyhedron and define

$$\mathbf{H} : X \times I \to X \times I$$

by $\mathbf{H}(x, t) = (h_t(x), t)$. Scholz observed that the fixed point classes of h_t are precisely the intersections of the fixed point classes of $\mathbf{H}$ with the subset $X \times \{t\}$. Take disjoint open sets in $X \times I$ about the fixed point classes of $\mathbf{H}$ and consider the intersection of one such set with $X \times \{t\}$, calling it U_i. For t' sufficiently close to t, the fixed point class of $h_{t'}$ that is the intersection with that same class of $\mathbf{H}$ is also contained in U_i. We may use the same simplicial approximation, with fixed points only in maximal simplices, for both h_t and $h_{t'}$, so the fixed point classes of h_t and of $h_{t'}$ that lie in U_i have the same index. Thus, as t changes, the value of the index of the fixed point class of h_t obtained by intersecting a given fixed point class of $\mathbf{H}$ with $X \times \{t\}$ is locally constant, so it is constant. In this way, the fixed point classes of $\mathbf{H}$ that at all levels have nonzero index determine the required one-to-one correspondence between the essential fixed point classes of f and those of g.

The immediate consequence of the homotopy invariance of the Nielsen number is the

NIELSEN FIXED POINT THEOREM. *Let $f : X \to X$ be a map of a finite polyhedron, then every map homotopic to f has at least $N(f)$ fixed points.*

Nielsen recognized in [45] that the calculation of what we now call the Nielsen number was a crucial question raised by his new theory, in fact he called it the "general fixed point problem". Nielsen's earlier paper [44] had calculated $N(f)$ for a map $f : T \to T$ of a torus when $I - F$ is nonsingular, where we recall that F depends on the homomorphism of the fundamental group of T induced by f. The calculation was that $N(f) = |det(I - F)|$ because Nielsen showed that f has $|det(I - F)|$ fixed point classes and it can be proved that all of them have nonzero index, as we will see.

There was little progress on the computation problem until Jiang published a significant extension of the torus calculation in the early 1960's [27]. We recall that the *cokernel* of a homomorphism $h : G \to G$ of an Abelian group G is the quotient group $G/h(G)$.

JIANG'S LIE GROUP THEOREM. *Let G be a compact Lie group and let $f : G \to G$ be a map. Let f_π denote the endomorphism of the fundamental group of G induced by f. If the Lefschetz number $L(f)$ is zero then $N(f) = 0$. If $L(f) \neq 0$, then $N(f)$ is the order of the cokernel of the homomorphism $Id - f_\pi$ of the fundamental group, where Id denotes the identity.*

The group structure on G was used by Jiang to prove his theorem in the following way. Choose a lift $\tilde{f}_0$ of f and let $\alpha^{-1}\tilde{f}_0$ be a lift of f, for some covering transformation α. Regard α as an element of the fundamental group of G based at the identity element e of the group G, so α is represented by a loop a based at e. Defining $h_t : G \to G$ by $h_t(x) = a(1-t)f(x)$ by means of the group multiplication on G gives a homotopy $H = \{h_t\}$ with $h_0 = h_1 = f$ under which the fixed point class of $f = h_0$ determined by the conjugacy class of $\tilde{f}_0$ and the fixed point class of $f = h_1$ determined by the conjugacy class of $\alpha^{-1}\tilde{f}_0$ are contained in the same fixed point class of $\mathbf{H}$. Thus both classes have the same index and, since α was arbitrary, we conclude that all the fixed point classes have the same index. Now $L(f) = 0$ implies $N(f) = 0$ because, as we noted, the sum of the indices of all fixed point classes equals $L(f)$. Otherwise, $N(f)$ equals the number of conjugacy classes of lifts. We have seen that if the fundamental group is Abelian, as it is for a Lie group, lifts $\alpha^{-1}\tilde{f}_0$ and $\beta^{-1}f_0$ are conjugate by ω if and only if

$$\alpha - \beta = \omega - f_\pi(\omega) = (Id - f_\pi)(\omega).$$

Thus, the number of conjugacy classes is the order of the cokernel of the homomorphism $Id - f_\pi$, as Jiang claimed.

In the case of the torus T, it is not difficult to show that $L(f) = det(I - F)$. If $I - F$ is nonsingular, so $L(f) \neq 0$, we choose generators for the fundamental group of T, the free Abelian group on two generators, in such a way that the homomorphism $Id - f_\pi$ is represented by the matrix $I - F$. There are unimodular matrices A and B such that $A(I - F)B = D = diag(d_1, d_2)$, a diagonal matrix. Then the order of the cokernel of $Id - f_\pi$ is the order of the direct sum of the cyclic groups of order d_1 and d_2 which is $|d_1 \cdot d_2| = |det(D)| = |deg(I - F)|$. Thus Jiang's theorem generalizes Nielsen's original calculation.

Since the time of Jiang's paper, much progress in calculating the Nielsen number has been made (see [42]), but Nielsen's "general fixed point problem" is still a challenging one.

5. The Reidemeister trace

We have seen that Lefschetz was able to simplify the main result of his fixed point theory in [39] by making use in [41] of the concept of the trace. Nielsen's paper [45] demonstrated the role of the universal covering space in fixed point theory. In a paper of 1936 [48], Reidemeister combined these notions by lifting the trace to the universal covering space, in a manner we will now describe.

We begin with a map $f : X \to X$ of a finite n-dimensional polyhedron. The simplicial structure on X can be lifted to its universal covering space $\widetilde{X}$. We choose a lift $\tilde{f}$ of f which we may assume is a simplicial map $\tilde{f} : \widetilde{X}' \to \widetilde{X}$, where $\widetilde{X}'$ is a subdivision of $\widetilde{X}$. We have the homomorphism $\tilde{f}_k : C_k(\widetilde{X}) \to C_k(\widetilde{X})$ of the k-chains of $\widetilde{X}$ obtained by composing the subdivision operator with the chain map induced by $\tilde{f}$. Denoting the fundamental group of X by π, the chains $C_k(\widetilde{X})$ may be viewed as a free module over the group ring $\mathbb{Z}[\pi]$, with a basis obtained by taking over each k-simplex s in X one k-simplex $\tilde{s}$ in $\widetilde{X}$. For each generator $\tilde{s}$ of $C_k(\widetilde{X})$ and element $\alpha \in \pi$, it can be shown that

$$\tilde{f}_k(\alpha\tilde{s}) = f_\pi(\alpha)\tilde{f}_k(\tilde{s}),$$

where, as in the previous section, f_π is the homomorphism of the fundamental group induced by f. The matrix with entries in $\mathbb{Z}[\pi]$ that represents $\tilde{f}_k$ with respect to that basis has a trace which we will denote by $trace(\tilde{f}_k) \in \mathbb{Z}[\pi]$.

Now we recall from the previous section the equivalence relation on π that resulted from considering lifts of f equivalent if they were in the same conjugacy class with regard to the covering transformations: Elements α and β of π are equivalent if and only if $\beta = f_\pi(\omega)\alpha\omega^{-1}$ for some $\omega \in \pi$. Let $R(f_\pi)$ be the set of equivalence classes and let $\rho : \pi \to R(f_\pi)$ send an element to its equivalence class, then ρ extends linearly to $\rho : \mathbb{Z}[\pi] \to \mathbb{Z}[R(f_\pi)]$. The *Reidemeister trace* of f is the element $\mathcal{R}(f) \in \mathbb{Z}[R(f_\pi)]$ defined by

$$\mathcal{R}(f) = \sum_{k=0}^{n} (-1)^k \rho\big(trace(\tilde{f}_k)\big).$$

The definition of the Reidemeister trace shows in its form Reidemeister's debt to Lefschetz, but there is more to it than that, as Wecken showed in a paper of 1942 [59]. We have seen that there is a one-to-one correspondence between $R(f_\pi)$ and the conjugacy class of lifts of f. On the other hand, if the lifts in a conjugacy class have fixed points, then their projection is a fixed point class $\mathbf{F}$ of f. In the previous section, we defined the index of a fixed point class, which we will denote by $i(\mathbf{F})$. Thus, corresponding to some of the elements of $R(f_\pi)$ we have an integer $i(\mathbf{F})$. We assign the integer 0 to the remaining elements of $R(f_\pi)$, which are the ones with the property that the corresponding lifts have no fixed points. Thus, for each $\alpha \in \pi$ we have the equivalence class $\rho(\alpha) \in R(f_\pi)$ and an integer for that class, which we now denote by $i(\rho(\alpha))$. The main result of Wecken's paper [59] is his

REIDEMEISTER TRACE THEOREM. *Let $f : X \to X$ be a map of a finite polyhedron, then*

$$\mathcal{R}(f) = \sum_{R(f_\pi)} i\big(\rho(\alpha)\big)\rho(\alpha),$$

where by the summation over $R(f_\pi)$ we mean that the α are chosen so that each element of $R(f_\pi)$ is represented by exactly one $\rho(\alpha)$.

As a consequence of Wecken's theorem, we easily see that the definition of the Reidemeister trace includes the calculation of the Lefschetz number of f. If we add up the coefficients in the formula of the theorem, we are adding up the indices of all the fixed point classes of f, which equals $L(f)$, as we noted above. The Reidemeister trace also contains the information in the Nielsen number because, by its definition, the Nielsen number

$N(f)$ is nothing other than the number of nonzero terms in the expansion of $\mathcal{R}(f)$ that is furnished by this theorem.

It is thus evident from Wecken's theorem that $N(f)$ is less than or equal to the number of terms on the right-hand side, that is, to the cardinality of $R(f_\pi)$. That number, which may be infinite, is called the *Reidemeister number* of f and denoted $R(f)$. When $R(f)$ is finite, the inequality $N(f) \leqslant R(f)$ can give useful information about the Nielsen number because $R(f)$ is usually much easier to compute than is $N(f)$. For instance, if π is Abelian, we can seen from the previous section that $R(f)$ is just the order of the cokernel of $Id - f_\pi$. Furthermore, since the Reidemeister number will, under certain conditions, equal the Nielsen number, for instance for maps on Lie groups, it will then furnish the solution to Nielsen's "general fixed point problem".

A discussion of Wecken's proof of his Reidemeister trace theorem is contained in [22], on which this section is largely based, and we will summarize it next. Modern proofs of the Reidemeister trace theorem can be found in [26, 19].

Applying a technique from [24], as we have done before, we may assume the fixed point set of f is finite and that each fixed point lies in a maximal simplex. We may further assume that the lift $\tilde{f}$ has the same properties. Let x be a fixed point of f. Over the simplex s containing x there is the simplex $\tilde{s}$ in the $\mathbb{Z}[\pi]$ basis for the chains of $\widetilde{X}$ and a unique point $\tilde{x} \in \tilde{s}$ that projects to x. Now $\tilde{f}(\tilde{x})$ lies in the same fiber, so we can write $\tilde{f}(\tilde{x}) = \alpha\tilde{x}$ for a covering transformation that we represent by $\alpha \in \pi$. This implies that the fixed point x produces a contribution to the trace of $\tilde{f}_k$ that consists of some multiple of α, and so it contributes that same multiple of $\rho(\alpha)$ to $\mathcal{R}(f)$. Now suppose x' is some other fixed point of f, contributing some multiple of $\beta \in \pi$ to the trace of $\tilde{f}_k$, then x' is in the same fixed point class of f as x if and only if $\rho(\beta) = \rho(\alpha)$. In this way we can see how the fixed point classes of f are reflected in the definition of $\mathcal{R}(f)$ so that $\mathcal{R}(f)$ can be expressed in terms of elements in $R(f_\pi)$. For the more delicate matter of why the multiples of $\rho(\alpha)$ coming from the points in the corresponding fixed point class add up to the index of the class, we refer the reader to [22] and, for more detail, to [19].

The Reidemeister trace has become an important tool for the calculation of the Nielsen number, again see [22].

6. The Wecken property

The Nielsen fixed point theorem of Section 4 tells us that every map homotopic to $f : X \to X$ has at least $N(f)$ fixed points. So $N(f)$ is a lower bound for the number of fixed points for all maps in its homotopy class, but the theorem does not furnish any information on how good a lower bound it is. In particular, it does not tell us when the bound is "sharp" in the sense that some map g homotopic to f has exactly $N(f)$ fixed points. Wecken took up this problem, in another paper of 1942 [60] which included the following result:

WECKEN'S MANIFOLD THEOREM. *Let $f : M \to M$ be a map where M is an n-manifold, $n \geqslant 3$. Then there is a map g homotopic to f with exactly $N(f)$ fixed points.*

A special case of this theorem is closely related to Lefschetz's theorem that a map $f : X \to X$ with nonzero Lefschetz number has a fixed point. Suppose there were a map g

homotopic to f that was *fixed point free*, that is, g had no fixed points, then its Lefschetz number would be zero, as we discussed in Section 3 with regard to the Lefschetz–Hopf theorem. But it is obvious from the definition that homotopic maps have the same Lefschetz number. Thus, the Lefschetz fixed point theorem could have been stated: if $L(f) \neq 0$, then every map homotopic to f has a fixed point. Now, if the manifold M in Wecken's manifold theorem is simply-connected, then M is its own universal covering space and f has just one lift, the map itself, so there is a single fixed point class, of index $L(f)$. Consequently, $L(f) = 0$ implies $N(f) = 0$ when M is simply-connected and Wecken's theorem implies the converse to the form of the Lefschetz theorem in which we just restated it:

CONVERSE LEFSCHETZ THEOREM. *Let $f : M \to M$ be a map where M is a simply-connected n-manifold, $n \geqslant 3$. If $L(f) = 0$, then there is a fixed point free map homotopic to f.*

Turning now to a discussion of the proof of Wecken's manifold theorem, we may, as usual, assume that f has only finitely many fixed points. Wecken's proof required two techniques. The first we will refer to as the "combining technique". Let $Fix(f)$ denote the set of fixed points of a map f and suppose p and q are fixed points of f in the same fixed point class $\mathbf{F}$. The combining technique will homotope the map f to a map f' such that $Fix(f') = Fix(f) - \{q\}$. The fixed point class structure of f' will be the same as that of f except that, for the class $\mathbf{F}'$ corresponding to $\mathbf{F}$ under the homotopy as in Section 4, we have $\mathbf{F}' = \mathbf{F} - \{q\}$. Wecken was "combining", rather than "removing" fixed points, even though f' has one fewer fixed point than f, because the Lefschetz–Hopf theorem implies that the index of p as a fixed point of f' is the sum of the indices of the fixed points p and q of f.

The other technique that Wecken used *was* a "removal technique". Given a fixed point of a map $f : X \to X$ of a finite polyhedron that has a finite number of fixed points, each in a maximal simplex, suppose a fixed point p is of index zero. This means that, taking a small ball B_p containing p, the map $(i - f)_p : S_p \to S_p$ of the boundary of B_p is of degree zero. A map of the boundary of a ball to itself that is of degree zero can be extended to the entire ball, as a map back to the boundary. Using this fact, it is not difficult to homotope f to a map f', identical to f except in a neighborhood of p, such that $Fix(f') = Fix(f) - \{p\}$. This homotopy therefore "removes" the fixed point p without affecting the fixed point structure in any other way.

Applying the combining technique a finite number of times, Wecken homotopes the given map f to a map f' that has the same number of fixed point classes as f, but each fixed point class of f' consists of a single point. After the removal technique is applied to every fixed point of f' of index zero, Wecken has a map g homotopic to f with one fixed point for each essential fixed point class of f and no other fixed points, so g has exactly $N(f)$ fixed points.

It is clear why a removal technique was available to Wecken, for a map of any finite polyhedron and not just for maps of manifolds. We will next use some concepts from geometric topology to understand why, in a manifold, Wecken was able to combine fixed points in the same class.

It will help us, just as it helped Wecken, to interpret the fixed point class concept in a somewhat more geometric manner than that furnished by Nielsen's original definition in [45]. In fact, Nielsen himself pointed out this alternate definition immediately after he defined a fixed point class as the projection of the fixed points of a conjugacy class of lifts. If fixed points p and q of a map $f : X \to X$ are in the same fixed point class, we have a lift $\tilde{f} : \widetilde{X} \to \widetilde{X}$ of f and fixed points $\tilde{p}$ and $\tilde{q}$ of $\tilde{f}$ that project to p and q respectively. Take any path $\tilde{\zeta}$ in $\widetilde{X}$ from $\tilde{p}$ to $\tilde{q}$, then $\tilde{f}(\tilde{\zeta})$ is another such path and, since $\widetilde{X}$ is simply-connected, these two paths are homotopic by a homotopy $\widetilde{H} : I \times I \to \widetilde{X}$ keeping the endpoints $\tilde{p}$ and $\tilde{q}$ fixed. The projection ζ of $\tilde{\zeta}$ to X is a path from p to q that is homotopic to its image $f(\zeta)$ under f by a homotopy keeping the endpoints fixed, because following $\widetilde{H}$ by projection to X defines just such a homotopy. Conversely, suppose $p, q \in X$ are fixed points of f and there is a path ζ connecting them that is homotopic to $f(\zeta)$ by a homotopy keeping the endpoints fixed. Choosing a point $\tilde{p}$ that projects to p and a lift $\tilde{f}$ of f that fixes $\tilde{p}$, a homotopy lifting theorem for covering spaces lifts the given homotopy up to $\widetilde{X}$ and establishes the existence of a point $\tilde{q}$ that projects to q and is fixed by that same $\tilde{f}$. Therefore, p and q are in the same fixed point class.

Returning to the problem of creating a "combining technique", we have a map $f : M \to M$ on a manifold and we assume f has a finite number of fixed points. We have fixed points p and q of f that are in the same fixed point class. Therefore there is a path ζ from p to q and a homotopy $H : I \times I \to X$ between ζ and $f(\zeta)$ that keeps the endpoints fixed. We may assume that ζ does not intersect any of the finite number of fixed points of f other than p and q. We will want to assume that M is of dimension at least 5, so it is clear that by general position (see, for instance, [49]), we may replace ζ by an arc, still avoiding other fixed points of f, and then modify the map by a homotopy, without changing the fixed point structure, so that $f(\zeta)$ is also an arc, disjoint from ζ except at the endpoints, that is, $\zeta \cup f(\zeta)$ is an embedded simple closed curve, and $f(\zeta)$ is still homotopic to ζ by a homotopy keeping the endpoints fixed. The dimension of M is high enough so that general position allows us to replace the singular disc in M determined by the homotopy by an embedded disc.

A result from geometric topology tells us that a regular neighborhood of an embedded disc in an n-manifold is homeomorphic to $\mathbb{R}^n$. Therefore we may view the map f in a neighborhood of the arc ζ as a map of a subset of $\mathbb{R}^n$ that maps into $\mathbb{R}^n$ and that has just two fixed points, p and q, the endpoints of the arc ζ. Without loss of generality, we can take ζ to be the line segment in $\mathbb{R}^n$ between p and q. Let B be an n-ball in R^n containing ζ on its interior such that, for each point x of $B - \{p\}$, the ray from p through x intersects the boundary S of B in a single point that we will denote by $\hat{x}$. We write $x = tp + (1-t)\hat{x}$ for some $t \in I$ and define a map f' on B by letting $f'(p) = p$ and setting $f'(x) = tp + (1-t)f(\hat{x})$, otherwise. Since f has no fixed points on S, the map f' has a fixed point only at p and, since f' agrees with f on S, we can extend f' to the rest of M thus completing the combining technique in this high-dimensional setting.

Wecken's combining technique was different from the one we have just sketched, and it could be applied to manifolds of dimension as low as 3. The purpose of the discussion above was to make it believable that such a technique could be developed, but without going into the more complicated procedures that Wecken actually used.

Wecken's combining technique was not limited to maps of manifolds but could be applied to a broader class of finite polyhedra. A (connected) polyhedron X is *three-dimensionally connected* if there is no subpolyhedron A of dimension less than or equal to one such that $X - A$ is disconnected. Wecken proved in [60] that the Nielsen number is a sharp lower bound for the number of fixed points in a homotopy class of maps on a polyhedron that is three-dimensionally connected and locally three-dimensionally connected.

We will not discuss Wecken's combining technique but instead describe a combining technique developed by Jiang, who used it in a paper of 1980 [28] to obtain a substantial improvement of Wecken's result. A *local cut point* of a polyhedron X is a point x for which there is a connected neighborhood U of x in X such that $U - \{x\}$ is disconnected.

WECKEN–JIANG THEOREM. *Let $f : X \to X$ be a map of a finite polyhedron X such that X has no local cut points and X is not a 2-manifold, with or without boundary, then there is a map g homotopic to f such that g has exactly $N(f)$ fixed points.*

A space X is said to have the *Wecken property* if, for every map $f : X \to X$, the Nielsen number $N(f)$ is a sharp lower bound for the number of fixed points in the homotopy class of f. Thus the Wecken–Jiang theorem tells us that a finite polyhedron that has no local cuts points and is not a surface has the Wecken property.

The reason Jiang had to exclude surfaces in his hypotheses was that his combining technique requires the presence in X of a 1-simplex that is the face of at least three 2-simplices. An n-dimensional polyhedron, $n \geqslant 3$, has such a 1-simplex because a 3-simplex can be triangulated to contain it. Thus the only finite polyhedra without local cut points that fail to have such a 1-simplex are those in which each 1-simplex is on the boundary of at most two 2-simplices, that is, the surfaces.

To describe Jiang's technique, we begin with a map $f : X \to X$, now on a polyhedron X satisfying the hypotheses of the theorem. We assume f has a finite number of fixed points, and we have fixed points p and q of f that are in the same fixed point class. Thus there is a path ζ from p to q and a homotopy between ζ and $f(\zeta)$ that keeps the endpoints fixed. Since there are only a finite number of fixed points and there are no local cut points, we can assume that there are no fixed points of f on ζ except for its endpoints. We can further assume that ζ is an arc by first making its self-intersections finite and then removing them one at a time, by general positions if the intersection is in a simplex of dimension at least three. Intersections in dimension 2 can be eliminated by sliding part of ζ off its endpoint. By hypothesis, we have a 1-simplex, call it s, in X that is on the boundary of at least three 2-simplices. We modify ζ so that it passes through s. To do this, we use an arc γ from $\zeta(t_0) = \gamma(0)$, for some $0 < t_0 < 1$, to s and then replace a small portion of ζ near $\zeta(t_0)$ by an arc that parallels γ to s, loops around $\gamma(1)$ and then returns paralleling γ.

Now suppose δ is a path from p to q. Jiang calls δ *special* (with respect to ζ) if $\delta(t) \neq \zeta(t)$ for all $0 < t < 1$. Note that since ζ contains fixed points only at its endpoints, the path $f(\zeta)$ is special with respect to ζ. Two special paths are *specially homotopic* if they are homotopic through special paths. Jiang's key result states that if two paths δ_1 and δ_2 that are both special with respect to ζ are homotopic, then they are specially homotopic. We will not discuss the proof of this result except to say that the fact that ζ passes through the 1-simplex s that is the face of at least three 2-simplices is crucial in constructing the maps h_t of the homotopy so that they are special.

Jiang constructs a path δ that is special with respect to ζ by just modifying the parameter of ζ, specifically he sets $\delta(t) = \zeta(t - \varepsilon \sin t\pi)$ for a small ε. Now this δ is homotopic to $f(\zeta)$, because ζ is, so we now know that they are specially homotopic, by a homotopy we denote by $\{h_t\}$. Letting $Z = \zeta(I)$ and setting $h'_t = h_t\zeta^{-1} : Z \to X$ defines a homotopy, between the map $\delta\zeta^{-1}$ and the restriction $f|Z$ of f to Z, with the property that the fixed point set of each h'_t consists of p and q.

A homotopy $\{h_t\}$ with the same fixed point set for each h_t was called by Jiang a *special homotopy*. He proved a "special homotopy extension theorem" that, in the present case, tells us that, since $f|Z$ extends to X, the map $\delta\zeta^{-1}$ also does so, by means of a special homotopy. Consequently, the fixed points of the extension of $\delta\zeta^{-1}$ are precisely those of f.

We still want to combine p and q, but now we have a map, which we will continue to call f, with the property that, on an arc ζ connecting them, the map is very close to the identity map, since on Z the map is defined to be $\delta\zeta^{-1}$ and ε can be made as small as desired. A map sufficiently close to the identity map is a *proximity map* and it is not difficult to combine fixed points of such a map. The arc ζ determines a chain of maximal simplices $s_1, s_2, \ldots, s_k$, where $p \in s_1$ and $q \in s_k$. Taking $r \in s_{k-1} \cap \zeta$, we can modify f on $s_{k-1} \cup s_k$ so that q is no longer a fixed point but r is. Furthermore, the map is still a proximity map on the portion of ζ that connects p to r. By repeating this construction, we replace the fixed point q by a fixed point, again call it r, that lies in s_1, as p does. Furthermore, the map is a proximity map on an arc in s_1 connecting them. We are now in the same Euclidean setting that the techniques of geometric topology obtained for us in the high-dimensional manifold case above so, as there, we can easily combine p and r.

Since the Wecken–Jiang theorem extends Wecken's manifold theorem to other polyhedra, it implies a corresponding extension of the converse Lefschetz theorem. That is, if X is a simply-connected polyhedron that has no local cut points and is not a surface and $f : X \to X$ is a map with $L(f) = 0$, then there is a fixed point free map homotopic to f.

7. The fixed point theory of surfaces

Nielsen himself was concerned with what is now called the Wecken property, that we discussed in the previous section. In 1927 he wrote "one might conjecture that any fixed point class might be made *either* to vanish *or* reduced to a single point by continuous deformation" [45]. However, the setting of [45] was that of surfaces, which are the finite polyhedra without local cut points specifically excluded from the Wecken–Jiang theorem of 1980. Thus, more than 50 years after he made his conjecture, it was still open.

The paper [45] was not concerned with all continuous functions, but only with homeomorphisms. Thus Nielsen's conjecture was that any homeomorphism $f : M \to M$ of a surface could be deformed so that there were only $N(f)$ fixed points. The surfaces in [45] were required to be closed and orientable, but Nielsen wrote in the introduction to the paper that this restriction was "for the sake of simplicity ... it is not difficult to see how one may extend the methods to other cases of two-dimensional manifolds". Since the context is that of homeomorphisms, a "deformation" would mean through homeomorphisms, that is an isotopy, so Nielsen's conjecture was that any homeomorphism f of a surface was isotopic to a homeomorphism g with exactly $N(f)$ fixed points. Jiang announced in 1981 [31] that Nielsen's conjecture was correct for all closed surfaces, orientable or not.

For the few surfaces that are of non-negative Euler characteristic, Jiang could verify Nielsen's conjecture on a case-by-case basis. The main part of Jiang's proof depended on a classification of the self-homeomorphisms of surfaces of negative Euler characteristic by William Thurston that was available in unpublished form. (It was not published by Thurston until 1988 [56].) Such a classification was pursued by Nielsen in [45] and subsequent papers, using many of the tools that Thurston was to discover independently many years later, see [20] and the introduction to [56]. One of the two basic types of homeomorphisms in Thurston's classification are the periodic homeomorphisms. A map $f : X \to X$ can be iterated as $f^2, f^3, \ldots$ by setting $f^2(x) = f(f(x))$ and, in general, letting $f^k(x) = f(f^{k-1}(x))$. A homeomorphism f is *periodic* if some iterate f^m is the identity function. The other basic type of homeomorphism is called *pseudo-Anosov* and we will not attempt to define it here. Thurston's classification theorem states that, for any homeomorphism $f : M \to M$ of a surface, there is a (possibly empty) set of disjoint simple closed curves $\mathcal{C}$ in the interior of M such that the restriction of f to each component of $M - \mathfrak{N}(\mathcal{C})$ is either periodic or pseudo-Anosov, where $\mathfrak{N}(\mathcal{C})$ is an f-invariant tubular neighborhood of $\mathcal{C}$.

Jiang's proof begins by putting the periodic and pseudo-Anosov homeomorphisms on the components of $M - \mathfrak{N}(\mathcal{C})$ in a standard form in which all their fixed point classes are connected. Furthermore, the homeomorphism can be made periodic on the boundary of $\mathfrak{N}(\mathcal{C})$. Jiang then puts the homeomorphism into a standard form on $\mathfrak{N}(\mathcal{C})$ so that he has a homeomorphism $f : M \to M$ all of the fixed point classes of which are connected. Furthermore, the fixed point classes are themselves manifolds, of dimension less than or equal to 2, and Jiang is able to list all the possible cases with regard to the behavior of f in a neighborhood of the fixed point class. Then, for each case, he describes how to deform f by an isotopy so that the fixed point class is removed, if it is of index zero, or, if it is essential, reduced to a single point, thus completing the verification of Nielsen's conjecture.

Jiang did not publish the proof of the Nielsen conjecture until 1993, in a paper with his student Jianhan Guo [34]. One cause of the delay was the unsatisfactory state of the conjecture in the case that M is a surface with nonempty boundary ∂M. As he had stated it, Nielsen's conjecture could not be true in that case, but for a rather shallow reason. Consider a reflection f of the unit disc about an axis, then certainly $N(f) = 1$. However, a homeomorphism of the disc must map the boundary circle to itself and since that is of degree -1, it must have at least two fixed points (by, for instance, Jiang's Lie group theorem of Section 4). The problem is that a homeomorphism of a surface with boundary is a map of pairs $f : (M, \partial M) \to (M, \partial M)$. We have seen in Section 3 that Bowszyc had extended the Lefschetz theory to maps of pairs, but there was no such theory of Nielsen type. The classification theorem of Thurston is valid for surfaces with boundary, so it seemed that the tools for establishing a form of Nielsen's conjecture in this setting were available, provided that the correct notion of Nielsen number could be found.

This problem was solved by Helga Schirmer in a paper of 1986 [51]. Let $f : (X, A) \to (X, A)$ be map of pairs, where X is a finite polyhedron and A is a subpolyhedron. It is easy to see from the geometric definition of fixed point classes that we discussed in the previous section that each fixed point class of $\bar{f} : A \to A$, the restriction of f, is contained in a fixed point class of $f : X \to X$. An essential fixed point class of the map $f : X \to X$ is called an *essential common fixed point class* if it contains a fixed point class of $\bar{f} : A \to A$ that is essential. Letting $N(f, \bar{f})$ denote the number of essential common fixed point classes of f, Schirmer defined $N(f; X, A)$, the *relative Nielsen number* of the map of pairs $f : (X, A) \to (X, A)$ by

$$N(f; X, A) = N(f) + N(\bar{f}) - N(f, \bar{f}).$$

This concept extends the usual Nielsen number because certainly $N(f; X, A) = N(f)$

when A is the empty set. She proved that $N(f; X, A)$ is a lower bound for the number of fixed points of all maps of pairs homotopic to f by a homotopy of maps of pairs.

Jiang was then able to use Thurston's classification to extend his solution to the Nielsen conjecture to all surfaces, with or without boundary, in [34] as

JIANG'S SURFACE HOMEOMORPHISM THEOREM. *Let M be a surface, with or without boundary, and let $f : M \to M$ be a homeomorphism. There is a homeomorphism g isotopic to f with exactly $N(f; M, \partial M)$ fixed points.*

Although the relative Nielsen number was used by Jiang only in the case of a pair (X, A), where X is a surface and A its boundary, there are many other interesting fixed point questions to which relative Nielsen theory can be applied. The survey [52] describes the first applications to be published, but much more has been done since that time in this very active area of fixed point theory.

Jiang was successful in settling the Nielsen conjecture concerning homeomorphisms of surfaces, but there was still no information on the sharpness of the Nielsen number as a lower bound for the number of fixed points in a homotopy class for maps in general. The Wecken–Jiang theorem states that all finite polyhedra without local cut points, except surfaces, have the Wecken property. Do surfaces also have the Wecken property? It was known that the seven surfaces with non-negative Euler characteristic do have the Wecken property, but nothing was known about the rest.

Then, in 1984, Jiang published a paper [32] containing an example of a map $f : M \to M$ on a surface such that $N(f) = 0$ but every map homotopic to f has at least two fixed points. In a paper published the following year [33], Jiang showed that he could embed a modified version of that example in any surface of negative Euler characteristic. Thus, with regard to the Wecken property, he could state that a surface M, with or without boundary, has the Wecken property if, and only if, the Euler characteristic of M is non-negative.

The surface in Jiang's original example from [32] is the disc with two open discs removed, often called the "pants surface" since it is homeomorphic to a surface in the shape of such a garment, and therefore denoted by P. Jiang's map $f : P \to P$ had two fixed points, of index 1 and -1, and they were evidently in the same fixed point class, so $N(f) = 0$. If P had the Wecken property, there would be a fixed point free g map homotopic to f.

To prove that no such fixed point free map g exists, Jiang assumed there was one and sought to establish a contradiction. He constructed loops w_0, w_1, w_2 in P by taking an arc from a base point to each of the three components of the boundary of P and defining each w_i to be the loop that follows the arc to the boundary component, goes once around the boundary component and then returns along the arc to the base point. The fixed points of f are on the interior of P, so it is easy to choose w_i that avoid them. The loops are oriented so that the loop $w_1 w_2$ can be deformed through P to w_0, thus $w_1 w_2 w_0^{-1}$ represents the identity element of the fundamental group $\pi_1(P)$. Therefore there is a homotopy $h_t : I \to P$ shrinking $w_1 w_2 w_0^{-1}$ to the constant loop. Letting Δ denote the diagonal in $P \times P$, the map $\hat{g}$ defined by $\hat{g}(x) = (x, g(x))$ maps each w_i to a loop in $P \times P - \Delta$ that represents an element $\tau_i \in \pi_1(P \times P - \Delta)$, the *braid group* (of pure 2-braids) on P. Since g is fixed point free, we have the homotopy $g \circ h_t : I \to P \times P - \Delta$ that shows us that $\tau_1 \tau_2 \tau_0^{-1}$ is the identity element of $\pi_1(P \times P - \Delta)$. Letting $\sigma_i \in \pi_1(P \times P - \Delta)$ be the element represented by the loop $\hat{f}(w_i)$, where $\hat{f}(x) = (x, f(x))$, Jiang showed that, since f and g

are homotopic, σ_i and τ_i are conjugate in $\pi_1(P \times P - \Delta)$, that is, $\tau_i = u_i \sigma_i u_i^{-1}$ for some u_i. Since $\tau_1 \tau_2 = \tau_0$ in $\pi_1(P \times P - \Delta)$, Jiang obtained from the existence of the fixed point free map g homotopic to f the following equation in the braid group:

$$u_1 \sigma_1 u_1^{-1} u_2 \sigma_2 u_2^{-1} = u_0 \sigma_0 u_0^{-1}. \tag{$*$}$$

The σ_i depend on the map f which Jiang had explicitly defined, so they can be expressed in terms of the generators of the braid group. He defined a homomorphism $\theta : \pi_1(P \times P - \Delta) \to G$, where G is the group with two generators, α and β, subject to the single relation $\alpha^2 = 1$, and he calculated that $\theta(\sigma_0) = \beta^3$, $\theta(\sigma_1) = 1$ and $\theta(\sigma_2) = \alpha\beta^2\alpha\beta$. Thus the equation $(*)$ leads to the equation

$$v_2 \alpha \beta^2 \alpha \beta v_2^{-1} = v_0 \beta^3 v_0^{-1} \tag{$**$}$$

in G, where $v_i = \theta(u_i)$. But $(**)$ implies that $\alpha\beta^2\alpha\beta$ is conjugate to β^3 in G, which is readily seen to be false. Thus there is no equation $(**)$ in G, which contradicts the equation $(*)$ in $\pi_1(P \times P - \Delta)$ that must exist if the fixed point free map g does. In this way, Jiang proved that there is no fixed point free map homotopic to the map f of his example. There could not be any map homotopic to f with a single fixed point either because it would be of index zero and so could be removed by a technique we described in the previous section. Consequently every map homotopic to f has at least two fixed points, even though $N(f) = 0$, so P does not have the Wecken property.

More discoveries about the Wecken property on surfaces followed Jiang's work and the study of this property continues to be an important part of fixed point theory (see [10]).

8. Some additional topics

In this final section, we will discuss briefly some of the many topics that are of current interest but that did not come up earlier. The topics have been chosen to illustrate the diversity of subject matter that has evolved from the sort of fixed point theory we have been considering. In recent years, this branch of topology has come to be known as "Nielsen theory". As we have seen, Nielsen's ideas had a profound impact on the subsequent development of fixed point theory, so it is appropriate that this area has become associated with his name. More significantly, a very substantial portion of the subject is concerned with the further development of those same ideas, through the characteristic methodology of Nielsen theory, which I will next describe in very general terms.

Nielsen theory is concerned with finding a lower bound, that is valid for maps in an appropriate homotopy class, for the number of solutions to some equation. The solutions are partitioned into equivalence classes, much like the fixed point classes of Nielsen's paper [45], and a criterion for the *essentiality* of a class is established. The Nielsen number, call it N, is defined to be the number of essential classes. Once the suitable Nielsen number concept has been found, there are two principal lines of inquiry. One of these is the development of techniques for the computation of N. For instance, some class of spaces may allow a relatively straightforward calculation of N, as is the case for the classical Nielsen number $N(f)$ if f is a map on a Lie group (see Section 4). The other line of inquiry concerns whether, as a lower bound, N is sharp, that is, whether there are maps in every homotopy class for which the number of solutions is exactly N. In the setting of classical Nielsen fixed point theory, that was the subject of the previous two sections, that is, the Wecken property.

We will not attempt to cover these additional topics in any detail but will just state briefly what they are and, to some extent, describe the work that initiated them. In general, references will be limited to the first papers in that line of inquiry. An adequate exposition of the subsequent development of these topics, which in all cases is substantial, must await another occasion.

In Section 3 we discussed the fact that intersection theory was the setting for the work of Lefschetz that contained his fixed point theorem and that, in that setting, it was natural to study coincidences. Specifically, in [40] Lefschetz considered maps $f, g : M \to N$ between oriented manifolds of the same dimension and found conditions for the existence of coincidences of the maps, that is, for solutions $x \in M$ to the equation $f(x) = g(x)$. In 1954, Wolfgang Franz [18] described a Nielsen theory of coincidences for maps of manifolds. Extending the geometric form of the Nielsen equivalence relationship that Wecken found so useful, coincidences x_1 and x_2 of maps $f, g : M \to N$ are *equivalent* if there is a path ζ in M from x_1 to x_2 such that the paths $f(\zeta)$ and $g(\zeta)$ are homotopic in N by a homotopy keeping the endpoints fixed. A "coincidence index", that is a local version of the Lefschetz coincidence number $L(f, g)$ defined in Section 3, may be used to define a Nielsen coincidence number $N(f, g)$ that is a lower bound for the number of solutions for the equation $f'(x) = g'(x)$ among all maps f' homotopic to f and g' homotopic to g. Although, as we noted, at its start coincidence theory seemed a somewhat unusual topic for topology, it has developed a significant body of information which continues to grow rapidly.

If the coincidence equation $f(x) = g(x)$ is specialized by taking g to be the identity map, we have the fixed point equation $f(x) = x$. On the other hand, if we take $f, g : X \to Y$ to be a map, where g is the constant map, say $g(x) = c$ for all $x \in X$, a solution to the coincidence equation is a point $x \in f^{-1}(c)$. The study of the sets $f^{-1}(c)$ for maps between orientable manifolds of the same dimension is related to the degree of Brouwer that we discussed in Section 2 because the absolute value of the degree can furnish information on the number of points in $f^{-1}(c)$. In a paper of 1930 concerned with this topic [25], Hopf introduced a Nielsen theory modelled, as he states, on the 1927 paper of Nielsen [45]. Points x_1 and x_2 in $f^{-1}(c)$ are equivalent if there is a path ζ in X from x_1 to x_2 such that the loop $f(\zeta)$ is contractible in Y by a contraction keeping c fixed. His notion of essentiality of an equivalence class is closely tied to Brouwer's degree. Although the degree theory of Hopf's paper, which applied even to maps of non-orientable manifolds, became an established part of algebraic topology, the part related to Nielsen's work did not. Then, in 1967, as part of his work on Nielsen coincidence theory, Robin Brooks [4] studied solutions to the equation $f(x) = c$, which he called *root theory*. Brooks independently developed Nielsen root theory in a more general context than Hopf had considered and, since that time, root theory has been recognized as an important part of Nielsen theory. (An exposition of Brooks's root theory can be found in [35].)

In principle, Lefschetz's intersection theoretic approach did not limit him to the study of single-valued functions. If f is a function that associates to each point x in a space X a subset of the space X, then the *graph* of f can be defined as

$$\Gamma_f = \{(x, y) \in X \times X \colon y \in f(x)\}.$$

A point (x, y) in the intersection of Γ_f and the diagonal in $X \times X$ corresponds to a point $x \in X$ with $x \in f(x)$, which is called a *fixed point* of the multivalued function f. But the history of fixed point theory for multivalued functions really begins with a paper of Samuel Eilenberg and Dean Montgomery in 1946 [15]. Let X be a finite n-dimensional polyhedron and f a function on X such that, for each $x \in X$, its image $f(x)$ is a nonempty closed subset of X that is acyclic, that is, its reduced rational homology is trivial. Suppose further that f is *upper semi-continuous* which means that for every open set V in X containing $f(x)$ there is a neighborhood U of x such that $f(U) \subset V$. Let $p, q : \Gamma_f \to X$ be the projections $p(x, y) = x$ and $q(x, y) = y$. Since, for each $x \in X$, the set $p^{-1}(x)$ is acyclic because it is homeomorphic to $f(x)$, a theorem of Vietoris of 1929 [57] implies that, for all k, the homomorphism $p_k : H_k(\Gamma_f) \to H_k(X)$ is an isomorphism. Eilenberg and Montgomery defined a Lefschetz number $L(f)$ for the multivalued function f by setting

$$L(f) = \sum_{k=0}^{n} (-1)^k \, trace\big(q_k \circ (p_k)^{-1}\big)$$

and proved that if $L(f) \neq 0$, then f has a fixed point, that is, $x \in f(x)$ for some $x \in X$.

The first Nielsen theory for multivalued functions appeared in a paper of Schirmer [50] of 1975. She considered upper semi-continuous functions ϕ taking points of a finite polyhedron X to *small* subsets of X, that is, such that $\phi(x)$ lies in the star of a vertex for each $x \in X$. Letting $N(\phi) = N(f)$ for a suitable single-valued approximation $f : X \to X$ to ϕ, she found conditions, in particular for multivalued functions that are acyclic-valued as in [15], so that $N(\phi)$ is a lower bound for the number of fixed points in an appropriate sense. Subsequent work on the fixed point theory of multivalued functions has extended the types of such functions that can be analyzed.

Given a map $f : X \to X$, a fixed point of the iterate f^m is called a *periodic point* (*of period* m). The application of fixed point theory to periodic points began in a paper of Brock Fuller in 1953 [17]. Fuller used the Lefschetz fixed point theorem to prove that if $h : X \to X$ is a homeomorphism of a finite polyhedron with nonzero Euler characteristic, then h has a periodic point. In fact, he proved more, namely that if R_k is the k-th Betti number, then h has a periodic point of period no larger than the larger of the sum of the R_k for k odd and the sum for k even. The argument depends on the fact that, if there is no such periodic point then, by the Lefschetz theorem, all the Lefschetz numbers $L(h)$, $L(h^2)$ and so on must be zero. This condition can be expressed in terms of the eigenvalues of the induced homology homomorphisms h_k^m, which are all nonzero since h is a homeomorphism. An algebraic argument then demonstrates that the condition implies that the Euler characteristic of X is zero.

The Nielsen theory of periodic points began much more recently, in a book published in 1983 by Jiang [30]. The purpose of such a theory should be to find a lower bound for the number of periodic points of a given period among all maps homotopic to the given map. But there is an issue that arises as soon as one considers the problem. For example, a fixed point of a map f is also a fixed point of any iterate of f. Thus, for instance, by the number of periodic points of period 2 do we include the fixed points or only consider the other points with the property $f^2(x) = x$? Jiang dealt with this problem by defining *two* Nielsen numbers for periodic points. One of these Nielsen numbers, now generally denoted by $N\Phi_m(f)$, is a lower bound for the number of periodic points of period m among all maps homotopic to f. Thus this number counts fixed points as periodic points of all periods. On the other hand, by a periodic point *of period exactly m* is meant a fixed point of f^m that is not a fixed point of f^j for any $j < m$. The Nielsen number $NP_m(f)$ is a lower bound for the number of periodic points of period exactly m among all maps homotopic to f. For a thorough exposition of this very active area of Nielsen theory, see [23].

The purpose of Brouwer's use of simplicial approximation and related combinatorial techniques was to eliminate the restriction, found in Poincaré's work, that maps had to be differentiable. In a sense then, a paper Jiang published in 1981 [29], in which he proved Wecken's manifold theorem (see Section 6) in the differentiable setting, reversed the direction of the development of fixed point theory. Jiang's theorem states that if M is a smooth manifold of dimension at least three and $f: M \to M$ is a smooth map, then f can be smoothly homotoped to a map g with exactly $N(f)$ fixed points. The principal tool of his proof was the smooth version of the "Whitney Lemma". By transversality, the fixed points of f can be made transverse, so each is of index (in the sense of Section 3) either 1 or -1. If two equivalent transversal fixed points have different indices, the Whitney lemma cancels them. After applying the lemma as much as possible, only essential fixed point classes remain and each consists just of a number k of points all with the same index. Jiang then deformed the map to create, within the class, $k + 1$ new fixed points of which k are transversal of sign opposite those already in the class and there is one more fixed point whose index makes the total index of those added equal to zero so that the construction is possible. A further application of the Whitney Lemma reduces the fixed point class to that single point and completes the proof. Thus, for this part of fixed point theory, the smooth category behaves just like the continuous one. On the other hand, Michael Shub and Dennis Sullivan had demonstrated in 1974 [54] that it is not always the case that restricting to smooth maps has no affect in fixed point theory. They proved that if $f: M \to M$ is a smooth map of a smooth manifold and the set of Lefschetz numbers $\{L(f^k)\}$ of all the iterates of f is unbounded, then there is an infinite set of points $x \in M$ that are periodic points of f. On the other hand, they defined a continuous function $f: S^2 \to S^2$, of degree 2, with the property that the only fixed points of f^k, for any k, are the poles. Thus $\{L(f^k) = 1 + 2^k\}$ is unbounded in this case and yet there are just two periodic points for f, which could not happen if f were smooth. The question of the distinction, or lack of it, between the smooth and continuous categories in the various parts of Nielsen theory continues to be of interest.

An immediate consequence of Jiang's smooth version of Wecken's manifold theorem is a smooth version of the converse Lefschetz theorem: if $f: M \to M$ is a smooth map on a smooth, simply-connected manifold and $L(f) = 0$, then there is a smooth fixed point free map smoothly homotopic to f. A smooth map on a smooth manifold is one that preserves the underlying structure on the manifold. In the same way, if a group G acts on a manifold M so that $g \cdot x$ is defined for $g \in G$ and $x \in M$, then M is called a *G-manifold* and a *G-map* is one that preserves that underlying structure because it obeys the rule $f(g \cdot x) = g \cdot f(x)$. In 1984, D. Wilczyński [61] extended the converse Lefschetz theorem to G-maps in the following way. A subgroup H of G is an *isotropy subgroup* if there exists some $x \in M$ such that $g \cdot x = x$ if and only if $g \in H$. Let H be an isotropy subgroup of G, then M^H denotes the submanifold of M consisting of the points that are fixed by H, that is $h \cdot x = x$, for all $h \in H$. The restriction of a G-map f takes M^H to itself and we denote that restriction by f^H. Letting NH denote the normalizer of H in G, its *Weyl group* WH is defined by $WH = NH/H$. Wilczyński proved that if, for every isotropy subgroup H of G with finite Weyl group, M^H is simply-connected and of dimension at least 3 and the Lefschetz number $L(f^H) = 0$, then there exists a fixed point free G-map that is G-homotopic to f. The Nielsen theory of G-maps began in 1988 with a paper of Edward Fadell and Peter Wong [16]. By using appropriate Nielsen numbers in place of the Lefschetz numbers of Wilczyński's paper, they were able to obtain a converse Lefschetz theorem for G-maps that did not require that the M^H be simply-connected.

The converse Lefschetz theorem of Section 6 has been extended in yet another way. If $H = \{h_t\}: X \times I \to X$ is a homotopy, then the fixed points of H are the fixed points of the maps h_t for all t, that is, by a *fixed point* of H is meant a point $(x, t) \in X \times I$ such that $H(x, t) = x$. A converse Lefschetz theorem in this setting would furnish conditions under which, given such a homotopy H between fixed point free maps, there is a homotopy without fixed points that is homotopic to H through maps from $X \times I$ to X. Dončo Dimovski and Ross Geoghegan considered this problem in a paper of 1990 [12] whose title "One-parameter fixed point theory" is the name given to this new branch of Nielsen theory.

Acknowledgements

I thank Ross Geoghegan, Evelyn Hart, Philip Heath and Boju Jiang for their comments. I especially thank Helga Schirmer who followed the work closely and offered many suggestions as this was being written. In addition, she made translations and furnished other assistance in dealing with the German-language literature.

Bibliography

[1] P. Bohl, *Über die Bewegung eines mechanischen Systems der Höhe einer Gleichgewichtslage*, J. Reine Angew. Math. **127** (1904), 179–276.
[2] K. Borsuk, *Sur les rétractes*, Fund. Math. **17** (1931), 152–170.
[3] C. Bowszyc, *Fixed point theorems for the pairs of spaces*, Bull. Acad. Polon. Sci. **16** (1968), 845–850.
[4] R. Brooks, *Coincidences, roots and fixed points*, Doctoral dissertation, University of California, Los Angeles (1967).
[5] L. Brouwer, *On continuous one-to-one transformations of surfaces into themselves*, Proc. Kon. Nederl. Akad. Wetensch. Ser. A **11** (1909), 788–798.
[6] L. Brouwer, *Über Abbildung der Mannigfaltigkeiten*, Math. Ann. **71** (1912), 97–115.
[7] L. Brouwer, *Über die Minimalzahl der Fixpunkte bei den Classen von eindeutigen stetigen Transformationen der Ringflächen*, Math. Ann. **82** (1921), 94–96.
[8] F. Browder, *Fixed point theory and nonlinear problems*, Bull. Amer. Math. Soc. **9** (1983), 1–39.
[9] R. Brown, *The Lefschetz Fixed Point Theorem*, Scott, Foresman, Glenview (1971).
[10] R. Brown, *Wecken properties for manifolds*, Nielsen Theory and Dynamical Systems, C. McCord, ed., Contemp. Math. vol. 152, Amer. Math. Soc., Providence, RI (1993), 9–22.
[11] J. Dieudonné, *A History of Algebraic and Differential Topology 1900–1960*, Birkhäuser, Basel (1989).
[12] D. Dimovski and R. Geoghegan, *One-parameter fixed point theory*, Forum Math. **2** (1990), 125–154.
[13] A. Dold, *Fixed point index and fixed point theorem for Euclidean neighborhood retracts*, Topology **4** (1965), 1–8.
[14] J. Dugundji and A. Granas, *Fixed Point Theory*, Polish Scientific Publishers, Warsaw (1982).
[15] S. Eilenberg and D. Montgomery, *Fixed point theorems for multi-valued transformations*, Amer. J. Math. **58** (1946), 214–222.
[16] E. Fadell and P. Wong, *On deforming G-maps to be fixed point free*, Pacific J. Math. **132** (1988), 277–281.
[17] B. Fuller, *The existence of periodic points*, Ann. Math. **57** (1953), 229–230.

[18] W. Franz, *Mindestzahl von Koinzidenzpunkten*, Wiss. Z. Humboldt-Univ. Berlin Math.-Nat. Reihe **3** (1954), 439–443.
[19] R. Geoghegan, *Nielsen fixed point theory*, Handbook of Geometric Topology, R. Daverman and R. Sher, eds, Elsevier, Amsterdam, to appear.
[20] J. Gilman, *On the Nielsen type and the classification for the mapping class group*, Adv. Math. **40** (1981), 68–96.
[21] J. Hadamard, *Sur quelques applications de l'indice de Kronecker*, Appendix to: J. Tannery, La Théorie des Fonctions d'une Variable, Paris (1910).
[22] E. Hart, *The Reidemeister trace and the calculation of the Nielsen number*, Nielsen Theory and Torsion Methods, J. Jezierski, ed., Banach Center Publications, to appear.
[23] P. Heath, *A survey of Nielsen periodic point theory (fixed n)*, Nielsen Theory and Torsion Methods, J. Jezierski, ed., Banach Center Publications, to appear.
[24] H. Hopf, *Über die algebraische Anzahl von Fixpunkten*, Math. Z. **29** (1929), 493–524.
[25] H. Hopf, *Zur Topologie der Abbildungen von Mannigfaltigkeiten II*, Math. Ann. **102** (1930), 562–623.
[26] S. Husseini, *Generalized Lefschetz numbers*, Trans. Amer. Math. Soc. **272** (1982), 247–274.
[27] B. Jiang, *Estimation of the Nielsen numbers*, Chinese Math. **5** (1964), 330–339.
[28] B. Jiang, *On the least number of fixed points*, Amer. J. Math. **102** (1980), 749–763.
[29] B. Jiang, *Fixed point classes from a differentiable viewpoint*, Fixed Point Theory: Proceedings, Sherbrooke, Quebec, Lecture Notes in Math. vol. 886, Springer, Berlin (1981), 163–170.
[30] B. Jiang, *Lectures on Nielsen Fixed Point Theory*, Contemporary Math. vol. 14, Amer. Math. Soc., Providence, RI (1983).
[31] B. Jiang, *Fixed points of surface homeomorphisms*, Bull. Amer. Math. Soc. **5** (1981), 176–178.
[32] B. Jiang, *Fixed points and braids*, Invent. Math. **75** (1984), 69–74.
[33] B. Jiang, *Fixed points and braids II*, Math. Ann. **272** (1985), 249–256.
[34] B. Jiang and J. Guo, *Fixed points of surface diffeomorphisms*, Pacific J. Math. **160** (1993), 67–89.
[35] T. Kiang, *The Theory of Fixed Point Classes*, Springer, Berlin (1989).
[36] L. Kronecker, *Über Systeme von Funktionen mehrerer Variablen*, Monatsh. Berlin Akad. (1869), 159–193 and 688–698.
[37] S. Lefschetz, *Continuous transformations of manifolds*, Proc. Nat. Acad. Sci. U.S.A. **9** (1923), 90–93.
[38] S. Lefschetz, *Intersections of complexes on manifolds*, Proc. Nat. Acad. Sci. U.S.A. **11** (1925), 287–289.
[39] S. Lefschetz, *Continuous transformations of manifolds*, Proc. Nat. Acad. Sci. U.S.A. **11** (1925), 290–292.
[40] S. Lefschetz, *Intersections and transformations of complexes and manifolds*, Trans. Amer. Math. Soc. **28** (1926), 1–49.
[41] S. Lefschetz, *Manifolds with boundary and their transformations*, Trans. Amer. Math. Soc. **29** (1927), 429–462.
[42] C. McCord, *Computing Nielsen numbers*, Nielsen Theory and Dynamical Systems, C. McCord, ed., Contemp. Math. vol. 152, Amer. Math. Soc., Providence, RI (1993), 249–268.
[43] C. Miranda, *Un'osservazione su una teorema di Brouwer*, Boll. Un. Mat. Ital. (1940), 527.
[44] J. Nielsen, *Ringfladen og Planen*, Mat. Tidsskr. **B** (1924), 1–22; English translation in *Jakob Nielsen Collected Mathematical Papers*, Birkhäuser, Basel (1986), 130–146.
[45] J. Nielsen, *Untersuchungen zur Topologie der geschlossenen zweiseitigen Flächen*, Acta Math. **50** (1927), 189–358; English translation in *Jakob Nielsen Collected Mathematical Papers*, Birkhäuser, Basel (1896), 223–341.
[46] H. Poincaré, *Sur certaines solutions particulières du problème des trois corps*, C. R. Acad. Sci. Paris **97** (1883), 251–252.
[47] H. Poincaré, *Sur les courbes définies par une équation différentielle IV*, J. Math. Pures Appl. **85** (1886), 151–217.
[48] K. Reidemeister, *Automorphismen von Homotopiekettenringen*, Math. Ann. **112** (1936), 586–593.
[49] B. Rushing, *Topological Embeddings*, Academic Press, New York (1973).
[50] H. Schirmer, *A Nielsen number for fixed points and near points of small multifunctions*, Fund. Math. **88** (1975), 145–156.
[51] H. Schirmer, *A relative Nielsen number*, Pacific J. Math. **122** (1986), 459–473.
[52] H. Schirmer, *A survey of relative Nielsen fixed point theory*, Nielsen Theory and Dynamical Systems, C. McCord, ed., Contemp. Math. vol. 152, Amer. Math. Soc., Providence, RI (1993), 291–310.
[53] K. Scholz, *The Nielsen fixed point theory for noncompact spaces*, Rocky Mountain J. Math. **4** (1974), 81–87.

[54] M. Shub and D. Sullivan, *A remark on the Lefschetz fixed point formula for differentiable maps*, Topology **13** (1974), 189–191.

[55] H. Siegberg, *Some historical remarks concerning degree theory*, Amer. Math. Monthly **88** (1981), 125–139.

[56] W. Thurston, *On the geometry and dynamics of surface diffeomorphisms*, Bull. Amer. Math. Soc. **19** (1988), 417–431.

[57] L. Vietoris, *Über den höheren Zusammenhang kompacter Räume und eine Klasse von zusammenhangstreuen Abbildungen*, Math. Ann. **97** (1927), 454–472.

[58] F. Wecken, *Fixpunktklassen I*, Math. Ann. **117** (1941), 659–671.

[59] F. Wecken, *Fixpunktklassen II*, Math. Ann. **118** (1942), 216–234.

[60] F. Wecken, *Fixpunktklassen III*, Math. Ann. **118** (1942), 544–577.

[61] D. Wilczyński, *Fixed point free equivariant homotopy classes*, Fund. Math. **123** (1984), 47–60.

This Page Intentionally Left Blank

CHAPTER 11

Geometric Aspects in the Development of Knot Theory

Moritz Epple

AG Geschichte der exakten Wissenschaften, Fachbereich 17 – Mathematik, University of Mainz, Germany
E-mail: Epple@mat.mathematik.uni-mainz.de

1. Introduction

§ 1. Among the most widely noticed achievements of knot theory are certainly the famous knot tables produced by the Scottish tabulating tradition in the late 19th century, the polynomial invariant invented by James W. Alexander in the 1920's, and the series of new polynomial invariants that came into existence after Vaughan F. Jones discovered a new knot polynomial in 1984. It might seem that these results easily fit into a story centered around plane knot diagrams, symbolical codings of such diagrams and the operations one can perform with them, and combinatorial techniques to draw conclusions from the information that is thereby encoded.[1] In this contribution, I will first outline such a narrative and then show that it fails to account for important causal and intentional links in the fabric of events in which these achievements were produced. Indeed, a striking feature of knot theory is that, even if a significant number of its results may be stated and proved in a direct, combinatorial fashion, the research that produced those results was often motivated by and directed toward geometric considerations of varying complexity. In many cases, these geometric ideas alone provided the links to other topics of serious mathematical interest and thus could induce mathematicians to devote their time to knots. Moreover, only by taking into account the surrounding geometric aspects can historians reach a position from which they may judge the relations between the steps in the formation of knot theory and the broader mathematical and scientific culture in which these steps were taken. These relations form part of the causal weave that needs analysis in order to attain a historical understanding of Tait's, Alexander's, or Jones's results.

§ 2. In what follows, I will pursue this subject in five steps (corresponding to Sections 2–6). In Section 2, some of the relevant combinatorial aspects of the history of knot theory will be sketched. This account is mainly intended to anchor the events relating to combinatorics in the historical chronicle, and to highlight the kinds of questions that remain unanswered if the history of knot theory is presented in a perspective that concentrates

[1] A sketch along these lines has been published by Przytycki [131].

HISTORY OF TOPOLOGY
Edited by I.M. James

exclusively on combinatorial issues. In Section 3, the main geometric ideas in the background of the first mathematical treatments of knots, up to and including the tabulations of Peter Guthrie Tait and his followers, will be discussed. It will be seen that, in an important sense, the knot tables of the 19th century did *not* represent an isolated and curious effort in the hardly existent science of topology. Section 4 will then be devoted to the making of what may be called *modern* knot theory in a historically specific sense, based on the new tools of Poincaré's *Analysis situs*, the fundamental group, torsion invariants, and Alexander's polynomial. Section 5 takes up the difficult task of choosing and describing some developments concerned with the further investigation of knots in the period up to Jones's breakthrough. Besides the background to his new invariant, I will mainly focus on researches formulating innovative ideas on knots as genuinely three-dimensional objects, rather than as objects given by diagrams. I will return to my general theme in the concluding Section 6, in a brief attempt to assess the historical role of geometric aspects in the mathematical treatments of knots.

§ 3. Due to limitations of competence and space, the selection of topics discussed cannot be exhaustive and may perhaps not even be representative with respect to the main theme of this article. This holds both with respect to the description of mathematical ideas and – even more so – with respect to the causal and intentional saturation of the historical narrative.[2] Among the many mathematical issues I have *not* dealt with are results about special classes of knots, investigations relating to the finer structure of knot groups, knots in higher dimensions, and the relations between knots and dynamical systems.[3]

It must also be emphasized that the following remarks are written from the perspective of a historian, and not from that of a mathematician engaged in active research on knots. This raises a particular difficulty when it comes to recent developments. Since there is little or no distance to view these events from, one is hard pressed to find *historical* criteria that would help to order the overwhelming amount of material that could be subjected to historical investigation. Since, on the other hand, it is not the historian's task to side with one or several of the engaged parties of active researchers in the assessment of this material, he is left with a huge and (from his perspective) largely unordered corpus of information. Under these circumstances, the best I can hope for is to propose some points of view that may prove useful for a better structuring and understanding of this corpus in subsequent historical work.[4]

2. A tale of diagram combinatorics

§ 4. It has been suggested that one of the earliest tools of combinatorial knot theory was forged by Carl Friedrich Gauss. Some posthumously published fragments of his *Nachlaß*,

[2] Taking descriptions of a complex of intellectual events in which certain mathematical ideas were produced as elements of the basic chronicle of a historical narrative, a historian has to "saturate" this chronicle in one of various possible ways. The idea that guides me in this enterprise is to produce an account of the *weave of mathematical action* in which these intellectual events actually happened. A historical narrative might thus be called "saturated" with respect to its basic chronicle if the causal and intentional context of the basic events in this weave is adequately captured. See also [42, Introduction].

[3] Interesting survey articles that offer information on these and still other questions are [60, 156, 67].

[4] A fuller treatment of the topics discussed here will be found in some of the papers referred to below and in my book *Die Entstehung der Knotentheorie*, forthcoming.

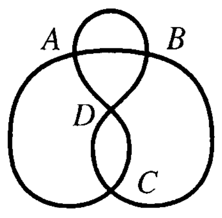


Fig. 1. A "Tractfigur" with crossing sequence $ABCDBADC$.

originally written in the years 1825 and 1844, document that the Göttingen mathematician tried to classify closed plane curves with a finite number of transverse self-intersections, sometimes called "Tractfiguren" (tract figures) by Gauss.[5] To do so, he invented a symbolical coding of such figures. He assigned a number or letter to each crossing and then wrote down the sequence of crossing symbols that resulted from following the curve in a given direction from a given point (see Figure 1).

To some extent, this symbol-sequence captured the characteristic features of the tract figure in the sense of *Geometria situs*, as Gauss preferred to call the as yet unexplored science of topology.[6] Gauss noticed that the sequences arising in this way had to satisfy certain conditions: Each of the symbols representing one of the n crossings had to appear exactly twice, once in an even and once in an odd place of the sequence. As Gauss noticed in 1844, however, these conditions were sufficient only for $n \leqslant 4$. He thus set out to write down a table of the admissible sequences for five crossings, but he did not find a method to solve the general problem of determining exactly which symbol sequences satisfying the above conditions actually represented crossing sequences of "Tractfiguren".[7]

What reasons did Gauss have for looking at this matter? Unfortunately, the fragments themselves do not give a clear answer. From the perspective of later knot theory, Gauss's attempt might look like a first step toward knot tables, but we will see that he had other reasons for studying "Tractfiguren".

Apparently unrelated to these considerations is another posthumous fragment that has often been cited as evidence for Gauss's interest in knots and links. This text, written in 1833, gives a double integral for counting

> the intertwinings of two closed or infinite curves. Let the coordinates of an undetermined point of the first curve be x, y, z; of the second x', y', z', and let
>
> $$\iint \frac{(x'-x)(\mathrm{d}y\mathrm{d}z'-\mathrm{d}z\mathrm{d}y')+(y'-y)(\mathrm{d}z\mathrm{d}x'-\mathrm{d}x\mathrm{d}z')+(z'-z)(\mathrm{d}x\mathrm{d}y'-\mathrm{d}y\mathrm{d}x')}{[(x'-x)^2+(y'-y)^2+(z'-z)^2]^{3/2}} = V$$
>
> then this integral taken along both curves is $= 4m\pi$ and m the number of intertwinings.[8]

Here, the situation is different than with the fragments on "Tractfiguren". A number that modern knot theorists might be inclined to calculate from a link diagram by adding "signs"

[5] [57, vol. VIII, pp. 271–286].

[6] Today, it is known that a reduced projection of a prime knot is indeed determined by its crossing sequence; see [29].

[7] This problem received new interest after Gauss's fragments were published in 1900 and described in [33]. An algorithm solving the problem was first published by Max Dehn in [36].

[8] [57, vol. V, p. 605].

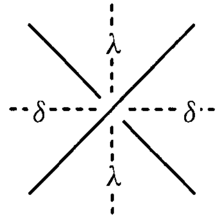


Fig. 2. Marking corners of diagrams.

of diagram crossings, was described by Gauss using *analytical* information. As it stands, also this fragment poses a historical riddle: Why, when, and how was Gauss led to consider linked space curves and this integral?

§ 5. While there is no direct evidence that Gauss actually studied the knot problem, his student and *protegé* Johann Benedikt Listing did.[9] In his 1847 essay, *Vorstudien zur Topologie* (in which he coined the term topology), Listing proposed to study, among other things, "Linearcomplexionen im Raume", roughly corresponding to 1-dimensional cell complexes embedded in ordinary space. The simplest case were knots (in the sense of smooth closed space curves without double points). Listing did not formulate the classification problem explicitly, but the general thrust of his essay suggests that he was interested in topologically invariant characteristics of "Complexionen" like knots. Guided by the Leibnizian idea of a symbolical calculus expressing "situation", as it was understood at the time,[10] Listing associated a "Complexionssymbol" with each knot diagram which, in slightly modernised notation, is a polynomial with integer coefficients in two variables. It was based on a rule for marking the corners of a diagram associated with Figure 2. Connecting two opposite regions by an axis running between the two arcs of the link, these arcs turn around the axis either like a right-handed or a left-handed screw. Accordingly, the regions were marked λ or δ, respectively [94, p. 52]. Listing's symbol was then defined to be the polynomial

$$\sum c_{ij} \lambda^i \delta^j,$$

where each term $c_{ij}\lambda^i\delta^j$ represented all diagram regions with precisely i marks λ and j marks δ; the coefficients c_{ij} were just the numbers of regions of type $\lambda^i\delta^j$, including the outer region. This polynomial was not a knot invariant, however, since diagrams of equivalent knots could have different polynomials. What Listing hoped was that the resulting identities could be made the basis of an *algebraic calculus* with diagram polynomials (in modern terms, one might interpret Listing's idea by considering the quotient of $\mathbb{Z}[\lambda, \delta]$ by the ideal generated by all diagram equivalences). The obvious problem was that the basic identities were unknown as long as the knot problem was unsolved, and Listing was unable to draw any interesting consequences from his definitions.

[9] On Listing, see [21]. A letter of Betti's reporting on his conversations with Riemann gives indirect evidence that during the last years of his life, i.e. after Listing's *Vorstudien* had appeared, Gauss studied knots, though without much success; see [170].

[10] The recurrent appeal to Leibniz' authority on *Analysis situs* is itself a historically interesting phenomenon, see [91, Introduction]. The particular conception of *Analysis situs* that Listing had in mind was in fact due to Euler [43].

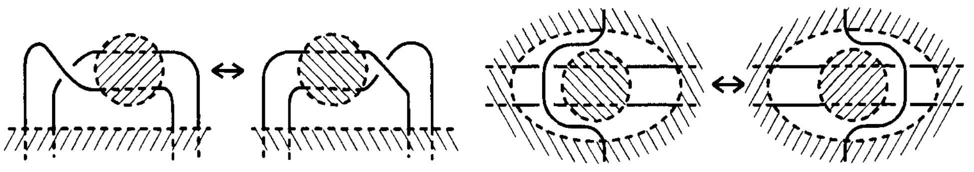


Fig. 3. "Twists" and "two-passes".

§ 6. The next visible scientific enterprise relating to knots was the construction of tables of alternating knots of up to eleven and of non-alternating knots of up to ten crossings by the Edinburgh physicist Peter Guthrie Tait, the Lancashire clergyman and mathematician Thomas P. Kirkman and the American civil engineer Charles N. Little in the last two decades of the 19th century. These tables, all but one published in the journals of the Royal Society of Edinburgh, were the outcome of hard combinatorial work.[11] Tait, who initiated the whole enterprise, outlined the strategy to be followed. It consisted of two separate tasks: first, all possible projections of prime knots (i.e. diagrams where over- and undercrossings were not distinguished) had to be enumerated; second, all possible choices of over- and undercrossings in these projections had to be checked, eliminating diagrams of equivalent knots.

The enumeration of knot projections was the easier part of this strategy. For the lowest crossing numbers, and independently of Gauss's still unpublished ideas, Tait first tried a method based on a refined version of crossing sequences. Later, he settled for a different technique, involving what today is called the "graph" of a knot. For higher crossing numbers, Kirkman took over this project, using another method for enumerating certain four-valent graphs from which knot projections could be derived. The harder task involved searching for duplications among the knot diagrams resulting from the enumerations of knot projections. Tait completed this task for alternating diagrams of up to ten crossings, and Little went on to deal with those having eleven crossings as well as the non-alternating diagrams. Two ideas about how diagrams of equivalent knots were related, implicit in Tait's work but only made explicit by Little, helped them to construct their tables, though both were acutely aware that their results were, to some extent, only tentative. For *alternating* diagrams without "nugatory" crossings,[12] Tait's implicit assumption and Little's explicit claim was that two such diagrams represented the same knot if and only if they could be related by a sequence of "twists" as in Figure 3 (left).[13] Only recently, and based on Jones's new invariant, has this conjecture been proved by Menasco and Thistlethwaite [108]. For *non-alternating* diagrams, Little argued that a sequence of twists and of additional operations, today called "two-passes" and illustrated in Figure 3 (right), would suffice to generate all diagrams of equivalent knots. Unfortunately, this claim was recognised to be wrong when K.A. Perko discovered a duplication in Little's tables in 1974 that the latter had missed because the two are not related by twists and two-passes.

[11] The main publications are [153–155, 88, 89, 95–97]. See [42] for a more detailed description of this work.

[12] A diagram crossing was called "nugatory" by Tait if a simple closed curve existed in the diagram plane intersecting the diagram only at this crossing. Diagrams without such crossings are today called "reduced".

[13] The name "flype" that modern authors tend to attach to this operation was used by Tait for a different operation.

Two obvious historical questions arise. First, why did these men spend so much of their time on knot tabulations? And, second, was their work causally linked to the Göttingen environment of the 1840's, where Gauss and Listing had dealt with similar issues?

§ 7. Not long after Poincaré had created the conceptual tools of modern topology – homological invariants and the fundamental group – presentations of the fundamental group of a knot complement, associated with a knot diagram, became known. In 1910, Max Dehn published a method for finding such a presentation and pointed out that a study of knots using group presentations would require solving some of the basic problems of combinatorial group theory [34]. In this connection, Dehn actually gave the first general and explicit formulations of the word and conjugacy problems in finitely presented groups.[14] Dehn also sketched a technique for treating the word problem by constructing what he called the "Gruppenbild", namely, the Cayley graph of a finitely presented group $G := \langle a_1, \ldots, a_n \mid r_1, \ldots, r_n \rangle$ consisting of the group elements as its vertices and oriented edges connecting group elements of the form g and $a_i g$. The cycles in this graph obviously correspond to all trivial words of the presentation, so that constructing the "Gruppenbild" and solving the word problem of a group presentation are equivalent tasks. Dehn managed to construct the graph of the group of a trefoil knot. The graph showed that this group was non-commutative, and hence a trefoil knot could not be deformed without self-intersections into an unknotted circle.

In the 1920's, Kurt Reidemeister and Emil Artin pointed out that another method for associating a group presentation with a knot diagram had been developed already around 1905 by the Vienna mathematician Wilhelm Wirtinger; and in fact the method had been described in a somewhat disguised fashion by Tietze in a paper of 1908.[15] Again, a question arises: what drew Wirtinger and Dehn to study knots and their groups in the first place? Was it just the wish to apply Poincaré's new tools to a "natural" particular case?

§ 8. The 1920's brought the first effectively calculable invariants of knots, and thus also a means for verifying the knot tables of the 19th century. More or less independently, the Princeton topologist James W. Alexander (together with his student G.B. Briggs), and Kurt Reidemeister, first at Vienna and then at Königsberg, showed how to associate certain matrices with knot diagrams in such a way that the elementary divisors of these matrices were knot invariants.[16] The model for this technique was clearly Poincaré's calculation of the torsion numbers of cell complexes, but both Reidemeister and Alexander presented their results in a completely independent way; Reidemeister even spoke of a new "elementary foundation" for knot theory. Both defined knots as equivalence classes of finite polygons in three-dimensional Euclidean space. Two such polygons were considered equivalent if and only if they could be deformed into each other by a sequence of applications of the following operation and its inverse: two incident edges AB, BC of a polygon may be replaced by an edge AC, provided the triangle ABC contains no further point of the polygon. Reidemeister and Alexander translated this into an equivalence relation between knot (or link) *diagrams*. Instead of just one operation, four had now to be considered: an analogue

[14] For a discussion of the relations between knot theory, topology in general, and early combinatorial group theory, see [28, Chapter I.4].

[15] See [9, Section 6; 160, p. 103f].

[16] See [134, 135, 7].

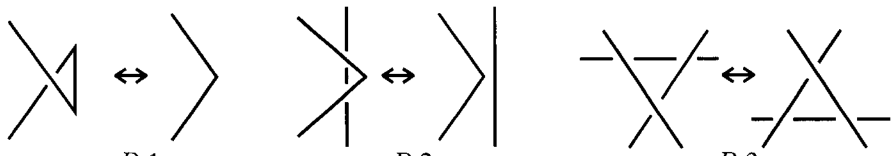


Fig. 4. Reidemeister's diagram moves.

of the above in the plane (where it is allowed that an arc of the polygon without crossings lies "below" or "above" the triangle involved), and three "diagram moves" involving modifications of diagram crossings which since have become known under the name of "Reidemeister moves" (see Figure 4).

Again, diagrams that could be deformed into each other by a finite sequence of such moves represented the same knot, and thus any mathematical object associated with a knot or link diagram invariant under Reidemeister's moves was a knot (or link) invariant. Reidemeister and Alexander showed this to be the case for the nontrivial elementary divisors of the matrices they were considering. Since this helps to understand Alexander's subsequent invention of a knot polynomial, let me present Alexander's version of a matrix associated with a knot diagram in condensed form. Thus, let the ν crossings of a knot diagram be denoted by $c_1, c_2, \ldots, c_\nu$, and let its $\nu + 1$ finite regions be denoted by $r_0, r_1, \ldots, r_\nu$.[17] The region r_0 should have a common border with the infinite region. After selecting an orientation of the diagram, the corners of diagram regions are marked according to the convention that the corners of the two regions to the *left* of each *undercrossing* arc receive a dot. Moreover, an integer $n \geqslant 2$ is chosen. Then an $(n\nu \times n\nu)$-matrix M is defined, consisting of $\nu \times \nu$ blocks a_{ij}, each of size $n \times n$. In order to abbreviate the definition, let I denote the $(n \times n)$-unit matrix, and let x be the $(n \times n)$-block given by

$$x := \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & & \vdots \\ 0 & 0 & 0 & \ddots & \vdots \\ \vdots & & & \ddots & 1 \\ 1 & 0 & \ldots & \ldots & 0 \end{pmatrix}.$$

Then M is defined by the following rules: (1) to each crossing c_i corresponds a row of blocks a_{ij}, and to each region r_j, $j = 1, \ldots, \nu$, corresponds a column of blocks a_{ij} in M; (2) if r_j, r_k, r_l, r_m are the regions incident with a crossing c_i, in cyclical order as one goes round c_i in counterclockwise sense, and such that the dotted corners belong to r_j and r_k, then $a_{ij} = a_{ik} = x$ and $a_{il} = a_{im} = I$.

It was now a matter of straightforward calculation to show that the elementary divisors of M different from zero and one – called the "torsion numbers" of the knot – remained invariant under Reidemeister's diagram changes. Alexander and Briggs calculated the elementary divisors of all of the 168 matrices associated with the 84 knots of nine or less

[17] These notations are taken from [5]. In the earlier [7], different notational conventions were adopted. In both papers, the matrix defined was viewed as a matrix of coefficients of a system of linear equations in the variables r_j.

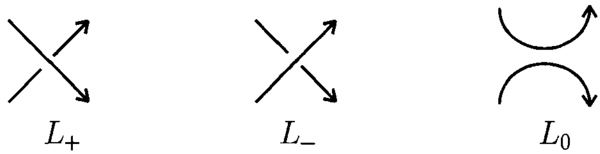


Fig. 5. Local modifications of links.

crossings in Tait's tables, corresponding to $n = 2$ and $n = 3$. Except for three pairs with identical torsion numbers, all of these knots were found to have distinct invariants.

Of course, the above description makes Alexander's construction of invariant numbers appear historically opaque. How could he possibly have found all this machinery? Why were these invariants called torsion numbers?

§ 9. In Alexander's and Briggs's first paper, the block structure of the above matrix was not introduced explicitly. Once it was recognized (and no doubt it was recognized during the extensive calculations needed for checking Tait's tables), it was but a small step to see that one could view the $(\nu \times \nu)$-matrix of *blocks* as a matrix with polynomials in the formal variable x as its entries (I being identified with 1). This step was taken in [5], where virtually the same arguments as before showed that the nontrivial elementary divisors of this new matrix, and therefore in particular its determinant $\Delta(x)$, were invariants under the Reidemeister moves up to a factor $\pm x^k$ $(k \in \mathbb{Z})$.[18] Normalizing $\Delta(x)$ by the requirement that the term of lowest degree became a positive constant, Alexander obtained the polynomial invariant of oriented knots that today carries his name. Again, the knots in Tait's tables were used to test the force of the new invariant. Alexander found that the polynomial, though of course much easier to calculate, was only slightly less effective in distinguishing knots than the torsion numbers. It turned out that both, however, could not distinguish knots from their reverse knots (obtained by reversing the orientation) or mirror images (obtained by switching all crossings).

Toward the end of his paper, Alexander included a side remark which probably resulted from his experiences with calculations of $\Delta(x)$. After noticing that his definition could equally well be applied to link diagrams (in this case, it gave rise both to a one-variable polynomial of oriented links, defined by the same rules as above, and a polynomial in as many variables as the link had components), Alexander established a relation of the one-variable polynomials of "three closely related links" [5, p. 301]. Using the modern subscripts L_+, L_-, and L_0 for oriented link diagrams that only differ at one crossing in the way indicated in Figure 5, Alexander's relation can be written as

$$\Delta_{L_-}(x) - \Delta_{L_+}(x) = (1 - x)\Delta_{L_0}(x). \qquad (\star)$$

For the time being, however, nothing was made of this relation.

§ 10. In 1961, Wolfgang Haken published a long and difficult paper in which an algorithm was described that allowed one to decide whether or not a given knot was equivalent

[18] Actually, Alexander used both this matrix and an equivalent one, in which certain signs were added to take care of orientations.

to the unknot [64]. Although this algorithm was extremely impractical, its existence made it probable that the classification of knots was possible by algorithmic means. Accordingly, and enhanced by the availability of powerful computers, the interest in computerised knot tabulations increased significantly during the 1960's. At a conference in 1967, John Conway surprised the tabulators by presenting an algorithm for enumerating knots and links which was much more effective than those used by the 19th-century tabulators and which enabled him to enumerate knots of up to eleven crossings and links of up to ten crossings by hand [31]. The main tool was a calculus of "tangles", parts of link diagrams with four open ends, that could be used to survey their possible combinations and closings. In order to distinguish the various links enumerated in this way, Conway had to calculate invariants, too. This led him to reconsider the Alexander polynomial, and he redefined it by a change of variables and a new normalization. Apparently without having read Alexander's earlier remark, Conway pointed out that his version of the polynomial satisfied an equivalent of ($\star$) and similar relations which he later came to call "skein relations".[19] In view of their usefulness in calculations, he counted these relations among "the most important and valuable properties" of his version of the Alexander polynomial, but he did not try to *define* the polynomial using a variant of ($\star$).

This was done in 1981 by L.H. Kauffman. He observed that up to a suitable normalization, Alexander's one-variable polynomial $\Delta_L(t)$ of oriented links L was uniquely determined as a symmetric element of $\mathbb{Z}[t^{1/2}, t^{-1/2}]$ (in the case of proper knots even of $\mathbb{Z}[t, t^{-1}]$) by the following two conditions (the symbol $\bigcirc$ represents the unknot):[20]

$$\Delta_{\bigcirc}(t) = 1, \tag{1}$$

$$\Delta_{L_+}(t) - \Delta_{L_-}(t) + \left(t^{1/2} - t^{-1/2}\right)\Delta_{L_0}(t) = 0. \tag{2}$$

Since any link diagram could be changed into a trivial one by switching its crossings, it was not difficult to see that these two rules would suffice to *calculate* the polynomial inductively, provided it was well-defined. This was shown to be the case using yet another description of $\Delta(x)$ as the determinant of a matrix associated with an oriented link diagram. Still, no further analysis of this seemingly peculiar property of Alexander's polynomial was undertaken, and Kauffman expressed his astonishment about the approach: "It seems nothing short of miraculous that such a scheme should produce good invariants" [83, p. 102].

§ 11. The view of the matter changed dramatically when Jones discovered his new polynomial invariant of oriented links in 1984 [77]. In discussions, Jones and Joan Birman found that also this invariant satisfied a skein relation similar to the one found by Alexander. In fact, if $V_L(t)$ denotes the Jones polynomial associated with a link L, then $V_L(t)$ was uniquely determined by the conditions

(i) $V_{\bigcirc}(t) = 1$,
(ii) $tV_{L_+}(t) - t^{-1}V_{L_-}(t) + \left(t^{1/2} - t^{-1/2}\right)V_{L_0}(t) = 0$.

---

[19] According to Lickorish [92]. In [31], this terminology was not used.

[20] Conventions on signs and variables in the literature on knot polynomials are far from consistent. I follow here [54, 67]. In [83], the polynomial $\Omega(t) := \Delta(t^2)$ was considered.

Fig. 6. Local modifications of unoriented links.

This striking similarity induced several mathematicians to investigate the conditions under which a *general* skein relation would define a link invariant. Almost simultaneously, at least eight mathematicians found an equivalent answer to this question. In the joint article (Freyd et al. [54]),[21] this answer was stated as follows: there is a unique invariant $P_L(x, y, z)$ of oriented links with values in the homogeneous Laurent polynomials of degree 0, satisfying

$$\text{(I)} \quad P_{\bigcirc}(x, y, z) = 1,$$
$$\text{(II)} \quad xP_{L_+}(x, y, z) + yP_{L_-}(x, y, z) + zP_{L_0}(x, y, z) = 0.$$

Moreover, this invariant is universal in the sense that *any* link invariant Q with values in a commutative ring A that equals one for the unknot and satisfies a linear skein relation $aQ_{L_+} + bQ_{L_-} + cQ_{L_0} = 0$, for arbitrary invertible coefficients $a, b, c \in A$, can be obtained from P by the canonical ring homomorphism from $\mathbb{Z}[x, x^{-1}, y, y^{-1}, z, z^{-1}]$ to A which sends x, y, z to a, b, c, respectively. In particular, both Alexander's and Jones's polynomials can be obtained from P (which is often expressed as the *inhomogeneous two-variable* polynomial $P_L(l, l^{-1}, m)$) by suitable substitutions of the variables.

It turned out that Jones's polynomial and its generalization were much stronger invariants than the Alexander polynomial. In many cases, these polynomials distinguished knots from their mirror images, and up to the time of writing, no nontrivial knot seems to be known with the Jones polynomial of the unknot.

Of course, once the surprising force of skein relations was recognized, variations of this combinatorial theme seemed promising and a whole series of related polynomials were found.[22] Kauffman's investigations were again particularly successful in this respect. By considering the *four* possible local modifications of *unoriented* links (see Figure 6) he found not only a new and extremely simple definition of Jones's polynomial (Kauffman [84]) but also a two-variable polynomial invariant of oriented links that was seen to be independent of P (Kauffman [85]).

§ 12. It is a historiographical commonplace that quite different historical narratives based on the same documentary material are possible. The above outline of some important combinatorial aspects in the development of knot theory is one of the stories that can be told about mathematical treatments of knots and links. Homogeneous as it may seem, though, it is clear that crucial historical questions remain unanswered and important parts of the documentary evidence have been passed over in silence. What were the actual motivating backgrounds for those contributing to this development? For whom, and in what contexts, did they work on knots? How, precisely, were physicists like Tait and mathematicians such as Dehn, Reidemeister, or Alexander led to form their ideas? How could an operator algebraist like Jones hit on a topological invariant of links? It is hardly imaginable that an

[21] See also [132].

[22] A concise survey is given in [93].

interest in the combinatorics of knot diagrams alone provided enough motivation and a sufficiently elaborated intellectual framework for constructing knot tables, for studying the knot group, or for inventing polynomial invariants. And indeed, in most cases, it turns out on closer inspection that quite varied and much richer impulses were at work in the historical development. Thus, a story about the mathematical study of knots can be told which is quite different from the above.

3. Mathematical treatments of knots before 1900

§ 13. Since prehistoric times, knots and interlacing patterns have been used in human cultures for practical, ornamental, and symbolical purposes. Against this background, the beginning of a *mathematical* interest for knots in the late 18th century marks a striking discontinuity. It has to be understood within the general context of the progressing mathematization of many domains of human knowledge and practice that characterizes this epoch. More precisely, knots and interlacing patterns found the attention of those few mathematicians that were interested in a vaguely conceived new exact science, tentatively called the "science of situation", *Analysis situs* or *Geometria situs*.[23] In fact, the first but vigorous attempt to bring knots within the reach of mathematical treatment bears all the marks of a typical Enlightenment attempt to mathematize a domain of human practice. It was made in 1771 by the Paris intellectual, A.T. Vandermonde, later a decided supporter of the French Revolution. In a short paper entitled "Remarques sur les problèmes de situation", he wrote:

> Whatever the convolutions of one or several threads in space may be, one can always obtain an expression for them by the calculus of magnitudes; but this expression would not be of any use in the arts. The worker who makes a *braid*, a *net*, or *knots*, does not conceive of them by relations of magnitude, but by those of situation: what he sees is the order in which the threads are interlaced. It would thus be useful to have a system of calculation that conforms better to the course of the worker's mind, a notation which would only represent the idea which he forms of his product, and which could suffice to reproduce a similar one for all times.[24]

Besides showing how some symmetrical weaving patterns (ones actually used in textile manufacture) could be described by means of a symbolical notation, Vandermonde did little to advance a veritable "system of calculation" relating to knotted or linked space curves. Nevertheless, it is significant that this kind of problem was incorporated into *Geometria situs* long before, say, the classification of surfaces became an important issue.

§ 14. Also in Gauss's case, it seems to have been the *uses* of the new science of *Geometria situs* that captivated his interest in the topic. For him, however, these uses were concerned not with the practical arts but rather the exact sciences, including traditional pure mathematics as well as sciences like astronomy, geodesy, and the theory of electromagnetism.[25] Gauss encountered linked space curves for the first time in his scientific career in an astronomical context. This happened in 1804, some twenty years before the first of the fragments described in the last section was written. After Gauss's successful

[23] See [129, 53].

[24] [163, p. 566]; my (rather literal) translation.

[25] The following paragraphs are a condensed version of [41]. For details and full references, the reader is referred to this article.

calculation of the orbit of the first observed asteroid, Ceres, had spread his fame over Europe in 1801, he continued to follow the discoveries of several other "small planets" made soon thereafter with increasing rapidity. In this connection, he published a small treatise entitled *Über die Grenzen der geocentrischen Örter der Planeten*, which took up a rather practical question, namely the determination of the celestial region in which a given new "planet" might possibly appear.[26] Taking the liberty of presenting Gauss's arguments in modern mathematical language, the problem of this paper can be described as follows. If the orbit of the earth's motion around the sun is given by $X \subset \mathbb{R}^3$, and if $X' \subset \mathbb{R}^3$ is the orbit of another celestial body (the sun being at the center of a suitable system of Cartesian coordinates), Gauss wanted to determine the region on the sphere given by

$$\left\{ \frac{\vec{x} - \vec{x}'}{\|\vec{x} - \vec{x}'\|} \in S^2 \,\middle|\, \vec{x} \in X,\ \vec{x}' \in X' \right\}.$$

This region he called the *zodiacus* of the celestial body in question. Its determination helped to limit the effort needed both in the observation of the celestial body and in the production of an atlas of the smallest part of the celestial sphere on which the orbit of the body could be drawn. In order to solve this problem, Gauss derived a differential equation for the *boundary curve or curves* of the *zodiacus*, implicitly assuming the orbits to be smooth curves. If $\vec{x} = (x, y, z) \in X$ and $\vec{x}' = (x', y', z') \in X'$ denote the coordinates of orbit points, a necessary condition that a pair of points $(\vec{x}, \vec{x}')$ corresponds to a boundary point of the *zodiacus* is that the triple consisting of the two tangent vectors to the orbits at $\vec{x}$ and $\vec{x}'$ and the distance vector $\vec{r} := \vec{x}' - \vec{x}$ is linearly dependent. Gauss expressed this condition by saying that the two tangents at $\vec{x}$ and $\vec{x}'$ had to be coplanar. Translating the condition into a formula led to the differential equation

$$(x' - x)(\mathrm{d}y'\mathrm{d}z - \mathrm{d}y\mathrm{d}z') + (y' - y)(\mathrm{d}z'\mathrm{d}x - \mathrm{d}z\mathrm{d}x') \\ + (z' - z)(\mathrm{d}x'\mathrm{d}y - \mathrm{d}x\mathrm{d}y') = 0.$$

Obviously, the differential form on the left-hand side is, up to a change of sign, nothing but the numerator of the integrand in the linking integral! At this point, Gauss inserted a typical remark: He had undertaken a mathematical study of this equation in its own right, but for the sake of brevity he did not wish to go into that now. However, Gauss pointed out that one case was of particular importance, namely that in which the two orbits were *linked*. (Even this was not just a mathematical fancy: While none of the orbits of the known "planets" was linked with that of the earth, Gauss reminded his readers that "comets of the sort exist in abundance".[27]) In this case, the *zodiacus* was, "for reasons of the geometry of situation", the whole celestial sphere.

As I have described elsewhere, it is probable that already in his study of the equation determining the boundary of the *zodiacus* Gauss began to understand the connection between the geometry of linked space curves and the integer calculated by his double integral – an integer which in modern mathematical language may also be described as the mapping degree of the mapping defining the *zodiacus*. Thus, geometric considerations that came up

---

26 The article is reprinted in [57, vol. VI, pp. 106–118].

27 [57, vol. VI, p. 111f] – in 1847, Listing counted 25 pairs of *asteroids*, whose orbits were known to be linked by then [94, p. 64f].

in a scientific context highly appreciated at its time induced Gauss to think about this kind of topological phenomenon.

A similar connection involving an exact science and *Geometria situs* probably first aroused Gauss's interest in tract figures. During the 1820's, his geodetic work related to the triangulation of the Kingdom of Hanover induced him to develop once again some new mathematics. While Gauss was directing this lucrative enterprise, he also worked on the *Disquisitiones generales circa superficies curvas*, published in 1827. In this concise treatise, he developed the basic ideas on curvature and geodesics on surfaces that formed the starting point of modern, intrinsic differential geometry. The crucial tool for studying curvature, however, depended on the consideration of surfaces *embedded* in ordinary space. This tool was a mapping that today carries Gauss's name: it associated to each point on a curved surface the direction of the surface normal at that point; this direction was then represented as a point on an auxiliary sphere. Using this mapping, Gauss introduced the notion of the total curvature of a portion of the surface bounded by a simple closed curve. By definition, this curvature was given by the area enclosed by the image of the boundary curve on the auxiliary sphere. Here, however, a problem arose: the image curve could have singularities – i.e. it could be a tract figure on the sphere (or even worse). Thus one had to clarify what "the area enclosed" by such a figure actually meant. In this way, Gauss was led to look at the topology of closed plane curves in more detail, and it was amidst his work on the *Disquisitiones* that he sketched his first ideas about tract figures. In the published treatise, he only alluded to this work (and the solution of the problem of defining the area enclosed by a tract figure by means of "characteristic" numbers given to the various regions of the figure). But in a letter to his friend Schumacher, he complained:

> Some time ago I started to take up again a part of my general investigations on curved surfaces, which shall become the foundation of my projected work on higher geodesy. [...] Unfortunately, I find that I will have to go *very far* afield [...]. One has to follow the tree down to all its root threads, and some of this costs me week-long intense thought. Much of it even belongs to *geometria situs*, an almost unexploited field.[28]

At about this time, Gauss also spent some thought on another peculiar object of *Geometria situs* – a four-strand braid. While the page in one of Gauss's notebooks documenting this astonishing attempt reveals that he knew how to determine the linking number of two curves by counting signs of diagram crossings, it remains unclear how this fragment relates to Gauss's other mathematical activities.[29]

The third exact science which brought Gauss back to the study of linked space curves was the theory of electromagnetism, which drew considerable scientific and public attention after Oersted's and Faraday's discovery of electromagnetism and electromagnetic induction. As is well known, Gauss was involved together with his friend and colleague, the physicist Wilhelm Weber, in setting up the first telegraph in Göttingen in April 1833. In connection with this work, Gauss studied intensively the mathematical formulation of the laws of electromagnetical induction. It could not have escaped his notice that the law describing the magnetic force induced by an electric current was governed by precisely the same differential form which he had encountered in his earlier investigation of the *zodiacus*. Conceiving magnetic forces as acting on particles, which behave mathematically like monopoles in some "magnetic fluid", the linking integral could be interpreted as ex-

[28] Gauss to Schumacher, 21 November 1825, in: [57, vol. VIII, p. 400].

[29] The fragment has been published and discussed in [41].

pressing the work needed to carry an "element of magnetic fluid" along a closed path in the magnetic field induced by a current running through another closed curve. Still, when Gauss wrote the passage on the linking integral in his notebook a few months before the telegraph was finished, he made no explicit remarks about electromagnetism.

Those with close contact to Gauss's work, including Wilhelm Weber and Ernst Schering, the later editor of Gauss's writings on electromagnetism, knew that he had thought of topological issues in connection with electromagnetism, and Schering thus decided that the fragment on the linking integral should be published in the fifth volume of Gauss's *Werke*, which appeared in 1867 and contained his unpublished notes on electromagnetic induction. It was there that another physicist learned of Gauss's interest in *Geometria situs*: James Clerk Maxwell. In his masterpiece, the *Treatise on Electricity and Magnetism* of 1873, Maxwell went to considerable lengths to explain the physical content of the linking integral [104, §§409–422].

From several passages in Gauss's letters as well as from certain writings of his scientific friends, we know that Gauss held the still barely existent science of *Geometria situs* in very high esteem and expected great developments to come from future research in this field. The reasons for his expectations certainly had little to do with his inconclusive results on tract figures or similar combinatorial ideas. Rather, they derived from his experience that certain types of objects and problems, like linked space curves and tract figures, that were geometric in nature but independent of "magnitude", continually reappeared in some of the leading sciences of his day, ranging from pure mathematics to electromagnetism.

§ 15. A similarly close relationship between important issues in the exact sciences and new ideas related to knots continued to hold throughout the 19th century. When Tait embarked on his tabulation enterprise, he was motivated by developments in natural philosophy in which topological ideas played a very fundamental role. The crucial mathematical device that brought topology into play came from Germany: potential theory in multiply connected domains. Thus, Tait's enterprise and the earlier topological ideas shaped under Gauss's hegemony at Göttingen were actually connected in the fabric of scientific practice, although in an indirect way. Three elements must be put together in order to understand this connection: the dynamical theories that many British natural philosophers held in the second half of the last century, H. von Helmholtz' researches on vortex motion in perfect fluids, and Riemann's notion of *connectivity*.[30]

Guided by the many "mathematical analogies" between physical phenomena that could be described by means of the Laplace equation (electrostatics, heat flow, etc.), many of the leading British physicists in the second half of the last century believed that all of physical theory could and should ultimately be based on some kind of Lagrangian dynamics that governed the flow of a continuous medium (or several media). However, an important challenge to this conception emerged with the continuous rise of atomistic conceptions in chemistry. From the 1860's onward, atomism was forcefully supported by what was perhaps the most advanced experimental technology of the time, spectrum analysis. This posed an immediate problem: how could the smallest units of matter possibly be explained by the dynamics of a continuous medium? One hint came from experiments with magnetism that seemed to imply that, on the molecular level, some kind of rotary motion

[30] Full references and a detailed account of the events described in the following paragraphs may be found in [42].

took place. Already in the 1850's, one of the important proponents of dynamical theories, William Thomson, the later Lord Kelvin, hoped to solve the riddle of atoms by detecting some kind of stable (presumably rotary) dynamical configuration in the motion of the universal medium (ether).

A crucial piece of knowledge for Thomson's pursuit of this line of thought was provided when Hermann von Helmholtz published a ground-breaking paper in which the dynamics of a perfect (i.e. incompressible and friction-free) fluid was investigated without supposing, as had been done earlier, that such a flow could be described by a globally defined potential function (Helmholtz 1858). In modern mathematical notation, Helmholtz discussed solutions $\vec{v}$ of the Euler equations,

$$\vec{F} - \frac{1}{h}\nabla p = \frac{\partial \vec{v}}{\partial t} + (\vec{v} \cdot \nabla)\vec{v},$$

$$0 = \nabla \cdot \vec{v},$$

for which the field of rotation, rot $\vec{v}$, did not necessarily vanish.[31] Helmholtz observed that the integral curves of rot $\vec{v}$ – called "Wirbellinien" (vortex lines) – possessed a kind of dynamical stability. During the motion of the fluid, the particles constituting a vortex line would continue to do so throughout the motion. In particular, if a vortex line was closed, it would remain closed, however altered in shape. This suggested looking for a particular kind of solution to the Euler equations. Helmholtz supposed the dynamics of a finite number of closed vortex lines (or vortex tubes, i.e. tubular bundles of vortex lines emerging from a small area) to be given. He then asked: could a solution to the Euler equations be found with precisely these vortex lines or tubes as its (discontinuous) rotation field? Since *outside* the vortices the rotation had to vanish, this problem amounted to finding a solution of the Laplace equation $\Delta\varphi = 0$ in the multiply connected complement of the vortices (possibly bounded by some closed surface). Here, a solution was understood to be a *many-valued* function φ defined on the complement of the vortices, the branches of which satisfied the Laplace equation locally. By invoking the analogy between the mathematics of hydrodynamics and of electromagnetism, Helmholtz made it clear that such solutions always existed. The mathematics of the situation corresponded to that of a system of closed currents (playing the role of the vortices) which induced a magnetic field (assuming the role of the flow) in their complement.[32] In fact it was easy to write down integral formulae representing the solutions locally. Helmholtz illustrated his results, which he viewed as a three-dimensional analogy to the "Abelian integrals of the first kind" on Riemann surfaces, by giving explicit formulae for a few concrete cases like circular vortex rings.

After a delay of more than eight years, and through the mediation of some experimental illustrations of Helmholtz' results by the Edinburgh physicist Peter Guthrie Tait, Thomson eventually understood that these findings provided one of the missing links in his earlier speculations on atomism. Were not these sorts of closed vortices the kind of stable dynamical configuration in the ether that made up atoms? Once this idea had taken shape in

[31] As a matter of fact, the vectorial notations for the rotation field and for the equations of fluid motion were indirectly inspired by Helmholtz' paper. When Tait read this paper in the fall of 1858, he was reminded of certain quaternionic formulae he had seen earlier in Hamilton's writings; this induced him to start a crusade for the use of quaternions – and thus, also vectorial notations – in physics.

[32] The analogy is strictly correct only in the stationary case.

early 1867, Thomson set out to pursue it with surprising energy. Two things were clear: first, if his speculation was correct, then the different topological types of knots and links provided a wealth of forms that should account for the variety of chemical elements. Second, if it was legitimate to hope for a precise mathematical treatment of "vortex atoms", as Thomson's ether singularities were now called, it would be necessary to extend Helmholtz' theory of vortex motion to a considerable degree.

Thomson took up the second of these tasks. Among the first goals he set himself was to determine the solution space of the problem Helmholtz had considered. Thus, given a multiply connected domain in three-dimensional space (the complement of a system of vortices), how many linearly independent solutions of the Euler equations for a perfect fluid existed with given boundary conditions (e.g., with the fluid flow tangential to all boundary surfaces)? The surprising answer that Thomson found was that the number of parameters determining a solution was dependent only on the topology of the domain considered. It equalled the "order of continuity" of this domain, as Thomson called it. In modern terms: The dimension of the linear space of harmonic vector fields in a given domain, with fixed boundary conditions, was equal to the domain's first Betti number. Thomson was aware of the fact that his result provided an analogy between integrals of the Euler equations for a perfect fluid and Abelian integrals on Riemann surfaces, an even closer analogy than that Helmholtz had seen earlier. Much later, this insight would be explored in a different direction by Hodge's theory of harmonic integrals.

Due to Thomson's theory, interest among British physicists in topological ideas began to surge. In late 1867 and during the following year, Maxwell also began to think about the topological issues involved in the theory of vortex atoms, although his interest stemmed perhaps more from the relevance of the mathematics of vortex motion for electromagnetism rather than because he believed it could be used to explain the structure of matter. Maxwell produced several manuscripts in which he sketched some of the topological ideas needed for dynamical theory.[33] One basic proposition concerned the first Betti number of a region in ordinary space, a "solid with holes" as he described such a region intuitively. If a solid with holes was bounded by one external surface of genus n_1 and several internal bounding surfaces of genus $n_2, \ldots, n_m$, then the first Betti number of the region was $b = n_1 + n_2 + \cdots + n_m$.

Both Thomson and Maxwell did not yet have a sufficiently clear language to give precise formulations and proofs of their topological results; neither the notion of the genus of a surface nor that of the "order of connectivity" of a space region were completely clear in their work. Maxwell and Thomson tried to explain their ideas mainly in terms of "irreconcilable curves" in a domain, but "reconcilability" meant for them something closer to *homotopical* rather than *homological* equivalence. Thus, a particular difficulty in determining the first Betti number of a space region arose again from knotting: why was the "order of connectivity" of a link complement equal to the number of components of the link, as Maxwell's result implied? A closer analysis shows that it was *physical thinking* rather than mathematical precision that helped Thomson and Maxwell to find the correct results.[34] In any case, an understanding of the topology of knots and links became a requisite part of their physical theories.

---

[33] These manuscripts were published only recently in [106, vol. 2].

[34] Briefly put: In technical arguments on multiply connected domains, cutting surfaces (interpreted as membranes stopping fluid motion) were used rather than "irreconcilable curves". See [42, Section II] for details.

In one of his manuscripts, and to the best of my knowledge for the first time, Maxwell explicitly formulated the classification problem for knots and links. Independently of Gauss's and Listing's earlier attempts, he then developed a method to represent link diagrams symbolically, and went so far as to look for obvious diagram modifications ordered according to the number of diagram crossings involved. Not surprisingly, this led him to uncover the "Reidemeister moves" – without, however, considering the question of whether or not these moves would generate all diagram equivalences.[35] Apparently, Maxwell did not pursue his reflections on knots very far in the years around 1868. However, in a very favourable review of Thomson's theory of vortex atoms, written for the ninth edition of the *Encyclopedia Britannica* in 1875, he pointed out that the classification of knots might actually turn out to be rather complicated: "The number of essentially different implications of vortex rings [that is: knot types] may be very great without supposing the degree of implication of any of them very high" [105, p. 471].

Soon afterward, Tait began to investigate the classification problem of knots along the lines described in Section 2. Throughout his work on the tabulations, Tait was motivated by the possible contributions these tables could make to the theory of vortex atoms, and he dropped his work when he felt the tables were sufficiently extensive to be compared with the requirements of physics – be it with a positive or, as became more and more probable, with a negative result.[36] It should be emphasized, however, that the scientific background of Tait's tabulation enterprise was anything but a scientific curiosity. Given the beliefs and methods of the period, Thomson's theory was considered a serious and even promising speculation. The fact that several of the leading British natural philosophers, including Maxwell, followed Thomson's ideas with interest, in itself offers ample evidence of this. Moreover, even if unsuccessful, the theory of vortex atoms was the first serious attempt to explain atoms on the basis of fundamental laws of motion rather than by postulating additional theoretical entities, like force centres or the like. Finally, the *mathematics* that had to be developed in order to pursue the theory was clearly perceived to be important, even if the physical core of the theory should turn out to be incorrect. One final point deserves attention. As in all earlier contributions to the mathematical study of knots, knots were thought of as physical objects in ordinary space. While Tait and his followers used diagrams to deal with these objects, the physical context implied that the *complement* of a knot or link was at least as interesting as the link itself. Thus, in connection with vortex atoms, it was the geometry of this spatial domain rather than the combinatorics of diagrams that "mattered".

4. The formation of "modern" knot theory

§ 16. By the time mathematicians of the twentieth century turned again to the investigation of knots and links, both the status of topology and the general horizon of mathematical culture had changed deeply – a new epoch of mathematics had dawned that may reasonably be called "mathematical modernity". Two aspects of these changes are particularly

[35] Cf. [42, §20; 104, vol. 2, pp. 433–438].

[36] During the 1880's, Thomson himself gradually abandoned the theory of vortex atoms. He repeatedly failed to prove that vortices possessed kinetic stability, and he began to feel that the difficulties to include other physical phenomena like light and gravitation into the picture were unsurmountable. See [152] for a concise description of Thomson's changing views.

relevant for our story. On the one hand, Poincaré's writings on *Analysis Situs* offered new ways to conceive topological objects and new mathematical tools to deal with them, however vague some of his proposals still were on the technical level.[37] On the other hand, the emergence of the modern, axiomatic style in mathematics, the power of which had been impressively demonstrated in Hilbert's *Grundlagen der Geometrie* of 1899, underlined the intellectual autonomy of mathematics and its increasing conceptual separation from the exact sciences. In such an environment, a continuation of the study of knots along the lines followed by Tait, Kirkman and Little seemed hardly promising.[38] Indeed, two quite different lines of thought brought knots to the fore of modern mathematics: the study of singularities of complex algebraic curves and surfaces, and Poincaré's attempt to give a topological characterization of ordinary three-dimensional space, known as the "Poincaré conjecture".[39]

§ 17. Around 1895, the Austrian function-theorist Wilhelm Wirtinger began to think about ways to generalize the approach to algebraic functions of a single complex variable based on harmonic functions on Riemann surfaces to the case of algebraic functions of two complex variables, i.e. "functions" $z = z(x, y)$ defined by a polynomial equation

$$f(x, y, z) = 0, \quad x, y, z \in \mathbb{C}.$$

Such an approach was very much in the spirit of Felix Klein's views on algebraic functions, and indeed Wirtinger regularly reported on his ideas in his correspondence with Klein. Soon, however, Wirtinger realized that among the many difficulties that had to be overcome, the *topological* ones were crucial. Viewing algebraic functions of two variables as branched coverings of $\mathbb{C}^2$,

$$p : \{(x, y, z) \in \mathbb{C}^3 \colon f(x, y, z) = 0\} \to \mathbb{C}^2, \quad (x, y, z) \mapsto (x, y),$$

Wirtinger tried to characterize the topological situation along the singularity set of such a function (a curve given by the discriminant of f). In particular, Wirtinger was interested in the *local monodromy* of such a covering along the branch curve, i.e. the group of permutations of the values $p^{-1}(x_0, y_0)$ over a point (x_0, y_0), induced by analytic continuation of the function values along small closed loops starting and ending at (x_0, y_0) and avoiding the branch curve. In modern terms, this meant considering, for a neighbourhood U of a branch point with the branch curve removed, the image of the fundamental group $\pi_1(U, (x_0, y_0))$ under the canonical mapping to the symmetric group acting on the fibre $p^{-1}(x_0, y_0)$. However, it should be emphasized that, in the beginning at least, Wirtinger's work was independent of Poincaré's, and the notion of a fundamental group did not appear explicitly.

While Wirtinger noticed that along *regular* pieces of the branch curve, the sheets of the covering were permuted in cyclical order, he recognized that at *singular* points of the

---

37 On Poincaré, see [141, 37, 166], and Chapter 6 in this volume.

38 The exception confirms the rule: In 1918, Mary G. Haseman of Bryn Mawr College published her dissertation on amphicheiral knots of 12 crossings in the footsteps of Tait, Kirkman and Little. She did not mention any of the modern contributions to knots by Tietze and Dehn that had appeared in the meantime.

39 The following paragraphs are mainly based on [40]. For a description of early work related to the Poincaré conjecture, see also [166].

branch curve the sheets could be connected in a more complicated way. Several years went by, however, before Wirtinger managed to work out a paradigmatic example, that given by the equation

$$f(x, y, z) = z^3 + 3xz + 2y = 0.$$

The discriminant of this polynomial is $D_f(x, y) = x^3 + y^2$, which yields a cubic with a cusp as branch curve. In 1905, Wirtinger presented this example to the annual meeting of the *Deutsche Mathematiker-Vereinigung*. The proceedings of the meeting give only the title of his talk, but from his correspondence with Klein and various remarks of later authors on Wirtinger's ideas the gist of what he said on that occasion is quite clear. In order to characterize the topological behaviour of a function like the one above in the neighbourhood of a singular point of its branch curve, Wirtinger brought a new idea into play which he probably had learned from Poul Heegaard's dissertation [69]. Therein, Heegaard described a similar program for a topological study of algebraic surfaces.[40] The idea that interested Wirtinger was to restrict the covering p to the boundary of a small 4-ball around the point in question, that is, to a covering of the sphere S^3, branched along a certain set K of real dimension one, namely the intersection of S^3 with the branch curve of the given algebraic function. In the particular example considered, this covering turned out to be a three-sheeted covering of S^3, branched along the trefoil knot [69, p. 85]! The point of this restriction was that it had the same monodromy as the algebraic function itself, and, moreover, that it allowed one to compute the monodromy group, as a matter of fact, to compute the fundamental group of the base space $S^3 - K$ of the restricted, *unbranched* covering. For his example, Wirtinger obtained the full symmetric group on three elements as monodromy group – and thus the first serious proof that the trefoil knot was actually knotted. It was soon realized, either by Wirtinger himself or by his younger Vienna colleague Heinrich Tietze, that Wirtinger's method actually gave a way to write down a presentation of the fundamental group of the complement of arbitrary knots and links, and not just of the trefoil knot. Moreover, it became clear that this approach could be used to describe the topology of singularities of algebraic curves (algebraic functions of *one* variable) by disregarding the covering obtained by Wirtinger and by taking the branch curve itself as the basic object to be studied.

The importance of the whole argument for the emergence of modern knot theory can hardly be overestimated. Not only had knots appeared in one of the central areas of mathematical interest, but the situation suggested a whole set of new ideas and questions. Together with a knot and its complement, *covering spaces* of knots – either coverings of the 3-sphere branched along a knot or unbranched coverings of knot complements – had come into the picture, including homomorphisms from the knot group to permutation groups. Among the obvious questions were: what kinds of knots could arise in situations like those considered by Wirtinger? What kinds of covering spaces could be obtained in such cases?

Since Wirtinger did not publish his ideas, it took some time before these problems were taken up by others. Wirtinger's basic insight and the main ingredients of the answer to the first question have often been attributed to Karl Brauner, who published a three-part article on the subject in 1928, based on his *Habilitationsschrift* under Wirtinger. Following Brauner, Kähler, Zariski, and Burau simplified and rounded off Brauner's arguments

[40] More information on Heegaard may be found in Chapter 34 in this volume.

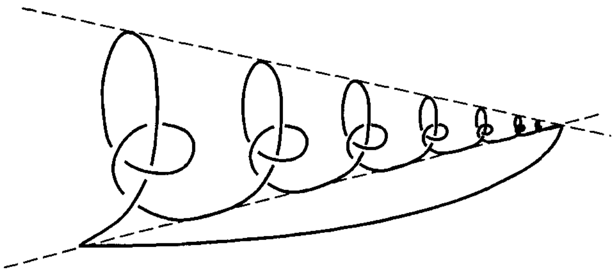


Fig. 7. Tietze's wild knot.

to obtain the final result that singularities of irreducible curves were topologically characterized by iterated torus knots while for reducible curves, links of such knots had to be considered.[41] Moreover, the knots and links arising from singularities of algebraic curves were classified by the pairs of integers determining the Puiseux expansions of the curve branches around the singularity. However, Wirtinger was certainly right when he pointed out in his (unpublished) review of Brauner's *Habilitationsschrift*: "More than twenty years ago, the reviewer has shown the way in which these difficult, but basic problems may be dealt with".[42]

Not only Brauner, but several other mathematicians who made significant contributions to modern knot theory in its early years were also inspired by Wirtinger's insights. Heinrich Tietze, Otto Schreier, Emil Artin, and Kurt Reidemeister all came in direct contact with Wirtinger at some time, and it will become clear below to what extent their ideas were influenced by Wirtinger's.

§ 18. Twenty years before Brauner, another young mathematician presented a *Habilitationsschrift* on topology, guided by Wirtinger in Vienna: Heinrich Tietze. His paper [162] marked a crucial step toward a clear technical understanding of Poincaré's topological ideas. Following a rather coherent, combinatorial approach to the topology of three-dimensional manifolds, Tietze re-established Poincaré's results, emphasizing that all then known invariants of three-dimensional manifolds could be derived from the fundamental group. Among the examples he discussed was Wirtinger's method for finding a presentation of the fundamental group of knot complements, including the example discussed by Wirtinger.[43] In addition, and in the thorough, critical spirit which marks the whole paper, Tietze formulated several basic questions related to knots and three-dimensional manifolds whose answers were unknown at the time.

First, Tietze pointed out in a discussion of Poincaré's definition of the homological invariants of manifolds that certain curves required special attention: For instance, a curve like that of Figure 7 could not be said to bound a finite two-dimensional cell complex in S^3 in Poincaré's sense. This example also made clear that the notion of a knot and of knot equivalence itself required additional care if "wild knots" were to be avoided.

[41] See [20, 82, 175, 25, 26].

[42] Quoted from [38, p. 247].

[43] However, Tietze's description of Wirtinger's ideas was scattered in different passages of his paper which made it hard for his readers to see just what Wirtinger's contribution had been. See [162, §§15, 18].

The second group of questions was inspired by Wirtinger's calculation of the group of the complement of the trefoil knot. Tietze viewed this as a region in $\mathbb{R}^3$, bounded by a torus which bounds on its other side a tubular neigbourhood of the knot. Clearly, both the right-handed and the left-handed trefoil had homeomorphic complements (and, consequently, isomorphic groups), but what about the converse? Could two knot complements be homeomorphic without one knot being isotopic to the other or its mirror image? As Tietze remarked, it was not even clear whether all submanifolds of $\mathbb{R}^3$ bounded by a torus were knot complements [162, § 15].

A third complex of questions arose from Tietze's consideration of the group of self-homeomorphisms of a (closed or bounded) manifold and its quotient by the group of those self-homeomorphisms homotopic to the identity [162, §16]. For oriented manifolds, one could also consider just the orientation-preserving self-homeomorphisms. These groups acted in a canonical way on the fundamental group of the manifold as well as on the fundamental group of its boundary. In the case of several boundary components, the group of permutations of these components induced by this action might also carry interesting information. In this way, a whole new set of topological invariants arose about which very little was known. Tietze illustrated these concepts by considering the complements of collections of disjoint right- and left-handed trefoil knots, pointing out that not even the intuitive belief that the two trefoil knots were inequivalent (a belief that he used in his illustrations) had been rigorously proved.

Finally, Wirtinger's construction suggested yet another way of looking at three-dimensional manifolds, namely as coverings of S^3, branched over a link. Manifolds described in this way were called "Riemann spaces" at the time, generalizing the idea of a Riemann surface (viewed as a branched covering of S^2). It was known that all closed, orientable surfaces could be described in this way; but, Tietze asked, could all closed, orientable 3-manifolds be described as Riemann spaces [162, § 18]?

All of Tietze's questions stressed the relations between knots (or links) and the general study of three-dimensional manifolds. In at least two ways, knots and links gave rise to interesting classes of such manifolds: by their complements, and by covering spaces. It turned out that some questions of Tietze's could be answered rather quickly by the next generation of topologists, while others resisted a solution until very recently.

§ 19. More or less simultaneously with Tietze, Max Dehn, a student of Hilbert who had started his mathematical career with brilliant results on the foundations of geometry, turned to an investigation of 3-manifolds which led him to study knots.[44] In the beginning, Dehn hoped to be able to prove an equivalent to Poincaré's conjecture that S^3 was the only closed, orientable 3-manifold with trivial fundamental group. However, a discussion with Tietze at the International Congress of Mathematicians in Rome in 1908 made clear to Dehn that his arguments were flawed.[45] Nevertheless, he continued to work on the topic. In 1910, he published a paper whose title "Über die Topologie des dreidimensionalen Raumes" indicated that he still hoped to find a topological characterization of ordinary 3-space or the 3-sphere. Instead of trying to prove the Poincaré conjecture, however, he showed how to construct infinitely many "Poincaré spaces", i.e. orientable 3-manifolds bounded by a two-sphere, with vanishing homological invariants but nontrivial fundamental group. In the

[44] On Dehn's career, see Stillwell's contribution to this volume.

[45] See [40] for more details on this.

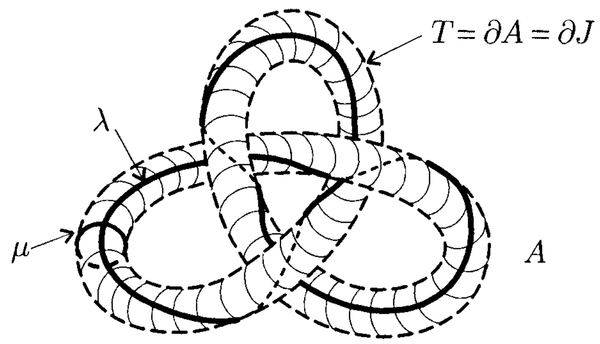


Fig. 8. Dehn's setup.

closing paragraph, Dehn outlined an argument by which he hoped to prove the Poincaré conjecture, directing the attention of his readers to the crucial gap.

While Dehn's paper made clear that the Poincaré conjecture was difficult, it broke new ground both for the study of knots and for combinatorial group theory. In order to construct his examples of "Poincaré spaces", Dehn introduced a presentation of the group of a knot different from Wirtinger's and a new criterion for knottedness. A knot is trivial, he claimed, if and only if its group is commutative. While the "only if" was obvious, the other implication required proof. Here, Dehn argued as follows (compare Figure 8): for all knots K, there exists a longitudinal curve λ on a torus T which bounds a tubular neighbourhood J of the knot, such that λ bounds in the knot complement $A := S^3 - J$.[46] If, in addition, the fundamental group of the knot complement is commutative, then λ must actually be null-homotopic in A. Hence it bounds a singular disk in the knot complement, in such a way, however, that all singularities may be removed from the boundary of the disk. At this point, Dehn invoked the famous, insufficiently proved "lemma" which today carries his name: if a closed curve in S^3 bounds a (piecewise linear) singular disk in such a way that an annulus along the boundary is free of singularities, then this curve even bounds a regularly embedded disk.[47] From the "lemma", then, it followed that λ and hence the knot itself were trivial. Regardless of the difficulties with the lemma, Dehn's criterion could be used to prove that certain knots were non-trivial by showing that their group was non-commutative.

Then Dehn proceeded to consider manifolds arising from the following construction. Let K be a knot in S^3, and let λ, J, T, and A be as above. The generators of the fundamental group of T may be represented by the longitude λ and a curve μ bounding a transversal disk in $J \cup T$. Any other element in this (commutative) group may then be represented by a curve $\lambda^l \mu^m$, or by the pair of integers, (l, m). Dehn now chose a curve ρ of class $(l, 1)$ in T and formed a new manifold $\Phi = \Phi_K(l, 1)$ by attaching to A a thickened disk D (i.e. a 3-cell, whose boundary is considered as the union of an annulus and two disks, see Figure 8) along a small strip on T that forms a neighbourhood of ρ. By construction, $\Phi_K(l, 1)$ is a manifold bounded by a sphere and with trivial homology, i.e. a "Poincaré

[46] My notation is a slightly modernized version of Dehn's.

[47] The gap in Dehn's rather involved argument was recognized in the 1920's, both by H. Kneser and Dehn himself. A sound proof was only given by Papakyriakopoulos [127]. For more details on this, see Chapter 36 in the present volume.

space". Its fundamental group arises from adjoining to the group of the knot the relation expressing the contractibility of ρ. Dehn's presentation of the knot group allowed him to express this relation in a straightforward fashion so that a presentation of the fundamental group of $\Phi_K(l, 1)$ could actually be found.

For the trefoil knot, Dehn managed to construct the graphs of the resulting groups. Of course, $\phi_K(0, 1)$ was just a 3-cell. The group of $\phi_K(\pm 1, 1)$ turned out to be a finite group of order 120, a binary extension of the group of rotational symmetries of the icosahedron.[48] All other manifolds possessed infinite groups. In this way, Dehn found an infinite family of "Poincaré spaces". Moreover, he observed that all of their fundamental groups (like the group of a trefoil knot itself) acted on the hyperbolic plane in a canonical way. Thus, he established a link between knot groups and hyperbolic geometry, a link that he exploited again in a paper of 1914 answering one of Tietze's questions. By a detailed consideration of the automorphisms of the group of a trefoil knot and their actions on longitudes λ and meridians μ of the knot, Dehn showed that the right- and left-handed trefoils were not isotopic.

One should note that Dehn's construction of "Poincaré spaces" is not *quite* the same as what today is usually called "Dehn surgery", since Dehn considered *bounded* manifolds and restricted himself to the case of attaching curves of type $(l, 1)$, a restriction that guaranteed that all manifolds obtained by his construction from knot complements were homologically trivial. I will describe below how the change to the modern point of view came about. It should also be noted that Dehn did *not* ask whether his construction might eventually produce not just a homology cell but even a counterexample to the Poincaré conjecture, i.e. a manifold bounded by a 2-sphere and with trivial fundamental group but topologically different from the 3-ball. Clearly, he still hoped he was on his way toward a proof of this conjecture, rather than a refutation.

§ 20. The geometric motivation behind Wirtinger's, Heegaard's, Tietze's, and Dehn's contributions is obvious. Neither of these mathematicians was motivated by building up a theory of knots *per se*. Rather, they were led to study knots by their research in other areas: research on the singularities of algebraic curves and surfaces, and investigations of three-dimensional manifolds as they had become tractable by Poincaré's new techniques of *Analysis situs*. Knots thus appeared in a rich geometric context, involving covering spaces or Dehn's method for constructing "Poincaré spaces". In both approaches, the knot group played a crucial role, but with different additional structures involved. Some of the problems related to these objects and structures turned out to be quite deep, and several were even too difficult to admit solutions with the methods available at the time. In many ways, later geometric-oriented research on knots, links, and in part also on 3-manifolds, tried to sort out and answer the questions raised in this first phase of a modern mathematical treatment of knots. Some crucial problems remain open even today, as we shall see.

World War I interrupted both Dehn's work and that of the Vienna mathematicians. After the war, two young mathematicians, James W. Alexander and Kurt Reidemeister, became increasingly interested in knots. In several respects, Alexander's and Reidemeister's work were strikingly parallel. Both were led to the same, indeed the first, effectively calculable invariants of knots in the mid-twenties. Moreover, both chose to present their results on the

[48] Using his theory of fibered 3-manifolds, Seifert later showed that Poincaré's original example of a homology sphere was homeomorphic to the closed version of Dehn's manifold $\phi_K(\pm 1, 1)$ [148, pp. 204ff.].

basis of the elementary, combinatorial approach to knots that has been sketched in the second section. To some extent, this parallelism may be traced to the common inspiration they found in Tietze's and Wirtinger's earlier ideas, and in particular to the idea of studying covering spaces of knots or links. On closer inspection, however, their approaches also reveal a basic difference. Probably guided by his earlier work, Alexander was mainly interested in *homological* invariants of covering spaces in his relevant contributions. Reidemeister's crucial insights, on the other hand, were concerned with the *fundamental groups* of such spaces.

§ 21. Alexander's contributions to knot theory began with several clarifications of issues Tietze had raised. During the war years, the Princeton topologist, who had already shown his talents in improving Poincaré's homological results, stayed as a volunteer in Paris and assisted in preparing a French translation of Heegaard's thesis [69].[49] It appears to have been around this time that the problems on 3-manifolds described in Tietze's paper of 1908 caught his attention. The first of Alexander's clarifications was only indirectly related to knots. Still in Paris, he showed that Tietze had been correct in conjecturing that the two "lens spaces" $L(5, 1)$ and $L(5, 2)$ were not homeomorphic [2]. Alexander defined these spaces in a way clearly influenced by Heegaard's dissertation and Dehn's paper of 1910, namely as the manifolds obtained by an identification of the boundaries of two solid tori in such a way that the meridian of one of them gets identified with a curve of type $(5, 1)$ or $(5, 2)$, respectively, in the boundary of the other. Since the fundamental group of both was cyclic of order 5, this showed that the fundamental group was not sufficient to distinguish 3-manifolds in all cases.

About a year later, Alexander claimed in a brief note that every closed, oriented 3-manifold given by a triangulation could indeed, as Tietze had suggested, be obtained as a covering of S^3 branched over a link [3]. His argument was strikingly simple, but incomplete. With each vertex of the triangulation, he associated a point in S^3 such that no four of these points were coplanar. By mapping the simplices of the triangulation onto the simplices of S^3 given by the corresponding vertices and respecting the orientations, Alexander obtained a covering of S^3 branched over a subcomplex of the 1-complex given by the chosen points in S^3 and the edges joining them. "It is easy to show", he continued, "that, without modifying the topology of the space, the branch system may be replaced by a set of simple, non-intersecting, closed curves such that only two sheets come together at a curve. These curves may, however, be knotted and linking" [3, p. 372]. As R.H. Fox later pointed out, the missing part of the argument could be filled in by appealing to a classical argument given by Clifford which showed that every closed Riemann surface – viewed as a branched covering of the complex number sphere – could be deformed into a covering in which only simple branch points of order 2 occur. Alexander's conclusion followed by applying this argument to a continuous family of generic plane sections of the covering obtained in the first step of his argument.[50]

Brief as the argument was, it gave a new and general construction technique for 3-manifolds. Such techniques were still rare and difficult, since triangulations were in some sense too general while the only other known method, Heegaard's decomposition

[49] On Alexander, see [90] and Chapter 32 in this volume.

[50] See [48, p. 213]. Other proofs of Alexander's claim were given by Birman, Hilden, and Montesinos, leading to sharper results, see [72, 117, 73, 118]. Today it is known that there even exist "universal knots", i.e. knots whose branched coverings exhaust all closed, orientable 3-manifolds (Hilden et al. [74]).

of a 3-manifold into two handlebodies, did not seem easy to use except in special cases. Accordingly, Alexander's result whet his interest in links and their covering spaces. Indeed, in November 1920, he presented a new idea for studying *finite cyclic* branched coverings of knots by calculating their torsion numbers. The paper, read to the US National Academy of Sciences, was not published, so that it is difficult to tell precisely what it contained. According to Alexander's own later account, he pointed out that these torsion numbers were actually invariants of the knot or link itself, and he calculated them for a few of the simpler knots. It remains unclear, however, whether he had developed a *general* method to calculate the new invariants.[51]

In 1923, Alexander further refined his picture of "Riemann spaces" by establishing a lemma showing that every link could be deformed into what was later called a closed braid. This lemma had already been demonstrated by Heinrich Brunn at the ICM in Zürich 1897, but Alexander was apparently unaware of Brunn's short note [23]. The implication of the lemma was that "every 3-dimensional closed orientable manifold may be generated by rotation about an axis of a Riemann surface with a fixed number of simple branch points, such that no branch point ever crosses the axis or merges into another" [4, p. 94].[52] A year later, Alexander settled yet another open question of Tietze's by showing that a piecewise linearly embedded torus in S^3 bounds a solid torus on at least one side, making the other side into a knot complement (Alexander 1924).

Up to this point, Alexander was clearly more interested in the 3-manifolds arising from knots or links than in the classification of links themselves. But this changed after Reidemeister's first papers appeared in 1926, describing both a general method for calculating the torsion numbers of a knot from a diagram and the "elementary foundation" of knot theory by diagram moves. In April 1827, Alexander and Briggs submitted their paper "On types of knotted curves" to the *Annals of Mathematics*, describing their own approach to torsion numbers. Although this method was presented in a combinatorial fashion, a closer analysis of the paper makes it clear that Alexander and Briggs were actually guided by Alexander's earlier ideas, and that the calculation was based on an analysis of a suitable cell decomposition of the branched cyclic covering spaces of a knot. I have described in Section 2 how this approach to torsion numbers led to the invention of the first polynomial invariant for knots. As the *infinite* cyclic covering of a knot does *not* appear in [5], it may well be that here, for the first time, Alexander was guided by the combinatorial approach rather than by a geometric one.

§ 22. Also in Reidemeister's case, the combinatorial presentation of his results gives a misleading picture of the actual course of his research. For him, it was an insight into the relation between the unbranched covering spaces of knot complements and the corresponding subgroups of the knot group that opened the way to calculable knot invariants. In 1922, Reidemeister obtained his first professorship in Vienna, and soon afterward he learned of his older colleague Wirtinger's ideas on knots. He began to study Poincaré's writings on *Analysis situs* and organized a seminar on topology and algebra in which he encountered

---

[51] See [7, p. 562]. This account figures in an argument with Reidemeister on priority and must thus be taken with some caution.

[52] Also this conclusion was mathematically related to an earlier idea which Alexander may or may not have known: In 1891, Hurwitz had published a substantial paper studying the deformations of Riemann surfaces (viewed as branched coverings of the complex number sphere) arising from braid-like deformations of their branch points. See below, § 23.

the young Otto Schreier, who was full of ideas about combinatorial group theory.[53] The breakthrough came in 1925, soon after he had accepted a position in Konigsberg (Kaliningrad). In correspondence with H. Kneser, Reidemeister announced that he had found a subgroup of the knot group that possessed nontrivial torsion invariants. This group was in fact the fundamental group of the double cyclic covering of the knot complement. In the following year, Reidemeister worked up his idea into a general method for writing down a presentation of the fundamental groups of finite cyclic coverings of knot complements. The method was based on a combination of Wirtinger's method for presenting the knot group and Poincare's method for calculating the fundamental group of a 3-manifold given by a cell decomposition. The surprising fact was that, in contrast to the knot group itself, many of the subgroups obtained in this way had nontrivial torsion invariants.

In preparing the publication of his results, Reidemeister tried to present his ideas in as abstract a fashion as possible. This led him to recognize that his method for calculating subgroups of the knot group could actually be made into the method for calculating subgroups of finitely presented groups that today is known as the "Reidemeister-Schreier method" (Reidemeister [134]).[54] Moreover, he developed his "elementary foundation" of knot theory [135], a manner of presentation that was at least partially motivated by his philosophical interests in the foundations of mathematics. In Vienna, Reidemeister had become one of the early members of the philosophical circle around Hans Hahn and Moritz Schlick. During the foundational debates of the twenties, he was engaged as a convinced "modernist", emphasizing that all exact knowledge (i.e. in his view, mathematics and logic) was ultimately rooted in "combinatorial facts" about signs.[55] Little wonder, then, that Reidemeister favoured a combinatorial approach to topology.[56]

§ 23. Before arriving in Vienna, Reidemeister had held an assistant professorship in Hamburg, where a new university had been founded in 1919. Its mathematical department was directed by Wilhelm Blaschke, who received his doctorate in Vienna under Wirtinger, and by Erich Hecke, a student of Hilbert. Hamburg quickly emerged as a lively mathematical center during the 1920's, and in 1922, the department began to publish its own journal, the *Abhandlungen aus dem Mathematischen Seminar der Hamburgischen Universitat.* Reidemeister's papers of 1926 were published in this journal, and it became the main forum for knot theoretical research during the following years. Hamburg's ties with Vienna were also particularly close. At about the time when Reidemeister left for Vienna, another Viennese mathematician came to Hamburg, Emil Artin, followed soon afterward by Otto Schreier. For a short period, the two worked together on a group-theoretical problem related to knots: the classification of braids, or the word problem in the braid group.

Since the paper describing the fruits of this work, [9], has often been taken as documenting the invention of the braid group, a few words should be said about earlier interest in braid-like topological objects and related groups. As pointed out in Section 3, Gauss was probably the first to consider braids (i.e. a collection of *n* disjoint, smooth curves in

[53] An outcome of this seminar was Schreier's very simple group-theoretical proof that the two trefoils were inequivalent [142]. On Schreier, who died in 1929 at age 28, see [28, Chapter II.3].

[54] For a historical description of the various stages in which this method reached its final shape, see [28, Chapter H.3].

[55] A pronounced statement of Reidemeister's philosophical views can be found in [136].

56 jk e "purely combinatorial" approach to topology was first advocated by Dehn and Heegaard [33], strongly inspired by the style of Hilbert's *Grundlagen der Geometrie.*

Euclidean space such that every member of a continuous family of parallel planes intersects each curve in precisely one point) as objects of topological interest. His unpublished fragment may actually be read as posing the problem of classifying braids up to a suitable notion of equivalence. Later, both Listing and Tait were interested in similar geometric objects but failed to prove substantial results about them. As with knots, it was the wish for a geometric understanding of algebraic functions that motivated mathematicians to dig deeper. Since the appearance of Puiseux's contributions [133], the idea became commonplace that the behaviour of an algebraic function of one complex variable, given by a polynomial equation $f(x, y) = 0$, could be studied by looking at the simultaneous motions of the finitely many values $y \in \mathbb{C}$ that arise when the argument x describes loops starting and ending at a given point $a \in X$, where $X \subset \mathbb{C}$ is the complement of the set of branch points of $f(x, y) = 0$. Puiseux and most authors following him were interested in the "monodromy group" of $f(x, y) = 0$, i.e. the group of permutations of the n roots of $f(a,y) = 0$ arising from all such loops. Once the conceptual apparatus of the fundamental group and the braid group became available, it was easy to see that the propositions proved by Puiseux actually yield homomorphisms

$$\pi_1(X,a) \to B_n \to E_n$$

where B_n is the n-strand braid group, E_n is the symmetric group on n elements, and the image of the composite homomorphism is the monodromy group. In other words, even if the notion of the braid group had not yet been defined, monodromy considerations led to knowledge concerning motions of configurations of complex numbers (we might say "braid motions") that was later encoded in the braid group.

In 1891, Adolf Hurwitz published a paper on (closed) Riemann surfaces, understood as branched coverings of the complex number sphere with finitely many sheets and a finite number n of branch points. Among other things, he investigated deformations of such coverings arising by a continuous change of the configuration of branch points, starting and ending at a given configuration. Thus he was again led to consider both braid motions and the special kind of braid motions where each point returns to its original position (in modern terms: motions corresponding to *pure* braids). Hurwitz went a step further than earlier authors by considering pure braid motions as loops in the configuration space

$$C^n \quad (x_1, \dots, x_n) \in \mathbb{C}^n \mid \prod (x_i - x_j) = 0$$

whereas he thought of general braid motions as loops in the quotient of this space by the canonical action of the symmetric group E_n. Still, the braid group did not appear explicitly in his paper. Instead, for two given natural numbers n and m, Hurwitz considered the set of Riemann surfaces with m sheets and n branch points, each surface being specified by the n sheet permutations $s_1, \dots, s_n \in E_m$ associated with the n branch points. He managed to give rules for determining the "monodromy groups" of permutations of the surfaces, induced by either braid or pure braid motions. Like Wirtinger, who seems to have studied the monodromy of coverings of a knot complement without explicitly discussing the knot group, Hurwitz seems to have been unaware that his rules actually determined the braid and pure braid group itself. In the case of braid motions, his result was that the "monodromy group" in question was generated by the permutations of surfaces,

$$\sigma_i := \begin{pmatrix} s_1 & \dots & s_i & s_{i+1} & \dots & s_n \\ s_1 & \dots & s_i s_{i+1} s_i^{-1} & s_i & \dots & s_n \end{pmatrix}, \quad 1 \leqslant i \leqslant n-1. \qquad (\star\star)$$

Since the sheet permutations s_k were left unspecified in Hurwitz' argument, the group generated by the σ_i may be understood as a group of automorphisms of the free group on n generators which is in fact a faithful representation of the braid group as Artin would show. Again, while the notion of braid group itself was absent, insights were developed that could immediately be transformed into knowledge about this group once it was defined.

The group-theoretic structure ($\star\star$) obtained by Hurwitz reappeared with a different interpretation in another context of ideas, relating to transformations of Riemann surfaces onto themselves Fricke and Klein [55, pp. 299ff.]. In this work, the idea of "braid motions" was less visible, but soon after the explicit definition of the braid group, Wilhelm Magnus showed how some of Fricke's and Klein's ideas could be translated into a connection between the mapping class group of the n times punctured plane or sphere and the braid group [98].

In the early twentieth century, the idea of what I have called "braid motions" was certainly well known to most mathematicians interested in algebraic functions and related issues. We have seen that it served Alexander to study 3-manifolds, and certainly Artin and Schreier were acquainted with it, too. In this light, Artin's geometric definition of the n-strand braid group B_n and his presentation of B_n as the group generated by $n-1$ elements $\sigma_1, \dots, \sigma_{n-1}$, with relations

$$\sigma_k \sigma_l = \sigma_l \sigma_k \quad \text{if } |k-l| \geqslant 2;$$

$$\sigma_k \sigma_{k+1} \sigma_k = \sigma_{k+1} \sigma_k \sigma_{k+1} \quad \text{for } k = 1, \dots, n-2,$$

appears less as an invention out of the blue but rather as a properly topological or group-theoretical definition of a known structure. The emphasis of Artin's paper was clearly on translating the geometric questions about braid motions into purely group-theoretical questions. In particular, the classification of braids up to the appropriate kind of isotopy was restated as the word problem of the braid group, while the classification of *closed* braids amounted to the conjugacy problem of the braid group. On the other hand, Artin probably knew Hurwitz' paper and its geometric techniques. In order to solve the main problem of his paper, the word problem in B_n, Artin used precisely the representation of braids as automorphisms of the free group on n generators that Hurwitz had (almost) defined. The crucial step was to show that this representation is indeed faithful, and here Artin relied on a topological argument quite close to some of Hurwitz' ideas. Moreover, when looking at closed braids, Artin used Wirtinger's presentation of a knot group as it had been communicated to him by Schreier.

Thus, on the whole it is clear that, as with the case of knots, interest in braids was closely tied to a geometric approach to algebraic functions; the latter provided the background for the investigations of the combinatorial and group-theoretic aspects that came into focus after Artin had published his paper.

§ 24. After Alexander's homological and Reidemeister's and Artin's group-theoretical contributions had shown how to construct calculable invariants of knots and links, the prospects for developing a "theory of knots" in its own right seemed promising. Many problems seemed tractable, and Dehn's and Alexander's results connecting knots with 3-manifolds, as well as the work inspired by Wirtinger on singularities of algebraic functions established sufficiently many links to other fields to convince others that knot theory was an interesting subject. At the same time, the piecewise linear, combinatorial approach to knots used by Reidemeister and Alexander made it possible to develop knot theory without entering the intricacies of these other fields too deeply. This enabled newcomers to join the enterprise. Indeed, both at Königsberg and Hamburg a number of students of Reidemeister and Artin started to work on knots, and the flow of papers to the *Hamburger Abhandlungen* increased steadily. In 1932, Reidemeister's monograph *Knotentheorie* summarized the results obtained until then (leaving out most connections to other fields, however) and provided a kind of "paradigm" in the sense of Thomas Kuhn for the young field. On the mathematical level, this period of flourishing activity was oriented toward a finer study of particular classes of knots or links (such as the links arising from singularities), a better understanding of the invariants that had been constructed, and a discussion of their power in distinguishing knots and links.

A significant contribution to the understanding of Alexander's invariants was made by Herbert Seifert, exploiting a geometric idea that lay dormant since Tait's days. While it had long been known that surfaces embedded in space and bounded by an arbitrary knot could be found (this followed for instance from Tait's observation that every knot diagram could be coloured in a chequerboard-like fashion), an additional argument was needed to show that *oriented* surfaces bounded by a given knot existed as well. A procedure to find such a surface was described by Frankl and Pontrjagin [52]. Seifert saw that one could use such surfaces – today called Seifert surfaces – for the construction of cyclic coverings of a knot complement and hence for a calculation of homological knot invariants [150]. In particular, Seifert was the first to describe the Alexander polynomial in terms of the first homology group of the *infinite* cyclic covering of a knot complement. He showed that this group could be viewed as a module over the ring $\mathbb{Z}[x, x^{-1}]$, and that the Alexander polynomial was given by the determinant of a presentation matrix of this module. Seifert's construction also made it possible to obtain information about the minimal genus g_K of Seifert surfaces, an invariant of the knot K which he called its "genus". For all knots, the degree of the Alexander polynomial was a lower bound for $2g_K$. Since a more or less sharp upper bound on g_K could be read off a diagram, this enabled calculations of the genus of many knots such as the torus knots and all knots of up to 9 crossings. Moreover, Seifert was able to describe a nontrivial knot all of whose cyclic coverings were homology spheres. This showed that the Alexander polynomial was not sufficient to detect knottedness [150, § 4].

Already in an earlier paper, Seifert had observed that the two composite knots presented in Figure 9, without being mirror images of each other, had the same group. Seifert distinguished these knots by a new type of "linking invariants", computed from the torsion subgroup of the first homology group of cyclic coverings of the knots [149]. This showed that the group of a knot was not a complete invariant, at least for composite knots. Moreover, Tietze's question whether a knot was determined by the homeomorphism type of its complement also remained a mystery. By a rather simple example, J.H.C. Whitehead pointed out in 1936 that the analogous statement for the case of *links* was false [171], and thus the answer to Tietze's question seemed quite unclear (see Figure 10).

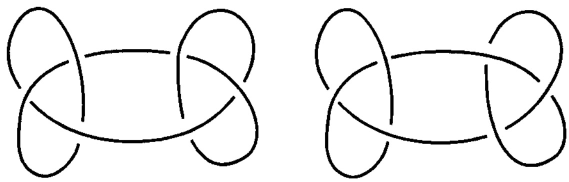


Fig. 9. Knots with the same group.

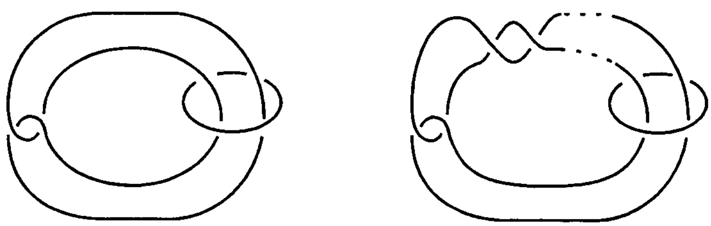


Fig. 10. Whitehead's links having homeomorphic complements.

Yet another way of looking at the Alexander polynomial came into view when Werner Burau found a rather surprising connection of this invariant to braids. Burau showed how the Alexander matrix of a link, represented by a closed braid of n strands, could be calculated from a linear representation of the braid group,

$$\beta : B_n \to GL(n, \mathbb{Z}[x, x^{-1}]),$$

that today carries his name [24].[57] In particular, if a knot K could be represented by closing a braid w, then up to a normalization, its Alexander polynomial was given by $\Delta_K(x) = \det(\beta(w) - I)$. Further light was thrown on the relation between braids and knots by a conjecture of the Russian mathematician A.A. Markov at the congress on topology in Moscow in 1935 [102]. He claimed that two closed braids, given by elements $v \in B_m$ and $w \in B_n$ in different braid groups, represented isotopic links if and only if v and w could be related by a sequence of modifications

$$a \leftrightarrow bab^{-1} \ (a, b \in B_k) \quad \text{or} \quad a \in B_k \leftrightarrow a\sigma_k^{\pm 1} \in B_{k+1}.$$

At the time, Markov's conjecture was not seriously pursued nor was it related to Burau's results. Only much later, Joan Birman included a full proof of it in her book [17].[58]

In Germany, this period of a rapid development of the young field was ended by the consequences of the Nazi regime's takeover. Already in April 1933, Reidemeister lost his professorship in Königsberg for being "politically unreliable".[59] After a lapse of a year, he obtained a new position in Marburg, but he had lost most of his Königsberg students and spent his Marburg years in growing isolation. Some of Reidemeister's students moved

[57] Joan Birman reports that Burau had learned of this representation either from Reidemeister or from Artin.

[58] The proof was based on notes taken at a seminar at Princeton University in 1954 [17, p. 49].

[59] See [39] for a description of the circumstances.

to Hamburg, but there, the situation was difficult as well. In 1937, Artin and his wife had to leave Germany because she was Jewish. After the pogroms of November 1938, Dehn, too, was forced to flee from Frankfurt under rather dramatic circumstances (Siegel [151]). Seifert was ordered by the German ministry of education to go to Heidelberg in 1937. There, the Nazis had driven the two Jewish professors of mathematics, Liebmann and Rosenthal, out of their positions. This interrupted Seifert's productive collaboration with William Threlfall. After the war broke out, research on knot theory was also abandoned outside Germany. Topologists like Alexander and Whitehead took over new tasks in the military and left knots and links behind.

§ 25. A look back on the events described in this section shows how deep the changes in the mathematical treatment of knots were that occurred between the late nineteenth and the early twentieth century. Research on knots needed no longer to be justified by its function in scientific contexts beyond mathematics. New kinds of mathematical objects and techniques definitely transcended the limits imposed by thinking of knots and knot complements as figures or regions in physical space. Moreover, an impressive range of problems could be dealt with in a rather rigorous way and with promising results. All these aspects point to the modernity which the new field shared with much contemporaneous mathematics.

A particular shade of this modernity is also visible in Reidemeister's successful attempt to build up knot theory in a very autonomous, combinatorial fashion, the Hilbertian roots of which can easily be discerned. Nevertheless, I hope to have made clear that both the main motivations and the complex mathematical objects that allowed mathematicians to reach a deeper understanding of knots did not originate in this "elementary" way. They came from the highly valued field of algebraic functions and from Poincaré's ideas on three-dimensional manifolds. Neither Alexander's nor Reidemeister's nor Artin's innovations would have been possible had they not been acquainted with the corresponding ideas of mathematicians like Hurwitz, Wirtinger, Dehn, or Tietze. For this reason, the combinatorial shade of modernity should not be overplayed in our understanding of the emergence of modern knot theory.

5. Some geometric topics in knot theory after 1945

§ 26. While the emergence of modern knot theory in the early decades of the 20th century can be described as a relatively coherent fabric of events, the further development of knot theory becomes increasingly complex. In part, this results from the fact that knot theory did not attain the status of a self-sustaining subfield of mathematics, with its own separate domain of problems and methods, with its own publication forums, institutional networks, etc. Rather, research on knots remained tied to the broader development of low-dimensional topology, especially the theory of 3-manifolds. This holds both with respect to mathematical ideas and with respect to the social setting of work on knots. Although there emerged a group of experts in knot theory, most of the important work was done by mathematicians who had interests in other areas as well. Consequently, a historical narrative sensitive to issues of motivation and context cannot isolate knot theory after 1945 from the spectrum of related mathematical activities. This makes our subject both interesting and difficult. The overwhelming proliferation of mathematical research during the second half of this century, in which the discipline of topology played a crucial role, is reflected in the

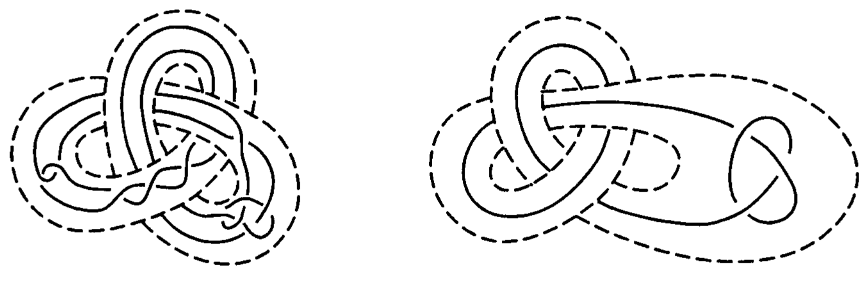


Fig. 11. Two satellites of the trefoil knot.

broad range of new concepts and techniques that were formed in order to deal with knots. The following paragraphs try to capture a part of this complexity. Guided by the main theme of this article, the following groups of issues will be discussed: the systematization of the categorical framework of knot theory by Ralph H. Fox and his Princeton group (§ 27); John Milnor's work related to knots (§ 28); the use of surgery techniques and the relation between knots and the Poincaré conjecture (§ 29); the ideas of Wolfgang Haken and others that led to a proof that knots can be classified algorithmically (§ 30); the discovery of the hyperbolic structure of most knot complements by Robert Riley and William Thurston (§ 31); and the path leading from von Neumann's construction of the "hyperfinite II_1 factor" to Vaughan Jones's new knot invariant (§ 32). Several further developments are passed over in silence, but the narrative should enable the reader to perceive the main geometric impulses that induced 20th-century mathematicians to investigate knots.

§ 27. For many years after the defeat of the Nazi regime in 1945, knots and links did not play a significant role in mathematical research in Germany. Reidemeister gathered a new group of research students with interest in low-dimensional topology and knot theory only after accepting a professorship in Göttingen in 1955, at the age of 61. In Heidelberg, where Seifert was teaching, Horst Schubert stood out as the major exception from the rule that mathematics in Germany now had other concerns than knots. Schubert had begun his studies during the war with Threlfall in Frankfurt, and followed him when Threlfall received a call to Heidelberg in 1946. In his dissertation, Schubert showed that knots formed a commutative semigroup with unique prime decomposition under the product operation given by tying two separate knots on the same string [143]. Schubert's *Habilitationsschrift* [144] treated knots K embedded in a solid torus J that formed a tubular neigbourhood of another non-trivial knot K', such that K could not be deformed into K' or the unknot by isotopies within J. The knot K' was called a "companion knot" of K by Schubert, while K later became to be called a "satellite" of K' (see Figure 11 for two examples; note that the right one indicates that all product knots are satellite knots). Schubert showed how certain invariants of satellite knots, such as their genus, were related to the corresponding invariants of their companions. A year later, Schubert introduced a new invariant, the "bridge index" of a knot [145]. This invariant could be defined from knot diagrams, namely as the least number of diagram arcs that extend from one undercrossing to the next while passing at least one overcrossing in between. Like the crossing number or the genus of a knot, the new invariant was in general difficult to calculate, but Schubert was able to give a complete classification of knots with bridge index 2 [146].

In contrast to the situation in Germany, knots quickly received attention in the United States. Soon after the war, a new center for research in knot theory developed in Princeton. There, the earlier local tradition in topology, the emigration of German mathematicians, and the new international contacts created a favourable environment for research on knots and related topics. From 1946 to 1958, Artin taught at Princeton University, and in the late forties, Reidemeister and Seifert also stayed at the Institute for Advanced Study for some time. But it was Ralph H. Fox who became the central figure in a group of young mathematicians interested in knots, links, and three-dimensional topology. Fox had obtained his doctorate under Solomon Lefschetz before joining the Princeton faculty in 1945. For a certain period, he closely collaborated with Artin, who was then reconsidering his earlier work on the braid group. Together, they published an article on "Some wild cells and spheres in three-dimensional space" which raised at least two important issues. On the one hand, they asked for a clear delineation of the domain of knot theory within three-dimensional topology; on the other, it hinted at a relation between the complements of (maybe wild) knots or knotted arcs and the Poincaré conjecture [49].[60] During the following years, many of Fox's students started their careers with contributions to knots and links, often with a view toward 3-manifold theory.[61]

At the International Congress of Mathematicians in 1950, Fox presented a first survey on the work of his group, which began with a criticism of the combinatorial fashion in which knot theory was conceived during the 1920's and 1930's:

> This description of what I may call classical knot theory tends, by its narrowness, to isolate the subject from the rest of topology. It is to be hoped that the various special theorems which make up classical knot theory will eventually turn out to be particular cases of general topological theorems. In working toward this end the following principles seem almost obvious: (A) *The class of polygons should be replaced by a suitable topologically defined class of curves.* [...] (B) *Euclidean 3-space should be replaced by other compact 3-manifolds.*[62]

The interest of Fox and some of his students in wild arcs was tied to this desire to redefine the objects of knot theoretical studies. When in the early 1950's, Edwin Moise proved that topological 3-manifolds could be triangulated and that, moreover, the "Hauptvermutung" of combinatorial topology was true in this case, it became clear that "classical" knot theory could indeed be reformulated according to Fox's ideas as the theory of (orientation preserving) homeomorphism classes of (oriented) tame simple closed curves in S^3 (or a different 3-manifold).[63] Before these clarifications, Fox had proposed to work with isotopy classes of smooth curves and conjectured that every smooth curve was actually tame. A proof of this conjecture was later included in [32].

This successful effort to readjust the foundation of knot theory must be seen in the context of a general reaction to the earlier, purely combinatorial style of low-dimensional topology. As R.H. Bing put the matter in an inspiring paper that will be discussed below,

---

[60] The terminology of "tame" and "wild" curves in a 3-sphere was introduced in this paper. A curve, surface, or domain in S^3 was said to be "tame" if and only if it could be transformed into a simple polygon, polyhedral surface, or solid polyhedron by a self-homeomorphism of S^3, respectively, and "wild" if this was not the case.

[61] A list of Fox's research students is given in the second volume of Milnor's *Collected Papers*, dedicated to Fox [114, vol. 2, p. xi]. A look at the bibliography of Burde and Zieschang [27] shows that almost all of them worked on topics related to knots, links, braids, or higher dimensional analogues.

[62] [45, p. 453]. Emphasis in the original.

[63] This was pointed out by Moise himself at the ICM 1954 [115]; a proof appeared the same year [116].

the trend was to "regard a 3-manifold as a concrete object [described by appropriate topological constructions] rather than an abstraction of combinatorially equivalent systems of symbols" [12, p. 17]. Of course, this desideratum was particularly easy to fulfill in the three-dimensional case, once it was clear that triangulations existed and the "Hauptvermutung" was true. At any rate, the new perspective on knots advocated by Fox tended to make explicit the integration of knot theory into the broader field of low-dimensional topology. As I have described above, a similar view had also guided the research of the pioneers of modern knot theory, but this perspective had more or less vanished from the printed texts of "classical" knot theory of the 1920's and 30's (note how quick a "modern" approach had become "classical").

In his talk at the ICM in 1950, Fox also reported on certain new ideas developed at Princeton concerning the algebraic structure of the knot group and their presentations. At the time of his talk, these ideas had not yet appeared in print, but during the following years Fox gradually unfolded them in a series of articles. One of his guiding ideas, for a knot in a 3-manifold M represented by a knotted solid torus J with boundary ∂J, was to consider the commuting diagram of homomorphisms:

$$\begin{array}{ccc} \pi_1(\partial J) & \rightarrow & \pi_1(J) \\ \downarrow & & \downarrow \\ \pi_1(M-J) & \rightarrow & \pi_1(M) \end{array}$$

where the arrows were given up to a conjugation in the respective image by the canonical embeddings of manifolds. In the case of knots in S^3, the information contained in such a diagram was already captured by the conjugacy class of subgroups of the knot group generated by the homotopy classes of a meridian and a longitude of the knot. Such subgroups Fox called the (maximal) peripheral subgroups of the knot group. He conjectured that all known knot invariants could be derived from the knot group together with the class of its peripheral subgroups. He also mentioned that Dehn's proof of the inequivalence of the two trefoil knots could be interpreted as an argument about peripheral subgroups. Moreover, he reported that he had been able to show that no automorphism of the group of the two knots discussed by Seifert (see Figure 9) preserved peripheral subgroups.[64] Fox further suggested that a proof of his conjecture might possibly depend on a proof of Dehn's lemma. As it turned out, he was right, but even after Dehn's lemma had been saved by C.D. Papakyriakopoulos in 1957 it took a long time and hard work to establish that the answer to Fox's question was affirmative, as will become clear from what follows.

Next, Fox mentioned an algebraic tool for the study of group presentations, his so-called "free differential calculus", by which not only Alexander's polynomial could be investigated but also the finer structure of the "elementary ideals" of the group ring $\mathbb{Z}[x, x^{-1}]$ of the abelianized knot group associated with a given presentation of the knot group. This calculus, first discussed in a series of papers starting to appear in 1953, was made popular by two publications that did much to disseminate the Princeton group's work on knot theory: Fox's "A quick trip through knot theory" (1962) and the *Introduction to Knot Theory* by Fox and his former student R.H. Crowell (1963), the first monograph on knot theory since Reidemeister's book. Together with his "Quick Trip", Fox published a list of open problems on knots. The two most fundamental were: (1) Tietze's old question, "Is the type

[64] The argument, based on a discussion of the representations of this group in the symmetric group on five elements, was published in [46].

of a knot determined by the topological type of its complement?", and (2), the new, complementary problem whether the topological type of a knot complement was determined by the knot group and its peripheral subgroups.

§ 28. Toward the end of his 1950 talk, Fox had also reported on the work of a then 19-year-old student that fell somewhat outside the range of topics otherwise described: John W. Milnor's study of the total curvature $\kappa(K)$ of a knot K [110]. Using a definition of total curvature applicable to *any* continuous closed curve, Milnor showed that, when K varied in its isotopy class $\mathfrak{K}$, the greatest lower bound of $\kappa(K)$ was a positive integer multiple of 2π, and equalled 2π only if the knot was isotopic to a circle.[65] Thus the integer

$$\mu_K := \inf_{K \in \mathfrak{K}} \frac{\kappa(K)}{2\pi}$$

was a knot invariant that, for the first time in the development of knot theory, involved a notion from differential geometry. Milnor showed how to relate this invariant to a Morse-theoretic view of knots. He began with the observation that every knot in a generic position in space attains a finite number of height maxima with respect to a given axis. The minimum number of such maxima, which Milnor called the "crookedness" of a knot, was just μ_K. As a matter of fact, it was not difficult to see that the crookedness of a knot and its bridge index (defined by Schubert a little later) were the same numbers. This gave a nice example of how a combinatorial knot invariant could have a geometric meaning.[66]

In his master's and doctoral theses [111, 112], written under Fox's direction, Milnor dealt with a new geometric idea concerning links. In order to describe this, it may help to look back at Gauss's linking number briefly. It was clear that this number was not only invariant under ambient isotopies of the link, but also under deformations where each component of the link might cross itself, but no two components were allowed to have mutual intersections. Such deformations were called "link homotopies" by Milnor. Invariants under this kind of deformation captured information about the proper "linking phenomena" in links, disregarding the possible knotting of individual link components.[67] By considering the factor group G/G_q of the fundamental group G of the link complement by its qth lower central subgroup, Milnor was able to define certain new numerical invariants of link homotopy, depending not just on two components of a link but on finitely many. These "higher linking numbers" represented a generalization of Gauss's invariant, i.e. for the special case where only two link components were considered the definitions were equivalent. A geometric ingredient in Milnor's technical arguments that documents the influence of Fox's ideas was the essential use of longitudes and meridians of the link components.

In 1957, Fox and Milnor together published a short note in the *Bulletin of the AMS* in which a new research theme was announced that would occupy Milnor's attention repeatedly during the following years. It concerned the relation between knots and singular points

[65] This confirmed the conjecture of Borsuk [19] that the total curvature of a non-trivial knot was bounded from below by 4π. Borsuk's conjecture was proved independently by Fáry [44].

[66] Strangely enough, Schubert originally claimed that his bridge number was independent of Milnor's crookedness [145, p. 245].

[67] The idea to look at this kind of deformations had already appeared in a dissertation by Erika Pannwitz in 1931, written under the direction of Otto Toeplitz. Using the idea, Pannwitz showed that there always exist lines in space intersecting a non-trivial two component link (or a knot) in at least four points [126].

of surfaces in a four-dimensional manifold, a problem that had stood at the beginnings of modern knot theory as we have seen in Section 3. Also with respect to this topic, Artin functioned as a mediator between the earlier generation of knot theorists and the Princeton mathematicians. In 1925, Artin wrote a brief paper in which he discussed the purely topological aspects of the situation as considered earlier by Heegaard and Wirtinger [10]. Artin pointed out that, contrary to the beliefs of some of his contemporaries, knotted surfaces in $\mathbb{R}^4$ (in particular, knotted spheres) did exist. To construct examples, he introduced a technique later called "spinning": a knot or a knotted arc in a half space in $\mathbb{R}^3$, bounded by a plane E, was "rotated" in $\mathbb{R}^4$ about E. The surface covered by the moving knot was then a knotted surface in 4-space. Moreover, Artin pointed out that the kind of singularities discussed by Heegaard and Wirtinger could be described in purely topological terms, without reference to algebraic functions. Given a point in a piecewise linear, closed surface F embedded in $\mathbb{R}^4$, the intersection of F with the boundary S^3 of a small 4-ball around the given point was a knot whose isotopy class in S^3 was a complete invariant of the surface point with respect to deformations in $\mathbb{R}^4$.[68] If the knot was non-trivial, the point could be considered as a "combinatorial singularity" of F, as Artin called it. Examples could be obtained by forming the "cone" on a given knot in $\mathbb{R}^3$, i.e. by joining all points of the knot by straight line segments with a vertex in $\mathbb{R}^4$ outside the hyperplane containing the knot.

In their research announcement, Fox and Milnor proposed to study these kinds of local singularities of piecewise linear embeddings of oriented 2-dimensional manifolds into piecewise linear, oriented 4-dimensional manifolds more closely. They claimed that a collection of knots $K_1, K_2, \ldots, K_n$ could arise from singularities of a 2-sphere in $\mathbb{R}^4$ if and only if the product knot $K_1 K_2 \ldots K_n$ could be obtained from a single singularity. This gave rise to the introduction of a new concept and a new equivalence relation among knots. A knot obtained from a *single* singularity of a 2-sphere, or, equivalently, as the boundary of a non-singular disc, embedded in a half space of $\mathbb{R}^4$ bounded by a hyperplane containing the knot, was called a "slice knot".[69] Two knots K_1 and K_2 were called equivalent, if and only if the product $K_1(-K_2)$ of K_1 with the "inverse" of K_2 (i.e. its mirror image with reversed orientation) was a slice knot. The equivalence classes of knots under this relation formed a commutative group. Fox and Milnor remarked that a necessary condition for a knot K to be a slice knot was that its Alexander polynomial had the form $\Delta_K(x) = p(x)p(x^{-1})$ for some $p \in \mathbb{Z}[x]$. This allowed them to conclude that the new group was not finitely generated.

In 1966, Fox and Milnor published a more detailed paper summarizing their ideas in a revised and extended form. There, they also showed that the new equivalence relation could be regarded as a kind of relative cobordism relation between knots: two oriented knots were equivalent if and only if they could be placed in two parallel hyperplanes in $\mathbb{R}^4$ such that in the region of 4-space between these hyperplanes, a non-singular, oriented annulus could be found which was bounded by the two knots (with correct orientations). Accordingly, Fox and Milnor proposed to call their group the *knot cobordism group*.

In the years between the authors' first announcement and the paper of 1966, their ideas on knot cobordism had been communicated to several other people, and in particular, to a group of mathematicians working in Japan. This connection had been established in the

[68] Here, Artin's claim was necessarily vague. As Fox and Milnor [51] pointed out, it was only clear that the knot was a "combinatorial" invariant of the embedding, i.e. unchanged by piecewise linear deformations.

[69] This last term was actually absent from Fox's and Milnor's announcement, but was introduced in Fox's "Quick Trip".

late 1950's by Fox, and during the 1960's a great number of articles on knots appeared in the *Osaka Journal of Mathematics*. Many of them focused on the brand new topic of slice knots. Perhaps the most important outcome of this research was a paper by Kunio Murasugi in which the signature of knots – the signature of a quadratic form derived from the first homology group of a Seifert surface of minimal genus – was discussed and shown to be a cobordism invariant [120]. Since then, slice knots and knot cobordism continued to form a focus of research at the interface between knot theory and 4-manifolds.[70]

Milnor's interest in the relation between knots and singularities took a new turn after Egbert Brieskorn, using techniques analogous to the Heegaard–Wirtinger construction, showed that certain algebraic varieties yielded examples of exotic spheres. Brieskorn considered the intersection of the varieties

$$V_n := \left\{(z_1, \ldots, z_{n+1}) \in \mathbb{C}^{n+1} \,\middle|\, z_1^3 + z_2^2 + \cdots + z_{n+1}^2 = 0\right\}$$

with the boundary S^{2n+1} of a ball centered at the origin, giving rise to smooth manifolds homeomorphic to S^{2n-1} and knotted in S^{2n+1}. Brieskorn showed that for certain n, for instance $n = 5$, these knotted spheres were even exotic, i.e. their differentiable structure inherited from $\mathbb{C}^{n+1}$ was inequivalent to the standard differentiable structure on S^{2n-1} [22]. Milnor set out to study the singularities of complex hypersurfaces, i.e. zero sets of polynomials, along similar lines [113]. His basic result was a fibration theorem: if S_ε was a sphere of sufficiently small radius ε around an arbitrary point $\vec{z}^0 = (z_1^0, \ldots, z_{n+1}^0)$ of a complex hypersurface V given by $f(\vec{z}) = 0$, and if K denoted the intersection $V \cap S_\varepsilon$, then $S_\varepsilon - K$ was a smooth fibre bundle over S^1, with projection mapping $\phi(\vec{z}) = f(\vec{z})/|f(\vec{z})|$, having a smooth, parallelizable $2n$-manifold F as fibre. From this theorem, further information on the algebraic topology of the singularity could be drawn. In the "classical case" of an isolated singularity of an irreducible, complex algebraic curve characterized by a knot K, Milnor's theorem implied that the complement $S^3 - K$ admitted a fibration by Seifert surfaces of minimal genus. Using another deep result of the Princeton school, a theorem of Neuwirth and Stallings characterizing knots with complements fibred over S^1, Milnor concluded that the commutator subgroup of the group of K was a finitely generated free group whose rank μ equalled the degree of the Alexander polynomial of K. Moreover, μ was twice the genus of the fibre F, i.e. the genus of the knot [113, p. 84]. Milnor's main interest, though, concerned higher-dimensional generalizations of this situation.

In all of Milnor's contributions to knot theory, a strong component of geometric thinking is clearly visible. The essential new ideas – curvature, link homotopy, knots as invariants of local singularities of surfaces in 4-manifolds, knot cobordism, Milnor's fibration – were all of a geometric character. There can be no doubt that it was this aspect that made his work so fruitful in stimulating further research. Also Brieskorn's examples, weaving together algebraic geometry, knotted spheres in higher dimensions, and exotic differentiable structures, gave a significant impulse to research in all fields concerned.[71]

§ 29. Another series of new researches on knots and their role in the theory of three-dimensional manifolds was initiated when Dehn's technique for constructing "Poincaré spaces" was elaborated in the early 1960's. As in Dehn's case, the main impulse to do this

[70] See, for instance, the long list of problems relating to this topic in [87].

[71] On the topology of singularities, see also Chapter 13 in this volume.

came from renewed attempts to decide the Poincaré conjecture (here always taken to refer to the three-dimensional case). In 1957, Christos D. Papakyriakopoulos, a mathematician supported by Fox although he had apparently been unproductive for several years, published proofs of Dehn's lemma and two other fundamental theorems, the loop theorem and the sphere theorem. These theorems provided new tools to draw geometric information on 3-manifolds from knowledge of their algebraic topology, and specialists in the field agree that they marked "the beginning of the modern period of growth in 3-dimensional topology" [114, p. xi]. In particular, substantial progress toward a resolution of Poincaré's long-standing problem seemed possible. In the late 1950's, rumours spread in Princeton that several independent proofs were on the way (Bing [14, p. 124]).

In 1958, a paper of Bing on "Necessary and sufficient conditions that a manifold be S^3" brought knots back into the discussion on Poincaré's conjecture. As a matter of fact, Bing, who came like Moise from R.L. Moore's school of general topology, tended to *disbelieve* the conjecture: "The conjecture has not been proved, and I suspect that perhaps being simply connected is not enough to insure that [a closed, orientable 3-manifold] M is topologically S^3".[72] Bing introduced his paper by describing an example of Whitehead of an open, bounded, simply connected subset U of $\mathbb{R}^3$ with connected boundary that nevertheless was topologically different from $\mathbb{R}^3$. In fact, Bing noticed that this open set failed to satisfy a topological property of ordinary 3-space that Artin and Fox had described in their paper on wild cells and arcs, for U contained a simple closed curve that could not be enclosed in a "topological cube", i.e. a 3-ball. Bing's main theorem then asserted that this property – that every simple closed curve can be enclosed in a "topological cube" – was indeed necessary and sufficient to conclude that a closed, connected 3-manifold was homeomorphic to S^3. Knots came into play both in the form of an ingenious trick in Bing's proof of this theorem (see [12, § 5]) as well as in his concluding discussion of various constructions that could perhaps produce counterexamples to the Poincaré conjecture. After discussing handlebody decompositions of 3-manifolds,[73] Bing considered 3-manifolds that could be decomposed into a solid torus and the complement of a tubular neighbourhood of the trefoil knot ("a cube with a knotted hole"). Such manifolds could be thought of as formed by removing a knotted solid torus J from S^3 and "sewing it back" in a different fashion by identifying the boundary torus of both components in various ways. The possible identifications were determined by the image of a meridian of the solid torus J on the boundary torus of $S^3 - J$. Indeed, the resulting manifolds were just Dehn's $\Phi_K(l, m)$ with a 3-sphere filled in to close the manifold. In contrast to Dehn, Bing now considered, for K the trefoil knot, *all* possibilities for the attaching curve and not just those with $m = 1$. (In the following, I will denote the *closed* manifold by $\Phi_K(l, m)$, too, and the construction will be referred to as "Dehn surgery on K".) A presentation of the fundamental group of the resulting manifold could easily be found by adding the relation that expressed the contractibility of the attaching curve to the relations defining the knot group. By analyzing these presentations, Bing showed that $\Phi_K(l, m)$ was simply connected if and only if $m = \pm 1$ and $l = 0$. Moreover, in these cases the manifold was homeomorphic to S^3. Thus, from Dehn surgery on the trefoil knot, no counterexample to the Poincaré conjecture could be formed.

Bing closed his paper with a series of questions. Papakyriakopoulos had informed him, he reported, that in the above construction, the trefoil knot could be replaced by an *arbi-*

---

72 [12, p. 18]. On Moore's school, see [173].

73 In particular, Bing pointed out that no manifold with a decomposition into handlebodies of genus one could lead to a counterexample to the Poincaré conjecture.

trary knot K with the result that any simply connected $\Phi_K(l, m)$ would still be homeomorphic to S^3. Was the same true, Bing asked, for manifolds from which two or more knotted and perhaps linked solid tori were removed and replaced differently? Moreover, did every simply connected compact 3-manifold belong to this class? If the answer to both questions were yes, the Poincaré conjecture would have been proved. If, on the other hand, the answer to either question were no, a counterexample might eventually be constructed.[74]

It turned out that the first question was difficult to answer. In fact, even Papakyriakopoulos' claim was only a conjecture as Bing pointed out in a correction to his paper [13]. The second question, however, was quickly answered in the affirmative. Using a general machinery of "modifications" of differentiable 4-manifolds, A.H. Wallace showed that every differentiable, closed and orientable 3-manifold could indeed be obtained by a finite number of Dehn surgeries on a link of disjoint solid tori [169]. Soon afterward, W.B.R. Lickorish gave an elementary and very geometric proof that the same could be shown in the piecewise linear category [92]. Lickorish's basic idea was to decompose a given oriented 3-manifold into two handlebodies and then to use a sequence of Dehn surgeries to simplify the boundary identification of these handlebodies until a 3-sphere was obtained.[75] In this way, a new technique for constructing and handling closed orientable 3-manifolds was established. The necessary data (what came to be called a "surgery description" of the manifold) were a link and, associated with each of its components, a rational number $r = m/l$ specifying the type of the surgery on a small tubular neighbourhood of this component.[76] It was quickly realized that Dehn surgery could be used to calculate invariants of 3-manifolds by controlling the effect of the surgery operations on the invariants in question. In particular, it became clear that Dehn surgery gave a powerful method for calculating homological knot invariants like the first homology group of the infinite cyclic covering of a knot complement, from which the Alexander polynomial could be derived. This method was heavily exploited in Rolfsen's textbook [140]. In 1978, Robion Kirby was even able to describe an equivalence relation on surgery descriptions, generated by two simple "diagram moves", which corresponded to orientation-preserving homeomorphism between the 3-manifolds thus defined.[77]

The result of Wallace and Lickorish also heightened the interest in Bing's other question: for which knots besides the unknot and the trefoil knot could one show that no Dehn surgery would ever produce a counterexample to the Poincaré conjecture? In 1971, Bing and Martin summarized the results obtained thus far. If the following two propositions about a given knot K were true, the knot was said to have "property P": (1) if Dehn surgery

[74] Soon afterwards, Fox reminded the community of low-dimensional topologists that there was, besides handlebody decompositions and surgery on links, a third way of constructing simply connected 3-manifolds, namely that indicated by Tietze and Alexander, using coverings of the sphere branched over a suitable link. Fox's free calculus allowed to give algebraic conditions on the sheet permutations of the covering that implied its simple connectivity [48].

[75] In order to show that this idea worked, Lickorish established a basic theorem on self-homeomorphisms of closed, orientable surfaces: every such homeomorphism is isotopic to a sequence of elementary self-homeomorphisms called "Dehn twists". Moreover, Lickorish showed that Dehn twists in the splitting surface of a given 3-manifold can be produced by Dehn surgeries.

[76] It is not hard to see that $|l|$ and $|m|$ have to be relatively prime since the corresponding curve must be simple. Moreover, only the quotient of the signs is relevant to fix the relative orientation of the two tori involved in the surgery. The "rational" notation seems to be due to Rolfsen [140].

[77] See [86]. More information on the developments initiated by Dehn surgery on 3-manifolds may be found in the article by Cameron McA. Gordon in this volume.

on K leads to a simply connected manifold Φ, then Φ is homeomorphic to the 3-sphere; (2) any piecewise linear homeomorphism of $S^3 - J$, where J is a small tubular neighbourhood of K, into S^3 can be extended to a piecewise linear self-homeomorphism of S^3. The first condition meant that no counterexample to the Poincaré conjecture could be obtained by Dehn surgery on K, while the second meant that the homeomorphism type of the complement of K determined the knot (up to orientations). Due to Alexander's theorem on embedded tori in S^3, property P could be reformulated as follows: a knot K had property P if and only if, for all nontrivial surgeries (i.e. for $l \neq 0$), the manifold $\Phi_K(l, m)$ was not simply connected. Since a presentation of $\pi_1(\Phi_K(l, m))$ could be found, this reduced the question to combinatorial group theory. By rather tricky constructions of homomorphisms onto known non-trivial groups, Bing and Martin showed that several classes of knots had property P, including twist knots, doubled knots, and all product knots (in this last case, a more geometric argument was used).[78] At the end of their article, Bing and Martin pointed out that in many cases, like that of the trefoil knot, a properly *geometric* understanding had not yet been reached for the fact that no non-trivial Dehn surgery yielded a simply connected manifold. Later work by various authors changed this to some extent. For instance, David Gabai showed by an argument involving foliations of 3-manifolds that all torus and satellite knots possessed property P [56].[79] It remains unclear, however, whether or not *all* knots share the property.

After a long series of partial results obtained by various authors, and relying on certain techniques of Gabai for studying foliations, Cameron McA. Gordon and John Luecke finally showed that no non-trivial Dehn surgery on a knot yields S^3 [62]. While this did not resolve the problem of property P, it answered Tietze's long-standing question: the topological type of a knot complement does indeed determine the type of a knot (with or without orientations).[80] Therefore, Gordon's and Luecke's result implied that the second clause in Bing's and Martin's original definition of property P could be dropped, so that the truth of the Poincaré conjecture would imply that all knots have property P. On the other hand, if a single knot could be found such that some Dehn surgery on it yielded a simply-connected manifold, a counterexample to the Poincaré conjecture would have been found, too. Thus, property P is still considered by several mathematicians as one of the major open problems of knot theory.

§ 30. By the end of the seventies, a further development in 3-manifold theory came to a certain end which had fundamental implications for knot theory: the general classification problem of knots was recognized to be solvable by algorithmic means. Following the first undecidability results in mathematical logic in the 1930's, logicians raised the question as to whether certain topological problems, among them the classification of knots, might

---

[78] In the proof for twist knots, matrix representations of Coxeter groups were used for this purpose. These representations were generated by certain matrices in which complex square roots of the numbers $4\cos^2(\pi/n)$, $n = 3, 4, 5, \ldots$ occurred. After Jones showed that these numbers were just the possible discrete values of the index of subfactors of the hyperfinite II_1 factor (see below, § 32), a relation between subfactors and Coxeter groups was immediately recognized, see, e.g., [77, p. 104]. I am not aware, however, of work relating this to the problems studied by Bing and Martin.

[79] For torus knots, a purely group-theoretical proof of property P had already been included in the textbook of Burde and Zieschang [27, Section 15.6].

[80] If a knot complement $S^3 - K_1$ would be homeomorphic to another, $S^3 - K_2$, without a homeomorphism of pairs $(S^3, K_1) \to (S^3, K_2)$, there would be a non-trivial Dehn surgery on K_1 yielding S^3. A survey of further known properties of Dehn surgeries on knots can be found in [61].

be algorithmically unsolvable as well (Church [30]). The reason for posing such a question was that Reidemeister's purely combinatorial approach had given the knot problem a form very much resembling a kind of word (or transformation) problem in symbolic calculi. When, in the mid-fifties, P.S. Novikov and W.W. Boone independently showed that the general word problem in finitely presented groups was unsolvable, Markov soon thereafter pointed out that this implied the algorithmic unsolvability of the general classification problem of manifolds of dimension greater than three [103]. Due to these circumstances, the case of three dimensions, and knot classification, gained even more interest. Reidemeister, at least, was prepared to wager that the problem of deciding whether or not two given knots were equivalent was solvable (see [11, p. 97]).

Notwithstanding such hopes, undecidability results rather than decidability proofs were high on the agenda of mathematical logicians, and Boone even seems to have tried to prove that the knot problem was unsolvable. Therefore, when Wolfgang Haken, then an almost complete outsider in the community of topologists, announced a theory which allowed to decide algorithmically whether or not a given knot was isotopic to an unknotted circle, he could be sure both of attention and of a certain amount of scepticism about the correctness of his results. Haken, who first presented his ideas at the ICM in 1954, was asked to work out his ideas in full detail. This task took him several years, but in 1961, his long and technically demanding "Theorie der Normalflächen" was finally published in *Acta Mathematica.* In the same year, a somewhat simplified and more intuitive presentation of Haken's ideas was given by Schubert [147]. In his article, Haken described an algorithm which enabled one to construct, for a given compact, triangulated 3-manifold, a finite set of "normal surfaces", characteristic of the manifold's topology. In the case of a knot complement (bounded by a torus along the knot), the algorithm could be adapted to produce a Seifert surface of minimal genus. Thus, in principle, the genus of the knot was computable and in particular, it was decidable whether or not the knot was trivial. However, even today Haken's highly complicated algorithm remains beyond the requirements of practical computation. Haken's result found great appreciation among logicians, however, and a few weeks after his paper appeared, he was offered a position at Urbana, Illinois, where Boone was gathering a research group working on the decidability of mathematical problems.[81]

In 1962, Haken announced that he could modify his algorithm in such a way that it could be used to classify a large number of 3-manifolds, including all knot complements [65]. The basic idea was to employ the algorithm to find so-called "incompressible surfaces" in a given manifold M along which the manifold could be split into pieces.[82] Haken sought to determine a class $\mathfrak{K}$ of compact, orientable 3-manifolds for which the process of finding such splitting surfaces could be iterated, decomposing the manifold after finitely many steps into a collection of 3-balls. Moreover, the class of manifolds should be such that the algorithm allowed, for each of the pieces obtained at a given stage, only finitely

[81] The historical details of this paragraph have been taken from an interview with Haken, conducted by T. Dale in 1994, that has kindly been communicated to me by D. MacKenzie.

[82] The technical definition of an incompressible surface underwent several modifications throughout the following years. Intuitively speaking, an incompressible surface cannot be simplified within M by cutting open handles or by deleting 2-spheres that bound a 3-ball. In [167], the following definition was chosen: An incompressible surface F in a compact, orientable 3-manifold M is either a properly embedded, compact surface (i.e. $F \cap \partial M = \partial F$) or a component of ∂M, such that the following two conditions are satisfied: (1) there does not exist an embedded disk D in the interior of M, bounded by a curve $\partial D \subset F$ which is not contractible in F; (2) no component of F is a 2-sphere bounding a 3-ball in M. See also [66] for a readable description of his procedure.

many possibilities for the next splitting surface (the splitting process thus had the structure of a finite, rooted tree). For two manifolds in such a class $\mathfrak{K}$, it could then be decided if they were homeomorphic by comparing the finitely many splitting trees (called "hierarchies" in the technical literature). The manifolds were topologically equivalent if and only if two of them ran completely parallel. However, in order to apply the algorithm to a given manifold M at all, it had to be known beforehand that 2-sided, incompressible surfaces $F \subset M$ existed (with $\partial F \subset \partial M$, if F was bounded), and, moreover, that M was irreducible, i.e. that every $S^2 \subset M$ bounded a 3-ball (otherwise, the unproved Poincaré conjecture would have prevented recognizing the 3-balls at the end of a splitting hierarchy). Thus it was reasonable to conjecture, and Haken in fact claimed, that $\mathfrak{K}$ could be taken to include all manifolds satisfying these two conditions (following Thurston, such manifolds are usually called "Haken manifolds" today). Among them were all knot complements, so that the algorithm implied a decision procedure for the homeomorphism problem of knot complements.

Unfortunately, Haken did not spell out the proofs of all the claims he made in his paper, so that the scope of his results was not completely clear. An announced sequel to his article, which should have given the missing technical details, never appeared. Indeed, further research by Haken and Friedhelm Waldhausen made clear that Haken's original arguments required either an additional restriction on the class $\mathfrak{K}$ or an algorithmic solution of the conjugacy problem in the group of isotopy classes of self-homeomorphisms of a compact, bounded surface (with respect to isotopies fixing the boundary).[83] It took another decade before a co-worker of Waldhausen, Geoffrey Hemion, solved this additional problem and thus established Haken's original claim as correct [71]. In a widely read survey article, Waldhausen summarized the overall results of the development. While these results were of great importance for 3-manifold theory in general, they had particularly striking consequences for knot theory. By a slight modification, Haken's procedure would not only classify knot complements and knot groups, but actually knots themselves.[84] A further consequence of Waldhausen's own contributions to the subject was the proof of Fox's conjecture that two knots whose groups could be mapped by an isomorphism respecting peripheral subgroups had homeomorphic complements. In view of Gordon's and Luecke's theorem, this implies that the knots are equivalent. The result can be strengthened in the case of prime knots: up to orientation, these knots are determined by their groups.[85]

So far, it seems, Haken's unwieldy algorithm itself has been less useful in further research on knots and 3-manifolds than the general theorems drawn from it by Waldhausen and others. Haken's contribution must thus first and foremost be viewed as a decidability proof. Nevertheless, the results established by working out Haken's ideas changed the outlook on knot theory. The search for simpler classifying algorithms or complete knot invariants was shown to be a meaningful enterprise. What should be emphasized in the present context is the fact that genuine three-dimensional ideas guided this line of research, and

[83] See [167, 66, 168, § 4]. The difficulty arose from the possibility that during the decomposition process, a fibre bundle over S^1 could arise, fibred by incompressible surfaces, with incompressible boundary, and containing only incompressible surfaces isotopic to a fibre or a boundary component. In this case, the decomposition process would be blocked. Since such manifolds were known to be representable as mapping tori of self-homeomorphisms of their fibre, a solution to the above-mentioned problem was required.

[84] See [168, § 4]. In carrying out the splitting procedure, one had to keep track of a meridian of the knot considered.

[85] See [168, p. 26], [172].

highly specific geometric tools, such as the notion of an incompressible surface, were provided by it. Once again, it was not by means of diagram combinatorics that a deep insight was found, and once again, knot theory profited from its status as a specialty within the well-established field of 3-manifold theory.

§ 31. In the mid-seventies, the mathematical community was surprised by a revival of the connection between knots, 3-manifolds, and hyperbolic geometry which had already been touched upon by Dehn. In 1973, a Ph.D. student at Southampton, Robert Riley, found that the complement $S^3 - K_4$ of the figure eight knot K_4 (see Figure 1) had a hyperbolic structure, i.e. it admitted a complete Riemannian metric of constant sectional curvature -1.[86] In fact, Riley showed that $S^3 - K_4$ was homeomorphic to a quotient $\mathbb{H}^3/G$ of three-dimensional hyperbolic space $\mathbb{H}^3$ by a discrete group G of hyperbolic isometries, acting freely on $\mathbb{H}^3$ and isomorphic to $\pi_1(S^3 - K_4)$ [138]. The proof relied essentially on Waldhausen's theorems on Haken manifolds. Riley then went on to construct, with the help of a computer, similar examples of hyperbolic structures in certain other knot complements. He conjectured that the complements of all knots except torus and satellite knots could be endowed with such a structure.[87]

In 1977, Riley met William Thurston, who was then in the course of working out his general programme of finding geometric structures on 3-manifolds, an outline of which began to circulate in the form of notes of Thurston's lectures at Princeton about a year later [160]. Riley's results inspired Thurston to look systematically for hyperbolic structures in knot complements and related 3-manifolds. Among many other things, Thurston pointed out in his lecture notes that Riley's example was closely related to a hyperbolic manifold that Hugo Gieseking, a student of Max Dehn, had discussed in 1912. In his dissertation, Gieseking had constructed a manifold whose fundamental group contained an isomorphic copy of the group of the figure eight knot as a subgroup of index two.[88] This manifold was constructed from a regular tetrahedron in three-dimensional hyperbolic space, all of whose vertices were on the sphere at infinity. By identifying the sides of this tetrahedron two by two, Gieseking had obtained a non-compact manifold with a complete hyperbolic metric, and with finite hyperbolic volume. Thurston now showed that the natural conjecture, suggested by the structure of the fundamental group of Gieseking's manifold, was indeed true: $S^3 - K_4$ was the twofold orientable covering of Gieseking's example. In particular, $S^3 - K_4$ could be decomposed into two copies of the hyperbolic tetrahedron defining Gieseking's manifold.

Thurston went on to prove a general result on the existence of hyperbolic structures on certain compact, bounded 3-manifolds which implied that a knot complement $S^3 - K$ (or, equivalently, the interior of the compact, bounded manifold obtained by removing an open tubular neighbourhood of K from S^3) admitted a such a structure if and only if K was not a torus knot or a satellite knot, as Riley had conjectured. If K was a torus knot, then its complement could be given a different geometrical structure, while if K was a satellite of a non-trivial knot K', the question of endowing the knot complement with a geometric structure could be asked separately for the two (simpler) pieces obtained by splitting $S^3 - K$ along a torus, bounding a tubular neighbourhood of K' and containing K.

86 A metric is called complete if every geodesic may be extended indefinitely.

87 See [139, 161, pp. 360, 366ff].

88 See [59]. A description of Gieseking's example was also given in [99, pp. 153ff]. For the delicate question how much Gieseking or Dehn knew about the relation with K_4, see [101, pp. 39f.].

Thus, knot complements provided a striking illustration of Thurston's main conjecture that "the interior of every compact 3-manifold has a canonical decomposition into pieces which have a geometric structure" [161, Conjecture 1.1]. A proof of this conjecture for the class of compact Haken manifolds, together with several surprising applications, earned Thurston a Fields medal in 1982, despite the fact that full details of the proofs had not yet appeared in print.

A direct application of Thurston's results to knot theory was made possible by another fundamental result on hyperbolic 3-manifolds that had been proved in the early 1970's, the "rigidity theorem" of hyperbolic manifolds.[89] It stated that if two 3-manifolds M and N of dimension $\geqslant 3$ with a hyperbolic structure of finite volume have isomorphic fundamental groups, then M and N are not only homeomorphic but even isometric to each other. This implied that every *isometric* invariant of such a manifold, for instance the volume, was necessarily also a *topological* invariant. Since all hyperbolic knot complements had finite volume, the rigidity theorem provided a way to introduce a whole basket of new invariants for knots with hyperbolic complements. Many of these new invariants turned out to be calculable by means of computers. Jeffrey Weeks, a student of Thurston, was particularly successful in this respect. In his Ph.D. thesis, he described an algorithm for calculating various hyperbolic knot invariants that has since been very useful in extending knot tables to ever higher crossing numbers.[90] Already the volume of a knot turned out to be a rather fine (though not complete) invariant of knots with hyperbolic complements. It seems to measure a kind of geometric complexity of knots, but not much is known about this as yet. Thurston has conjectured that the complement of the figure eight knot K_4 might be the hyperbolic manifold with the least volume [161, p. 365].

Evidently, the rigidity theorem and the work of Riley and Thurston not only related knot theory to 3-dimensional topology in a deeper way but also led to a variety of more specifically geometric issues. For instance, representations of knot groups by discrete subgroups of $PSL(2, \mathbb{C})$ can be investigated, or the details of the hyperbolic structure of knot complements may be looked at. Here, too, Dehn surgery turns out to be a particularly helpful tool. It allows to construct new hyperbolic manifolds from given ones, and to address questions such as: which Dehn surgeries on a given knot do produce hyperbolic manifolds and which do not? The connection between knot theory and hyperbolic geometry has opened up a rich and still rather unsurveyable field of inquiry.

§ 32. Up to this point, the "geometry" involved in the investigation of knots and links was mainly that of three-dimensional manifolds associated with knots, be it in the sense of their topological structure or, as in the last paragraph, in the more specific sense of a Riemannian metric on the knot complement. In Vaughan Jones's discovery of a new knot polynomial, a completely different kind of geometry came into play: that of lattices of projections on a Hilbert space and the algebras generated by them. In order to explain why this may with reason be called a variety of *geometry*, a short look back to the beginnings of the field in which Jones was working is necessary. In the 1930's, John von Neumann and his collaborator Francis J. Murray embarked on a programme investigating what they called "rings of operators" on a separable Hilbert space (today called von Neumann algebras). In the course of this work, they invented a mathematical object that represented a close

[89] See [119] for a proof in the case of compact manifolds and [130] for the non-compact case.

[90] See [1]. Recently, tables of prime knots of up to 16 crossings have been constructed by Thistlethwaite, Hoste and Weeks, using a modification of Weeks's program.

infinite-dimensional analogy to complex projective space, although in an important respect it had very different properties. Since it was an investigation of the fine structure of this object which led to Jones's breakthrough, a more detailed account is necessary.[91]

Von Neumann's and Murray's work was motivated by earlier research on the spectral theory of linear operators and the wish to understand the mathematical foundations of quantum mechanics. It concentrated on so-called "factors" of operator rings, i.e. rings $\mathcal{M} \subseteq B(\mathcal{H})$ of bounded operators acting on a Hilbert space $\mathcal{H}$, closed under the adjoining operation * and under pointwise convergence on $\mathcal{H}$, containing the identity operator $\mathbf{I} \in B(\mathcal{H})$, and with a trivial center. Obvious examples of factors were the rings of *all* operators on a (separable) Hilbert space. Up to algebraic isomorphism, these factors were classified by the dimension of the underlying Hilbert space, i.e. they were all isomorphic to the rings $M_n(\mathbb{C})$ of all $(n \times n)$-matrices over the complex numbers or to the set $B(\mathcal{H})$ of all bounded linear operators on a separable, infinite-dimensional Hilbert space $\mathcal{H}$. However, the theory acquired depth by the fact that more, and different, examples could be constructed. In particular, Murray and von Neumann described a class of factors which in their view represented in many ways a better analogy to the finite-dimensional factors $M_n(\mathbb{C})$ than $B(\mathcal{H})$. Using a simplified construction method that Murray and von Neumann published in 1943, these factors can be defined as follows. For a finite or countably infinite group G, the Hilbert space $l^2(G)$, consisting of all square-summable sequences of complex numbers indexed by the elements of G may be formed. On this Hilbert space, G acts by its left regular representation U, given by

$$(U_g\xi)_h := \xi_{gh} \quad \text{for all } \xi \in l^2(G);\ g, h \in G.$$

Then the smallest closed subring $\mathcal{M} \subseteq B(l^2(G))$ containing all operators U_g is a von Neumann algebra, whose elements can all be represented in the form $\sum_{g \in G} \eta_g U_g$ for certain $\eta \in l^2(G)$. Murray and von Neumann pointed out that for finite G, the ring $\mathcal{M}$ was equivalent to Frobenius' "group numbers" (in today's language, the group ring $\mathbb{C}G$). In contrast, it was not too difficult to show that for countably infinite groups, $\mathcal{M}$ was a factor if and only if all conjugacy classes of G were infinite [122, § 5.3].

The rings constructed in this way had a very particular property, though. The function $\mathrm{tr}(\sum_{g \in G} \eta_g U_g) := \eta_e$, where e was the neutral element in G, defined a *finite trace*, i.e. a linear function $\mathrm{tr}: \mathcal{M} \to \mathbb{C}$, satisfying $\mathrm{tr}(\mathbf{I}) = 1$, $\mathrm{tr}(x^*x) \geqslant 0$, and $\mathrm{tr}(xy) = \mathrm{tr}(yx)$, for all $x, y \in \mathcal{M}$. This, in turn, made it possible to define a *dimension function* (relative to $\mathcal{M}$) on the lattice of all closed linear subspaces $E \subseteq l^2(G)$ of the form $E = p_E(l^2(G))$ for some orthogonal projection $p_E \in \mathcal{M}$, by putting

$$\dim_{\mathcal{M}}(E) := \mathrm{tr}(p_E).$$

For finite groups of order n, this function measured the dimension of a subspace of $\mathbb{C}G$, normalized in the sense that for a subspace E of dimension k, $\dim_{\mathcal{M}}(E) = k/n$. For the factors constructed from groups with infinite conjugacy classes, however, the range of this dimension function was the closed interval $[0, 1]$. In general, Murray and von Neumann showed that for all factors a similar dimension function could be constructed, with but few possibilities for the range of its values [121, § 8.4]. Corresponding to these possibilities,

[91] Based mainly on [121, 122]. Again, notations have been slightly modernized.

factors with a dimension function like the above (or, equivalently, infinite-dimensional factors with a finite trace) were called factors of type II_1.

Von Neumann realized that this construction came very close to a view of projective geometry that had been advocated by Karl Menger and Garrett Birkhoff a few years earlier. In 1928, Menger, like Reidemeister a mathematician with strong ties to the Vienna circle in philosophy, had proposed to reformulate projective geometry as the theory of linear subspaces of a finite-dimensional vector space [109]. Through the use of homogeneous coordinates, this idea had been implicit in many researches of 19th-century analytic geometers such as Felix Klein, but it was only under the influence of Hilbert's axiomatics that Menger proposed to shift the perspective on projective geometry and to make the properties of linear subspaces of a vector space the basis of the theory. Accordingly, he characterized these by a suitable system of axioms. Besides axioms governing the intersection and linear span of two subspaces, a crucial ingredient of Menger's approach was an axiom asserting the existence of a *dimension function*, associating with each subspace a positive integer that behaved correctly under intersection and linear span of subspaces. The value of this function then specified whether a given subspace corresponded to a point, or to a plane, etc. Seven years later, and independently of Menger's work, Garrett Birkhoff pointed out that the system of subspaces of a finite-dimensional vector space defining "a projective geometry" represented a particular kind of what he had come to call a *lattice* [16].[92] The lattice of projections of a factor of type II_1 satisfied virtually all of Menger's or Birkhoff's axioms except those securing that the structure defined was finite-dimensional. In this perspective, these factors represented a kind of infinite-dimensional complex projective space, or else, a geometry "without points", since no elements of least dimension existed. Von Neumann set out to show that one could indeed characterize the lattices of subspaces arising in the above way in an abstract fashion [123, 124]. For some time, he had great hopes that these "continuous geometries", as he decided to call them, provided the right framework to do infinite-dimensional projective geometry, and even quantum mechanics.[93] Accordingly, he devoted a significant effort to the further investigation of factors of type II_1.

In 1943, Murray and von Neumann were able to show that not all II_1-factors constructed as above were algebraically isomorphic, depending on the properties of the group G used in the construction. If G was the set theoretic union of an ascending sequence of finite groups, then, and only then, the associated factor M was "approximately finite" (or, in today's terminology, hyperfinite), i.e. generated by an ascending sequence of finite-dimensional algebras. Moreover, all such factors were algebraically isomorphic. In other words, up to isomorphism, there was just *one* of them, say, the factor $\mathfrak{R}$ constructed from the group Σ_∞ of permutations of the integers such that each $\sigma \in \Sigma_\infty$ permuted only finitely many integers. Since Σ_∞ was the union of the finite symmetric groups, $\Sigma_1 \subset \Sigma_2 \subset \Sigma_3 \subset \cdots$, the factor $\mathfrak{R}$ was indeed hyperfinite; the group rings $\mathbb{C}\Sigma_n$ could be taken as the approximating sequence of finite-dimensional algebras. If, on the other hand, G was taken to be a free group on two generators, then the associated factor was not hyperfinite [122, § 6.2]. Thus, the hyperfinite II_1 factor $\mathfrak{R}$ had acquired a rather singular position in the theory. It represented, so to speak, the closest infinite-dimensional analogue to the finite-dimensional factors $M_n(\mathbb{C})$; in other words, its lattice of projections represented the closest analogue to the lattice of subspaces of a finite-dimensional, complex vector space. Moreover, its construction showed that it had a rich but complicated inner structure.

92 See [107] for information on the origins of the theory of lattices.

93 See, e.g., the introduction to [121, 125].

For a long period after World War II, the attention of operator algebraists turned to more general issues, and the "continuous geometry" of $\mathfrak{R}$ moved in the background. A crowning achievement of much of this work was Alain Connes's completion of the classification of factors up to algebraic isomorphism, which earned him a Fields medal in 1982. Connes also took up the study of $\mathfrak{R}$ again by classifying its automorphisms.[94] Finally, the time seemed ripe to look at the inner structure of $\mathfrak{R}$ in more detail. It was Vaughan Jones who set himself the task of investigating *subfactors* of $\mathfrak{R}$, i.e. other infinite-dimensional factors $\mathcal{N}$ embedded in $\mathfrak{R}$. As such subfactors were automatically equipped with a finite trace, they also were of type II_1. In such a situation, i.e. given a pair of II_1 factors $\mathcal{N} \subseteq \mathcal{M}$ with the same unit, von Neumann's theory of dimension functions could be used to define an "index" $[\mathcal{M} : \mathcal{N}]$ which equalled the index of groups, $[G : H]$, if the factors $\mathcal{N}$ and $\mathcal{M}$ were constructed from groups $H \subseteq G$ as above. The suprising result found by Jones was that for II_1 subfactors of $\mathfrak{R}$, the possible values of this index did not consist of the interval $[1, \infty)$, as the definition of the index would have allowed, but only of the continuous interval $[4, \infty)$ and the discrete set $\{4\cos^2 \pi/n \mid n = 3, 4, 5, \ldots\}$ [76].

In the proof of his result, Jones calculated the index in a different way. If a pair of II_1 factors $\mathcal{N} \subseteq \mathcal{M}$ with the same unit was given, the inner product on $\mathcal{M}$ given by $(x, y) \mapsto \mathrm{tr}(y^*x)$ allowed for a completion of $\mathcal{M}$ to a Hilbert space, denoted by $L^2(\mathcal{M}, \mathrm{tr})$. On this Hilbert space, $\mathcal{M}$ acted by the left regular representation, given by left multiplication on the dense subspace $\mathcal{M}$. Similarly, $L^2(\mathcal{N}, \mathrm{tr})$ could be formed as a closed linear subspace of $L^2(\mathcal{M}, \mathrm{tr})$. Introducing the projection $e_{\mathcal{N}} : L^2(\mathcal{M}, \mathrm{tr}) \to L^2(\mathcal{N}, \mathrm{tr})$, Jones considered the von Neumann algebra $\mathcal{M}_1 \subset B(L^2(\mathcal{M}, \mathrm{tr}))$, generated by $\mathcal{M}$ and $e_{\mathcal{N}}$. It turned out that $\mathcal{M}_1$ was again a II_1 factor, with a trace extending the trace on $\mathcal{M}$, and such that $[\mathcal{M}_1 : \mathcal{M}] = [\mathcal{M} : \mathcal{N}] = \beta$, where $\beta^{-1} = \mathrm{tr}(e_{\mathcal{N}})$. By iterating this construction (which had already been studied by C. Skau and E. Christensen in the late 1970's), Jones was able to find the possible values of the index. Repeating the process by which $\mathcal{M}_1$ had been formed, Jones obtained both an infinite tower of II_1-factors $\mathcal{M}_i$ $(i = 1, 2, \ldots)$ and an infinite sequence of orthogonal projections $e_{i+1} : L^2(\mathcal{M}_i, \mathrm{tr}) \to L^2(\mathcal{M}_{i-1}, \mathrm{tr})$ (here, $\mathcal{M}_0 := \mathcal{M}$ and $e_1 := e_{\mathcal{N}}$). These orthogonal projections satisfied a remarkable set of relations:

$$e_i e_{i\pm 1} e_i = \beta^{-1} e_i, \qquad e_i e_j = e_j e_i \quad \text{for } |i - j| \geqslant 2;$$

moreover, for all words w in $\mathbf{I}, e_1, \ldots, e_{i-1}$, the relation

$$\mathrm{tr}(w e_i) = \beta^{-1} \mathrm{tr}(w) \qquad (\star\star\star)$$

held, and β was restricted to the set of values mentioned above. Thus, a necessary condition on the values of the index $[\mathcal{M} : \mathcal{N}]$ had been found. But more than that: Jones showed that whenever $\mathcal{P}$ was the von Neumann algebra generated by a system of orthogonal projections satisfying the above relations, then $\mathcal{P}$ was isomorphic to $\mathfrak{R}$ – it was approximated by the ascending sequence of the canonical images $\overline{\mathcal{A}}_{\beta,n} \subseteq \mathcal{P}$ of the abstract finite-dimensional algebras $\mathcal{A}_{\beta,n}$, generated by $\mathbf{I}, e_1, \ldots, e_n$ and satisfying the above relations –, and the double commutant of the set $\{e_2, e_3, \ldots\}$ in $\mathcal{P} = \mathfrak{R}$ was a II_1 subfactor with index β. Consequently, the condition also was sufficient.

[94] See [8] for a brief description of Connes's work.

The system of projections arising in this argument and the finite-dimensional algebras $\mathcal{A}_{\beta,n}$ soon turned out to form the core of a web of surprising relations to other mathematical topics. In fact, these algebras had been encountered by several other people in quite different fields. For instance, H. Temperley and Elliott H. Lieb had used a representation of $\mathcal{A}_{\beta,n}$ on $\mathbb{C}^{2n+2}$ in a study of certain models of statistical mechanics already in 1971. Moreover, for the discrete values of β, the algebras $\mathcal{A}_{\beta,n}$ were in some way or other related to Coxeter groups.[95] Finally, and most important, it was pointed out to Jones by D. Hatt and Pierre de la Harpe that the relations bore a strong resemblance to those defining the braid groups B_n. In fact, it was not difficult to show that after a change of variables,

$$g_i := te_i - (1 - e_i), \quad 2 + t + t^{-1} = \beta,$$

the algebras $\mathcal{A}_{\beta,n}$ were presented by the relations

$$g_i^2 = (t-1)g_i + t,$$

$$g_i g_{i+1} g_i = g_{i+1} g_i g_{i+1},$$

$$g_i g_j = g_j g_i \quad \text{for } |i-j| \geqslant 2,$$

$$g_i g_{i+1} g_i + g_i g_{i+1} + g_{i+1} g_i + g_i + g_{i+1} + 1 = 0.$$

Consequently, the mapping ρ sending a braid generator $\sigma_i \in B_n$ to the element $\rho(\sigma_i) := g_i \in \mathcal{A}_{\beta,n}$ defined a representation of the braid group B_{n+1} within $\mathcal{A}_{\beta,n}$ or, similarly, within $\overline{\mathcal{A}}_{\beta,n} \subseteq \mathfrak{R}$. In this way, a connection to topology was opened up which no one had expected. "For the first time", Jones remarked in a contribution to a conference in July 1983, "II_1 factors have begun to exhibit their geometric and combinatorial nature. This rich structure can only be expected to deepen as one answers further simple questions about subfactors of finite index" [78, p. 270]. It was not immediately clear, though, how to exploit this connection, as the same paper shows. For a short period, Jones hoped that the determinant of his family of representations of B_n could be used in a way similar to that in which the Burau representation had been used to get new information about the Alexander polynomial or perhaps a related invariant of links.[96] To discuss this question, Jones turned to an expert in braid groups, Joan Birman. In her earlier book [17], written shortly after Garside had solved the conjugacy problem for the braid group, Birman had collected and refined the available knowledge for the study of knots and links via an analysis of the relation between links and closed braids. She was therefore a natural partner for discussing Jones's new ideas. In the discussions, however, Birman pointed out to Jones that his first idea would not work out.[97] In a popular article, Jones later recalled:

> I went home somewhat depressed after a long day of discussions with Birman. It did not seem that my ideas were at all relevant to the Alexander polynomial or to anything else in knot theory. But one night the following week I found myself sitting up in bed and running off to do a few calculations. Success came with a much simpler approach than the one I had been trying. I realized I had generated a polynomial invariant of knots.[98]

[95] See footnote 78 above.

[96] See above, § 24, and [78, p. 244].

[97] See note (2), added in proof, in [78, pp. 244 and 273].

[98] From [81].

It is not difficult to tell what Jones had found. Among other things, Birman had explained to Jones Markov's equivalence relation, characterizing the braid words that represented isotopic knots or links as a closed braid (see above, § 24). Jones realized that the traces $\mathrm{tr}: \mathcal{A}_{\beta,n} \to \mathbb{C}$ furnished by his theory of subfactors automatically satisfied

$$\mathrm{tr}\big(vwv^{-1}\big) = \mathrm{tr}(w) \quad \text{and} \quad \mathrm{tr}\big(wg_{i+1}^{\pm 1}\big) = \beta^{\mp 1}\mathrm{tr}(w),$$

for all words v, w in the generators $g_1, g_2, \ldots, g_i$, since two arguments of the trace could be interchanged and since it satisfied property ($\star\star\star$). Thus, only a slight correction was needed in order to make the trace itself into an invariant of links. Indeed, Jones showed that if $w \in B_n$ was a word with exponent sum e in the braid group generators σ_i, representing an oriented link L, then

$$V_L(t) := \left(-\frac{t+1}{\sqrt{t}}\right)^{n-1} t^{e/2}\,\mathrm{tr}\big(\rho(w)\big)$$

was invariant under the moves generating Markov's equivalence relation and thus $V_L(t)$ was an isotopy invariant of oriented links. Moreover, the finite-dimensional algebra involved showed that V_L was a Laurent polynomial in the variable t for knots and links with an odd number of components, while it was a Laurent polynomial in $\sqrt{t}$ for links with an even number of components. Further discussions with Birman brought the next surprise. Examples showed that the new invariant was *not* equivalent to the Alexander polynomial. However, Birman and Jones found that V_L satisfied a skein relation similar to the Alexander polynomial, as described in § 11 above.[99] This meant that Jones's polynomial could be defined independently of its original context in von Neumann algebras, a fact heavily exploited in subsequent work.

Let me reconsider the remarkable chain of arguments leading from von Neumann's construction of the hyperfinite II_1 factor $\mathfrak{R}$ to Jones's new link invariant. While completely independent of low-dimensional topology, the beginnings of this development were clearly motivated by the wish to understand a particular kind of infinite-dimensional geometry, extending the approach to projective geometry by Menger and Birkhoff. Moreover, these beginnings were related to von Neumann's attempt to clarify the mathematical basis of quantum mechanics. Jones took up the problem of subfactors of $\mathfrak{R}$, continuing this investigation along lines close to those indicated by Murray's and von Neumann's work. When Jones found his towers of finite-dimensional algebras (the canonical images of $\mathcal{A}_{\beta,n}$ inside $\mathfrak{R}$), he was inclined to think of them in terms of von Neumann's variety of geometry: "The situation is thus very geometric and [the] relations [defining $\mathcal{A}_{\beta,n}$] can be thought of as defining special configurations of subspaces" [79, p. 377]. From this point of view, however, the outcome of Jones's research generated perhaps even more riddles than it solved. By arguments which in the end boiled down to exploiting a surprising similarity in the combinatorial structures of $\mathfrak{R}$ and the braid groups, a relation between the geometry of configurations of subspaces of a Hilbert space and the topology of low-dimensional objects such as braids and links was established. But what was – apart from this combinatorial resemblance – the geometric reason for this connection? Was there a kind of structure

[99] Interview with Joan Birman, Oberwolfach 1995. For this interview and further private communications about her involvement in the invention of the new polynomial invariants, I wish to express my sincere thanks to Joan Birman.

which bridged the algebra and the topology in question, in a similar way than the homology of cyclic coverings related knots and links to the Alexander polynomial? In the years following Jones's breakthrough, such questions were asked repeatedly. In a contribution on statistical mechanical models of link invariants, published in 1989, Jones himself conceded that the riddle was still unsolved: "Our main reason for doing this work was as a step towards a useful and genuinely three-dimensional understanding of the invariants. So far we have not succeeded. The situation is the same as that of the poor prisoners in Plato's allegory of the cave" [80, p. 312].[100]

While Jones's new invariant has been used and generalized by many people in a broad spectrum of directions (such as: further polynomial link invariants, statistical mechanical models, quantum field theory, invariants of 3-manifolds constructed on the basis of Kirby's calculus of surgery descriptions), it seems that a deeper understanding of the relations between the two kinds of geometry involved is still lacking. However, an important new idea which might eventually change the situation came into play through work of V.A. Vassiliev [164, 165]. Following a general approach outlined by V.I. Arnold, Vassiliev proposed to study the space $\mathcal{V}$ of all smooth mappings $S^1 \to S^3$. In this space, the isotopy classes of knots are separated by a system Σ of "walls" representing *singular* maps, and thus the homology of $\mathcal{V} - \Sigma$ in dimension zero, which can be studied by means of a spectral sequence, characterizes all numerical knot invariants. After Birman and X.-S. Lin found a connection between Jones's and Vassiliev's ideas in fall 1990, a substantial amount of research was done on this connection which might provide the starting point for a better understanding of the topology underlying the new link polynomials.[101]

6. Conclusion

§ 33. From the account given in the previous three sections it will be clear that a "tale of diagram combinatorics" such as that told in Section 2 reduces the complex weave of scientific and mathematical practice in which knot theory was formed to a rather thin narrative, in which the intentional and causal aspects of the development become almost unrecognizable. This can already be seen from the periodization which is suggested by the developments discussed. Four major stages of the history of knot theory can be discerned. In the first stage, extending from Vandermonde's first remarks to the late 19th-century tabulations, the mathematization of the knot problem stood in the foreground. This mathematization was called for by various developments in the exact sciences, ranging from astronomy and the theory of electromagnetism to Thomson's speculations on the structure of matter. In the second period, from 1900 to the late 1930's, modern knot theory emerged as a subfield of the discipline of topology, culminating in Alexander's, Reidemeister's and Seifert's contributions. On the one hand, we have seen that this emergence of modern knot theory was motivated by the desire to understand singularities of algebraic curves and surfaces – a topic deeply rooted in 19th-century pure mathematics – and to solve several major problems thrown up by Poincaré's new *Analysis situs*. On the other hand, the formation of knot theory was influenced by the modernist impulse toward autonomous, formal theories, an impulse which found its clearest expression within the developments considered here in Reidemeister's *Knotentheorie* of 1932. The third period, extending roughly from 1945 to

[100] For statements in a similar spirit, see [18].

[101] See the survey of this development in [18].

Jones's invention of a new knot polynomial, is characterized by the close interplay between knot theory and the growing field of low-dimensional topology. The various ways in which knots gave rise to 3-manifolds were explored in detail, and the surprising resistance of the three-dimensional Poincaré conjecture only contributed to motivate topologists to clarify the structure of knot complements, manifolds obtained from those by Dehn surgery, and 3-manifolds in general. The fourth period set on with Jones's discovery, a breakthrough which remains surprising even today, and which changed the structure of the field very deeply. Knot theory is no longer more or less exclusively tied to low-dimensional topology, but also to a variety of other fields among which mathematical physics certainly stands out.

A look at the intellectual contexts which I have touched upon (restricting mainly to the mathematical ideas involved) allows us to recognize that the actual motivations for mathematical investigations of knots and links were very complex. In a more or less direct way, and like so many other fields of mathematics, research on knots was related to the small sets of highly appreciated and contested research themes, the "big issues" that occupied the attention of the scientists of a given period. What is the nature of the small planets, and what are their orbits? What is an atom, and how are observed spectra to be explained? How do algebraic curves or surfaces behave at singular points? What are the objects of the new science of topology and how can ordinary space be characterized in purely topological terms?[102] What is the right mathematical framework to be used in quantum physics? Which mathematical problems are solvable by algorithmic means? The appreciation of such problems, and even more of the candidates for their solutions, has continually changed and will often be found not to coincide with today's valuations. William Thomson's theory of vortex atoms which inspired Tait's tabulation enterprise did not sustain its original attraction for long. Nevertheless, it was in relation to such larger themes that the knot problem has continued to occupy the attention of mathematical minds. In their day, and for a shorter or longer period, they represented hard and deep problems in rich intellectual constellations; constellations which reached far beyond the narrow focus of a particular piece of knot-theoretical work. Moreover, a study of the temporal modifications of the interplay between the grand scientific themes and more concrete research allows us to gain a deeper insight into the historical changes influencing the development of mathematics. It is significant that after 1900 the interest in knots no longer arose from physics but from pure mathematics, and that in the wake of Jones's work, mathematical physics again came to play a major role in motivating research on knots and related topics. I have also indicated in which way the move toward a combinatorial style of "modern" knot theory (or "classical", depending on the perspective) was at least partially inspired by philosophical debates on the foundations of mathematics. To spell out all these influences and interrelations in the details of mathematical, scientific, and cultural practice would mean to produce yet another, and still much "thicker" historical narrative on the formation of knot theory than the one I have presented here.[103]

Returning to the proper subject of this contribution, let me close by recalling the truism that it is not the historian's task to predict the future. However, it is less a prediction than a reasonable expectation to suppose that geometric aspects will continue to play a crucial role in the further development of knot theory. After all, the hierarchy of knots in ordinary

[102] A more detailed analysis would show that around 1900, the second half of this theme was not without cosmological overtones.

[103] For the interesting notion of "thick narratives", see [58].

space or of similar placements in manifolds will continue to remain, first and foremost, a hierarchy of *geometrical complexity* of a certain kind. This hierarchy remains only very partially understood. The bare fact that it is possible in principle to enumerate all types of knots and links in that hierarchy does not tell too much about its finer structure (the comparison has been made with a listing of all prime numbers and a deeper understanding of number theory).[104] Thus, knot theory will continue to be interesting and useful in all situations within and outside mathematics where this kind of geometrical complexity is involved. If the history of knot theory tells us anything, it is that this has always been the case.

Bibliography

The following abbreviations have been used: Abhandlungen MSHU = Abhandlungen aus dem Mathematischen Seminar der Hamburgischen Universität; AMS = American Mathematical Society; DMV = Deutsche Mathematiker-Vereinigung, LMS = London Mathematical Society; NAS = National Academy of Sciences USA; RSE = Royal Society of Edinburgh.

[1] C. Adams, M. Hildebrand and J. Weeks, *Hyperbolic invariants of knots and links*, Transactions of the AMS **326** (1991), 1–56.
[2] J.W. Alexander, *Note on two three-dimensional manifolds with the same group*, Transactions of the AMS **20** (1919), 339–342.
[3] J.W. Alexander, *Note on Riemann spaces*, Bulletin of the AMS **26** (1920), 370–372.
[4] J.W. Alexander, *A lemma on systems of knotted curves*, Proceedings of the NAS **9** (1923), 93-95.
[5] J.W. Alexander, *Topological invariants of knots and links*, Transactions of the AMS **30** (1928), 275–306.
[6] J.W. Alexander, *Some problems in topology*, Verhandlungen des Internationalen Mathematiker-Kongresses, Zürich 1932, Vol. 1, W. Saxer, ed., Orell Füssli, Zürich/Leipzig, 1932, 249–257.
[7] J.W. Alexander and G.B. Briggs, *On types of knotted curves*, Annals of Mathematics **28** (1927), 562–586.
[8] H. Araki, *The work of Alain Connes*, Proceedings of the International Congress of Mathematicians, Warszawa 1983, Vol. 1, North-Holland, Amsterdam, 1984, 3–9.
[9] E. Artin, *Theorie der Zöpfe*, Abhandlungen MSHU **4** (1925), 47–72.
[10] E. Artin, *Zur Isotopie zweidimensionaler Flächen im* $\mathbb{R}_4$, Abhandlungen MSHU **4** (1925), 174–177.
[11] R. Artzy, *Kurt Reidemeister*, Jahresberichte der DMV **74** (1972), 96–104.
[12] R.H. Bing, *Necessary and sufficient conditions that a 3-manifold be* S^3, Annals of Mathematics **68** (1958), 17–37.
[13] R.H. Bing, *Correction to "Necessary and sufficient conditions that a 3-manifold be* S^3*"*, Annals of Mathematics **77** (1963), 210.
[14] R.H. Bing, *Some aspects of the topology of 3-manifolds related to the Poincaré conjecture*, Lectures on Modern Mathematics, Vol. 2, T.L. Saaty, ed., Wiley, New York, 1964, 93–128.
[15] R.H. Bing and J.M. Martin, *Cubes with knotted holes*, Transactions of the AMS **155** (1971), 217–231.
[16] G. Birkhoff, *Combinatorial relations in projective geometries*, Annals of Mathematics **36** (1935), 743–748.
[17] J.S. Birman, *Braids, Links, and Mapping Class Groups*, Annals of Mathematics Studies vol. 82, Princeton University Press, Princeton, 1974.
[18] J.S. Birman, *New points of view in knot theory*, Bulletin of the AMS New Series **28** (1993), 253–287.
[19] K. Borsuk, *Sur la courbure totale des courbes fermées*, Annales de la Société Polonaise **20** (1947), 251–265.
[20] K. Brauner, *Zur Geometrie der Funktionen zweier komplexer Veränderlicher*, Abhandlungen MSHU **6** (1928), 1–55.
[21] E. Breitenberger, *Gauß und Listing. Topologie und Freundschaft*, Mitteilungen der Gauss-Gesellschaft Göttingen **30** (1993), 2–56.

[104] See, e.g., [159].

[22] E. Brieskorn, *Beispiele zur Differentialtopologie von Singularitäten*, Inventiones Mathematicae **2** (1966), 1–14.
[23] H. Brunn, *Über verknotete Kurven*, Verhandlungen des Ersten Internationalen Mathematikerkongresses, Zürich 1897, F. Rudio, ed., Teubner, Leipzig, 1898, 256–259.
[24] W. Burau, *Über Zopfgruppen und gleichsinnig verdrillte Verkettungen*, Abhandlungen MSHU **11** (1936), 179–186.
[25] W. Burau, *Kennzeichnung der Schlauchknoten*, Abhandlungen MSHU **9** (1933), 125–133.
[26] W. Burau, *Kennzeichnung der Schlauchverkettungen*, Abhandlungen MSHU **10** (1934), 285–297.
[27] G. Burde and H. Zieschang, *Knots*, de Gruyter, Berlin/New York, 1985.
[28] B. Chandler and W. Magnus, *The History of Combinatorial Group Theory*, Springer Verlag, New York, 1982.
[29] N. Chaves and C. Weber, *Plombages de rubans et problème des mots de Gauss*, Expositiones Mathematicae **12** (1994), 53–77.
[30] A. Church, *Comment on Th. Skolem's review of "Zur Reduktion des Entscheidungsproblems" by L. Kalmár*, Journal of Symbolic Logic **3** (1938), 46.
[31] J. Conway, *An enumeration of knots and links and some of their related properties*, Computational Problems in Abstract Algebra, Proceedings of Oxford 1967, J. Leech, ed., Pergamon, New York, 1970, 329–358.
[32] R.H. Crowell and R.H. Fox, *Introduction to Knot Theory*, Ginn, Boston, 1963.
[33] M. Dehn and P. Heegaard, *Art. "Analysis situs"*, Encyklopädie der Mathematischen Wissenschaften, III AB, Teubner, Leipzig, 1907–1910, 153–220; completed January 1907.
[34] M. Dehn and P. Heegaard, *"Uber die Topologie des dreidimensionalen Raumes*, Mathematische Annalen **69** (1910), 137–168.
[35] M. Dehn and P. Heegaard, *Die beiden Kleeblattschlingen*, Mathematische Annalen **75** (1914), 402–413.
[36] M. Dehn and P. Heegaard, *Über kombinatorische Topologie*, Acta Mathematica **67** (1936), 123–168.
[37] J. Dieudonné, *A History of Algebraic and Differential Topology*, Birkhäuser, Basel, 1989.
[38] R. Einhorn, *Vertreter der Mathematik und Geometrie an den Wiener Hochschulen, 1900–1940*, thesis, Technische Universität Wien, 1983.
[39] M. Epple, *Kurt Reidemeister. Kombinatorische Topologie und exaktes Denken*, Die Albertus-Universität Königsberg und ihre Professoren, D. Rauschning and D. van Nerée, Duncker and Humblot, Berlin, 1994, 567–575.
[40] M. Epple, *Branch points of algebraic functions and the beginnings of modern knot theory*, Historia Mathematica **22** (1995), 371–401.
[41] M. Epple, *Orbits of asteroids, a braid, and the first link invariant*, The Mathematical Intelligencer **20/1** (1998), 45–52.
[42] M. Epple, *Topology, matter, and space, I. Topological notions in 19th-century natural philosophy*, Archive for History of Exact Sciences **52** (1998), 297–382.
[43] L. Euler, *Solutio problematis ad geometriam situs pertinentis*, Commentarii Academiae Scientiarum Imperialis Petropolitanae **8** (1736), 128–140; English translation in: N.L. Biggs et al., *Graph Theory, 1736–1936*, Clarendon Press, Oxford, 1976, 3–21.
[44] M.I. Fáry, *Sur la courbure totale d'une courbe gauche faisant un nœud*, Bulletin de la Société Mathématique de France **77** (1949), 128–138.
[45] R.H. Fox, *Recent development of knot theory at Princeton*, Proceedings of the International Congress of Mathematicians 1950, Vol. 2, L.M. Graves et al., eds, AMS, Wiesbaden, 1952, 453–457.
[46] R.H. Fox, *On the complementary domains of a certain pair of inequivalent knots*, Indagationes Mathematicae **14** (1952), 37–40.
[47] R.H. Fox, *A quick trip through knot theory*, Topology of 3-Manifolds and Related Topics, M.K. Fort Jr., ed., Proceedings of the University of Georgia Institute 1961, Prentice-Hall, Englewood Cliffs, NJ, 1962, 120–167.
[48] R.H. Fox, *Construction of simply connected 3-manifolds*, Topology of 3-Manifolds and Related Topics, M.K. Fort Jr., ed., Proceedings of the University of Georgia Institute 1961, Prentice-Hall, Englewood Cliffs, NJ, 1962, 213–216.
[49] R.H. Fox and E. Artin, *Some wild cells and spheres in three-dimensional space*, Annals of Mathematics **49** (1948), 979–990.
[50] R.H. Fox and J.W. Milnor, *Singularities of 2-spheres in 4-space and equivalence of knots*, Bulletin of the AMS **63** (1957), 406.

[51] R.H. Fox and J.W. Milnor, *Singularities of 2-spheres in 4-space and cobordism of knots*, Osaka Mathematical Journal **3** (1966), 257–267; reprinted in [114, pp. 59–69].

[52] F. Frankl and L. Pontrjagin, *Ein Knotensatz mit Anwendung auf die Dimensionstheorie*, Mathematische Annalen **102** (1930), 785–789.

[53] H. Freudenthal, *Leibniz und die Analysis Situs*, Studia Leibnitiana **4** (1972), 61–69.

[54] P. Freyd, D. Yetter, J. Hoste, W.B.R. Lickorish, K. Millett and A. Ocneanu, *A new polynomial invariant of knots and links*, Bulletin of the AMS New Series **12** (1985), 239–246.

[55] R. Fricke and F. Klein, *Vorlesungen über die Theorie der Automorphen Funktionen*, Vol. 1, Teubner, Leipzig, 1897.

[56] D. Gabai, *Surgery on knots in solid tori*, Topology **28** (1989), 1–6.

[57] C.F. Gauss, *Werke*, 12 vols, Königliche Gesellschaft der Wissenschaften, Göttingen, 1863–1933.

[58] C. Geertz, *Thick description. Toward an interpretive theory of culture*, The Interpretation of Cultures. Selected Essays, C. Geertz, Basic Books, New York, 1973, 3–30.

[59] H. Gieseking, *Analytische Untersuchungen über topologische Gruppen*, thesis, University of Münster, 1912.

[60] C. McA. Gordon, *Some aspects of classical knot theory*, Knot Theory, Proceedings Plan-sur-Bex 1977, J.C. Hausmann, ed., Lecture Notes in Mathematics vol. 685, Springer, Berlin, 1978, 1–60.

[61] C. McA. Gordon, *Dehn surgery on knots*, Proceedings of the International Congress of Mathematicians, Kyoto 1990, I. Satake, ed., Vol. 1, Springer, Berlin, 1991, 631–642.

[62] C. McA. Gordon and J. Luecke, *Knots are determined by their complements*, Bulletin of the AMS New Series **20** (1989), 83–87.

[63] W. Haken, *Über Flächen in 3-dimensionalen Mannigfaltigkeiten. Lösung des Isotopieproblems für den Kreisknoten*, Proceedings of the International Congress of Mathematicians, Amsterdam 1954, Vol. 1, North-Holland, Amsterdam, 1957, 481–482.

[64] W. Haken, *Theorie der Normalflächen*, Acta Mathematica **105** (1961), 245–375.

[65] W. Haken, *Über das Homöomorphieproblem der 3-Mannigfaltigkeiten, I*, Mathematische Zeitschrift **76** (1962), 89–120.

[66] W. Haken, *Connections between topological and group-theoretical decision problems*, Word Problems, W.W. Boone et al., eds, North-Holland, Amsterdam, 1968, 427–441.

[67] P. de la Harpe, *Introduction to knot and link polynomials*, Fractals, Quasicrystals, Chaos, Knots and Algebraic Quantum Mechanics, Proceedings Acquafredda di Maratea 1987, A. Amann et al., eds, Kluwer, Dordrecht, 1988, 233–263.

[68] M.G. Haseman, *On knots, with a census of the amphicheirals with twelve crossings*, Transactions RSE **52** (1918), 235–255.

[69] P. Heegaard, *Forstudier til en topologisk teori for de algebraiske fladers sammenhæng*, Det Nordiske Forlag, København, 1898. French translation: *Sur l'Analysis situs*, Bulletin de la Société Mathématique de France **44** (1916), 161–242.

[70] H. von Helmholtz, *Ueber Integrale der hydrodynamischen Gleichungen, welche der Wirbelbewegung entsprechen*, Journal für die Reine und Angewandte Mathematik **55** (1858), 25–55.

[71] G. Hemion, *On the classification of homeomorphisms of 2-manifolds and the classification of 3-manifolds*, Acta Mathematica **142** (1979), 123–155.

[72] H.M. Hilden, *Every closed orientable 3-manifold is a 3-fold branched covering space of S^3*, Bulletin of the AMS **80** (1974), 1243–1244.

[73] H.M. Hilden, *Three-fold branched coverings of S^3*, American Journal of Mathematics **98** (1976), 989–997.

[74] H.M. Hilden, M.T. Lozano and J.M. Montesinos, *Universal knots*, Bulletin of the AMS New Series **8** (1983), 449–450.

[75] A. Hurwitz, *Über Riemannsche Flächen mit gegebenen Verzweigungspunkten*, Mathematische Annalen **39** (1891), 1–61.

[76] V.F.R. Jones, *Index for subfactors*, Inventiones Mathematicae **72** (1983), 1–25.

[77] V.F.R. Jones, *A polynomial invariant for knots via von Neumann algebras*, Bulletin of the AMS New Series **12** (1985), 103–111.

[78] V.F.R. Jones, *Braid groups, Hecke algebras, and type II_1 factors*, Geometric Methods in Operator Algebras, Proceedings Kyoto/Los Angeles 1986, H. Araki and E.G. Effros, eds, Longman, Harlow, 242–273.

[79] V.F.R. Jones, *Hecke algebra representations of braid groups and link polynomials*, Annals of Mathematics **126** (1987), 335–388.

[80] V.F.R. Jones, *On knot invariants related to some statistical mechanical models*, Pacific Journal of Mathematics **137** (1989), 311–334.
[81] V.F.R. Jones, *Knot theory and statistical mechanics*, Scientific American **262/11** (1990), 98–103.
[82] K. Kähler, *Verzweigungen einer algebraischen Funktion zweier Veränderlicher in der Umgebung einer singulären Stelle*, Mathematische Zeitschrift **30** (1929), 188–204.
[83] L.H. Kauffman, *The Conway polynomial*, Topology **20** (1981), 101–108.
[84] L.H. Kauffman, *State models and the Jones polynomial*, Topology **26** (1987), 395–407.
[85] L.H. Kauffman, *An invariant of regular isotopy*, Transactions of the AMS **318** (1990), 417–471.
[86] R. Kirby, *A calculus for framed links*, Inventiones Mathematicae **45** (1978), 35–56.
[87] R. Kirby, *Problems in low-dimensional manifold theory*, Algebraic and Geometric Topology, Proceedings of Symposia in Pure Mathematics vol. 32, R.J. Milgram, ed., AMS, Providence, RI, 1978, 273–312.
[88] T.P. Kirkman, *The enumeration, description, and construction of knots with fewer than 10 crossings*, Transactions RSE **32** (1884–1885), 281–309; read 1884.
[89] T.P. Kirkman, *The 364 unifilar knots of ten crossings, enumerated and described*, Transactions RSE **32** (1884–1885), 483–506; read 1885.
[90] S. Lefschetz, *James Waddell Alexander (1888–1971)*, Yearbook of the American Philosophical Society (1973), Philadelphia, 1974, 110–114.
[91] G.W. Leibniz, *La Caractéristique Géométrique*, J. Echeverria, ed., J. Vrin, Paris, 1995.
[92] W.B.R. Lickorish, *A representation of orientable combinatorial 3-manifolds*, Annals of Mathematics **76** (1962), 531–540.
[93] W.B.R. Lickorish, *The panorama of polynomials for knots, links and skeins*, Braids, Proceedings of the AMS–IMS–SIAM Joint Summer Research Conference 1986, J.S. Birman and A. Libgober, eds, Contemporary Mathematics vol. 78, AMS, Providence, RI, 1986, 399–414.
[94] J.B. Listing, *Vorstudien zur Topologie*, Göttinger Studien **2** (1847), 811–875; separately published: Vandenhoeck und Ruprecht, Göttingen, 1848.
[95] C.N. Little, *Non-alternate ± knots of orders eight and nine*, Transactions of the RSE **35** (1889), 663–664.
[96] C.N. Little, *Alternate ± knots of order 11*, Transactions of the RSE **36** (1890) 253–255.
[97] C.N. Little, *Non-alternate ± knots*, Transactions of the RSE **39** (1900), 771–778.
[98] W. Magnus, *Über Automorphismen von Fundamentalgruppen berandeter Flächen*, Mathematische Annalen **109** (1934), 617–646.
[99] W. Magnus, *Noneuclidean Tesselations and Their Groups*, Academic Press, New York, 1974.
[100] W. Magnus, *Braid groups. A survey*, Proceedings of the Second International Conference on the Theory of Groups, 1973, M.F. Newman, ed., Lecture Notes in Mathematics vol. 372, Springer Verlag, Berlin, 1974, 463–487.
[101] W. Magnus, *Max Dehn*, Mathematical Intelligencer **1** (1978), 32–42.
[102] A.A. Markov, *Über die freie Ăquivalenz der geschlossenen Zöpfe*, Receuil Mathématique (Nouvelle Série) = Matematiceski Sbornik **1 (43)** (1936), 73–78.
[103] A.A. Markov, *Insolubility of the problem of homeomorphy*, Proceedings of the International Congress of Mathematicians, Edinburgh 1958, Cambridge University Press, Cambridge, 1960, 300–306.
[104] J.C. Maxwell, *A Treatise on Electricity and Magnetism*, 2 vols, Clarendon Press, Oxford, 1873; reprint of the third edition 1891, Dover, New York, 1954.
[105] J.C. Maxwell, *Art. "Atom"*, Encyclopedia Britannica, 9th edition, Vol. 3 (1875); reprinted in: J.C. Maxwell, Scientific Papers, 2 vols, W.D. Niven, Cambridge University Press, Cambridge, 1890.
[106] J.C. Maxwell, *Scientific Letters and Papers*, 2 vols, P.M. Harman, ed., Cambridge University Press, Cambridge, 1990/1995.
[107] H. Mehrtens, *Die Entstehung der Verbandstheorie*, Gerstenberg, Hildesheim, 1979.
[108] W. Menasco and M.B. Thistlethwaite, *The classification of alternating links*, Annals of Mathematics **138** (1993), 113–171.
[109] K. Menger, *Bemerkungen zu Grundlagenfragen, IV. Axiomatik der endlichen Mengen und der elementargeometrischen Verknüpfungsbeziehungen*, Jahresberichte der DMV **37** (1928), 309–325.
[110] J.W. Milnor, *On the total curvature of knots*, Annals of Mathematics **52** (1950), 248–257.
[111] J.W. Milnor, *Link groups*, Annals of Mathematics **59** (1954), 177–195; reprinted in [114, 7–25].
[112] J.W. Milnor, *Isotopy of links*, Algebraic Geometry and Topology. A Symposium in Honor of S. Lefschetz, R.H. Fox et al., eds, Princeton University Press, Princeton, 1957, 280–306; reprinted in [114, 27–53].
[113] J.W. Milnor, *Singular Points of Complex Hypersurfaces*, Annals of Mathematics Studies vol. 61, Princeton University Press, Princeton, 1968.

[114] J.W. Milnor, *Collected Papers*, Vol. 2, Publish or Perish, Inc., Houston, 1995.

[115] E.E. Moise, *The invariance of the knot-types*, Proceedings of the International Congress of Mathematicians, Amsterdam 1954, Vol. 2, North-Holland, Amsterdam, 1957, 242–243.

[116] E.E. Moise, *Affine structures in 3-manifolds, VII. Invariance of the knot types; local tame embedding*, Annals of Mathematics **59** (1954), 547–560.

[117] J.M. Montesinos, *A representation of closed, orientable 3-manifolds as 3-fold branched covers of* S^3, Bulletin of the AMS **80** (1974), 845–846.

[118] J.M. Montesinos, *Three-manifolds as 3-fold branched coverings of* S^3, Quarterly Journal of Mathematics Oxford **27** (1976), 85–94.

[119] G.D. Mostow, *Strong Rigidity of Locally Symmetric Spaces*, Annals of Mathematics Studies vol. 78, Princeton University Press, Princeton, 1973.

[120] K. Murasugi, *On a certain numerical invariant of link types*, Transactions of the AMS **117** (1965), 387–422.

[121] F.J. Murray and J. von Neumann, *On rings of operators*, Annals of Mathematics **37** (1936), 116–229.

[122] F.J. Murray and J. von Neumann, *On rings of operators, IV*, Annals of Mathematics **44** (1943), 716–808.

[123] J. von Neumann, *Continuous geometry*, Proceedings of the NAS **22** (1936), 92–100.

[124] J. von Neumann, *Examples of continuous geometry*, Proceedings of the NAS **22** (1936), 101–108.

[125] J. von Neumann, *On an algebraic generalization of the quantum mechanical formalism, part I*, Receuil Mathématique (Nouvelle Série) = Matematiceski Sbornik **1 (43)** (1936), 415–484.

[126] E. Pannwitz, *Eine elementargeometrische Eigenschaft von Verschlingungen und Knoten*, Mathematische Annalen **108** (1933), 629–672.

[127] C.D. Papakyriakopoulos, *On Dehn's lemma and the asphericity of knots*, Annals of Mathematics **66** (1957), 1–26.

[128] K. Perko, *On the classification of knots*, Proceedings of the AMS **45** (1974), 262–266.

[129] J.-C. Pont, *La Topologie Algébrique des Origines à Poincaré*, Presses Universitaires de France, Paris, 1974.

[130] G. Prasad, *Strong rigidity of* $\mathbb{Q}$*-rank 1 lattices*, Inventiones Mathematicae **21** (1973), 255–286.

[131] J. Przytycki, *History of the knot theory from Vandermonde to Jones*, Aportaciones Matemáticas Comunicaciones **11** (1992), 173–185.

[132] J. Przytycki and P. Traczyk, *Invariants of links of Conway type*, Kobe Journal of Mathematics **4** (1987), 115–139.

[133] V. Puiseux, *Recherches sur les fonctions algébriques*, Journal des Mathématiques Pures et Appliquées (I) **15** (1850), 365–480.

[134] K. Reidemeister, *Knoten und Gruppen*, Abhandlungen MSHU **5** (1926), 7–23.

[135] K. Reidemeister, *Elementare Begründung der Knotentheorie*, Abhandlungen MSHU **5** (1926), 24–32.

[136] K. Reidemeister, *Exaktes Denken*, Philosophischer Anzeiger **3** (1928), 15–47.

[137] K. Reidemeister, *Knotentheorie*, Springer, Berlin, 1932.

[138] R. Riley, *A quadratic parabolic group*, Mathematical Proceedings of the Cambridge Philosophical Society **77** (1975), 281–288.

[139] R. Riley, *Discrete parabolic representations of link groups*, Mathematika **22** (1975), 141–150.

[140] D. Rolfsen, *Knots and Links*, Publish or Perish, Inc., Berkeley, 1976.

[141] E. Scholz, *Geschichte des Mannigfaltigkeitsbegriffs von Riemann bis Poincaré*, Birkhäuser, Basel, 1980.

[142] O. Schreier, *Über die Gruppen* $A^a B^b = 1$, Abhandlungen MSHU **3** (1924), 167–169.

[143] H. Schubert, *Die eindeutige Zerlegbarkeit eines Knotens in Primknoten*, Sitzungsberichte der Heidelberger Akademie der Wissenschaften, Math.-Nat. Klasse (1949), 57–104.

[144] H. Schubert, *Knoten und Vollringe*, Acta Mathematica **90** (1953), 131–286.

[145] H. Schubert, *Über eine numerische Knoteninvariante*, Mathematische Zeitschrift **61** (1954), 245–288.

[146] H. Schubert, *Knoten mit zwei Brücken*, Mathematische Zeitschrift **65** (1956), 133–170.

[147] H. Schubert, *Bestimmung der Primfaktorzerlegung von Verkettungen*, Mathematische Zeitschrift **76** (1961), 116–148.

[148] H. Seifert, *Topologie dreidimensionaler gefaserter Räume*, Acta Mathematica **60** (1932), 147–238.

[149] H. Seifert, *Verschlingungsinvarianten*, Sitzungsberichte der Preussischen Akademie der Wissenschaften Berlin **26** (1933), 811–828.

[150] H. Seifert, *Über das Geschlecht von Knoten*, Mathematische Annalen **110** (1934), 571–592.

[151] C.L. Siegel, *Zur Geschichte des Frankfurter Mathematischen Seminars*, Frankfurter Universitätsreden **34** (1964); reprinted in C.L. Siegel, Gesammelte Abhandlungen, Vol. III, de Gruyter, Berlin, 1966, 462–474.

[152] D.M. Siegel, *Thomson, Maxwell, and the universal ether in Victorian physics*, Conceptions of Ether, G.N. Cantor and M.J.S. Hodge, eds, Cambridge University Press, Cambridge, 1981, 239–268.

[153] P.G. Tait, *On knots*, Transactions of the RSE **28** (1877), 145–190; reprinted in: P.G. Tait, *Collected Scientific Papers*, Vol. 1, Cambridge University Press, Cambridge, 1898, 273–317.

[154] P.G. Tait, *On knots. Part II*, Transactions RSE **32** (1884–1885), 327–339; read 1884; reprinted in: P.G. Tait, *Collected Scientific Papers*, Vol. 1, Cambridge University Press, Cambridge, 1898, 318–333.

[155] P.G. Tait, *On knots. Part III*, Transactions RSE **32** (1884–1885), 493–506; read 1885; reprinted in: P.G. Tait, *Collected Scientific Papers*, Vol. 1, Cambridge University Press, Cambridge, 1898, 335–347.

[156] M.B. Thistlethwaite, *Knot tabulations and related topics*, Aspects of Topology. In Memory of Hugh Dowker, I.M. James and E.H. Kronheimer, eds, Cambridge University Press, Cambridge, 1985, 1–76.

[157] W. Thomson, *On vortex atoms*, Proceedings of the RSE **6** (1866–1869), 94–105; read 1867; reprinted in: W. Thomson, *Mathematical and Physical Papers*, Vol. 4, Cambridge University Press, Cambridge, 1910, 1–12.

[158] W. Thomson, *On vortex motion*, Transactions of the RSE **25** (1869), 217–260; reprinted in: W. Thomson, *Mathematical and Physical Papers*, Vol. 4, Cambridge University Press, Cambridge, 1910, 13–66.

[159] A. Thompson, *Review of "The classification of knots and 3-dimensional spaces", by G. Hemion*, Bulletin of the AMS New Series **31** (1994), 252–254.

[160] W.P. Thurston, *The Geometry and Topology of 3-Manifolds*, Lecture notes, Princeton University, 1978–1979.

[161] W.P. Thurston, *Three dimensional manifolds, Kleinian groups, and hyperbolic geometry*, Bulletin of the AMS New Series **6** (1982), 357–381.

[162] H. Tietze, *Über die topologischen Invarianten mehrdimensionaler Mannigfaltigkeiten*, Monatshefte für Mathematik und Physik **19** (1908), 1–118.

[163] A.T. Vandermonde, *Remarques sur les problèmes de situation*, Mémoires de l'Académie Royale des Sciences de Paris (1771), 566–574.

[164] V.A. Vassiliev, *Topology of complements to discriminants and loop spaces*, Theory of Singularities and its Applications, V.I. Arnold, ed., AMS, Providence, RI, 1990, 9–21.

[165] V.A. Vassiliev, *Cohomology of knot spaces*, Theory of Singularities and its Applications, V.I. Arnold, ed., AMS, Providence, RI, 1990, 23–69.

[166] K. Volkert, *Das Homöomorphieproblem insbesondere der 3-Mannigfaltigkeiten in der Topologie*, Habilitationsschrift, 2 vols, University of Heidelberg, 1994.

[167] F. Waldhausen, *On irreducible 3-manifolds which are sufficiently large*, Annals of Mathematics **87** (1968), 56–88.

[168] F. Waldhausen, *Recent results on sufficiently large 3-manifolds*, Algebraic and Geometric Topology, R.J. Milgram, ed., Proceedings of Symposia in Pure Mathematics vol. 32, AMS, Providence, RI, 1978, 21–38.

[169] A.H. Wallace, *Modifications and cobounding manifolds*, Canadian Journal of Mathematics **12** (1960), 503–528.

[170] A. Weil, *Riemann, Betti and the birth of topology*, Archive for History of Exact Sciences **20** (1979), 91–96.

[171] J.H.C. Whitehead, *On doubled knots*, Journal of the LMS **12** (1937), 63–71.

[172] W. Whitten, *Knot complements and groups*, Topology **26** (1987), 41–44.

[173] R.L. Wilder, *The mathematical work of R.L. Moore. Its background, nature, and influence*, Archive for History of Exact Sciences **26** (1982), 93–97.

[174] W. Wirtinger, *Über die Verzweigungen bei Funktionen von zwei Veränderlichen*, Jahresberichte der DMV **14** (1905), 517.

[175] O. Zariski, *On the topology of algebroid singularities*, American Journal of Mathematics **54** (1932), 453–465.

This Page Intentionally Left Blank

CHAPTER 12

Topology and Physics – a Historical Essay

Charles Nash

Department of Mathematical Physics, National University of Ireland, Maynooth, Ireland
E-mail: cnash@thphys.may.ie

1. Introduction and early happenings

In this essay we wish to embark on the telling of a story which, almost certainly, stands only at its beginning. We shall discuss the links between one very old subject, *physics*, and a much newer one, *topology*. Physics, being so much older, has a considerably longer history than does topology. After all the bulk of topology did not even exist before the beginning of the twentieth century. However, despite this disparity of antiquity between the two subjects, we shall still find it worthwhile to examine the situation before the twentieth century. We should also remind the reader that the term topology is not usually found in the literature before 1920 or so (cf. footnote 5 of this article), one finds instead the Latin terms *geometria situs* and *analysis situs*.

1.1. *From graph theory to network theory*

Perhaps the earliest occurrence of some topological noteworthiness was Euler's solution [1736] of the *bridges of Königsberg problem*. This was nothing if not a physical problem though one with a rather amusing aspect. The problem was this: in the city of Königsberg[1] were seven bridges and the local populace were reputed to ask if one could start walking at any one of the seven so as to cross all of them precisely once and end at one's starting point. Euler had the idea of associating a graph with the problem (cf. Figure 1). He then observed that a positive answer to this question requires the vertices in this graph to have an even number of edges and so the answer is *no*[2].

In devising his answer Euler gave birth to what we now call graph theory and, in so doing, was considering one of the first problems in combinatorial topology; and though graph

[1] Since 1945, when the Potsdam agreement passed the city to Russia, Königsberg has been called Kaliningrad. Kaliningrad is a Baltic sea port on a separate piece of Russian territory that lies between Lithuania and Poland.

[2] The point is that if one successfully traverses the graph in the manner required then the edges at each vertex divide naturally into pairs because one can label one as an entry edge and the other as an exit edge.

HISTORY OF TOPOLOGY
Edited by I.M. James

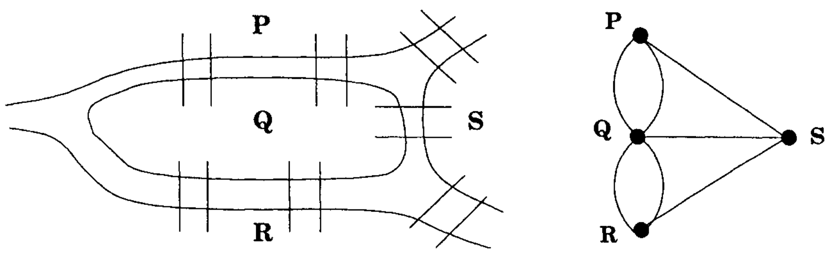


Fig. 1. The seven bridges of Königsberg and its associated graph with seven edges.

theory could once be viewed as a combinatorial study of the topology of one dimensional complexes it is now an independent subject in its own right.

Among the parts of physics with close connections to graph theory is network theory. The earliest connection occurs in the work of Kirchhoff [1847] who, as well as formulating his two famous laws for electric circuits, made use of a graph theoretic argument to solve the resulting equations for a general electrical network. From these beginnings the links between graph theory and physics have strengthened over the centuries.

1.2. *Electromagnetic theory and knots*

In the nineteenth century we encounter a more substantial example of a physical phenomenon with a topological content. The physics concerned electromagnetic theory while the topology concerned the linking number of two curves.

In electromagnetic theory the magnetic field **B** produced by a current I passing through a wire obeys the pair of equations

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J}, \qquad \nabla . \mathbf{B} = 0, \tag{1.1}$$

where **J** is the current density. Now let us briefly summarise what is entailed in deriving the famous Ampère law. We integrate **B** round a closed curve C which bounds a surface S, then we have

$$\int_C \mathbf{B} \cdot \mathbf{dl} = \int_S \nabla \times \mathbf{B} \cdot \mathbf{ds} = \mu_0 \int_S \mathbf{J} \cdot \mathbf{ds} = \mu_0 I, \tag{1.2}$$

where in the last step we use the fact that the integral $\int_S \mathbf{J} \cdot \mathbf{ds}$ gives the total current passing through S and this will be precisely I if (as we assume) the wire cuts the surface S just once. However, the wire could cut the surface more than once – say n times. When this is the case at each cut the the integral receives a contribution of $\mp\mu_0 I$ depending on whether **J** is parallel or anti-parallel to **ds** at the cut. Thus, if $m\mu_0 I$ denotes the algebraic sum of all these n contributions then the most general statement for Ampère's law in this case is

$$\int_C \mathbf{B} \cdot \mathbf{dl} = \mu_0 m I, \quad m \in \mathbb{Z}. \tag{1.3}$$

This integer m is a familiar feature of textbook calculations of the magnetic field due to a solenoid.

Now let us take an alternative route to calculating $\int_C \mathbf{B}\cdot \mathbf{dl}$. If we introduce the vector potential $\mathbf{A}$ where

$$\mathbf{B} = \nabla \times \mathbf{A}, \tag{1.4}$$

and impose the Coulomb gauge condition

$$\nabla.\mathbf{A} = \mathbf{0}, \tag{1.5}$$

then, to find $\mathbf{B}$, we only have to solve, the equation

$$\nabla^2\mathbf{A} = -\mu_0\mathbf{J} \tag{1.6}$$

and this has the solution

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0}{4\pi}\int_{\mathbb{R}^3} \mathbf{d}^3\mathbf{r}'\,\frac{\mathbf{J}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}. \tag{1.7}$$

But if the wire forms a closed curve C', say, with infinitesimal element of length $\mathbf{dl}'$ then $\mathbf{J}$ has support only on the wire and so is related to the current I by

$$\mathbf{J}(\mathbf{r}')\mathbf{d}^3\mathbf{r}' = I\mathbf{dl}'. \tag{1.8}$$

This allows us to express $\mathbf{B}$ as a line integral around C' giving us

$$\mathbf{B}(\mathbf{r}) = \frac{I\mu_0}{4\pi}\nabla \times \int_{C'}\frac{\mathbf{dl}'}{|\mathbf{r}-\mathbf{r}'|} = -\frac{I\mu_0}{4\pi}\int_{C'}\frac{(\mathbf{r}-\mathbf{r}')\times\mathbf{dl}'}{|\mathbf{r}-\mathbf{r}'|^3}. \tag{1.9}$$

Now we introduce a second curve C and integrate $\mathbf{B}$ round C thereby obtaining

$$\int_C \mathbf{B}\cdot\mathbf{dl} = -\frac{I\mu_0}{4\pi}\int_C\int_{C'}\frac{(\mathbf{r}-\mathbf{r}')\times\mathbf{dl}'\cdot\mathbf{dl}}{|\mathbf{r}-\mathbf{r}'|^3}. \tag{1.10}$$

Perusal of (1.3) and (1.10) together shows us immediately that

$$-\frac{1}{4\pi}\int_C\int_{C'}\frac{(\mathbf{r}-\mathbf{r}')\times\mathbf{dl}'\cdot\mathbf{dl}}{|\mathbf{r}-\mathbf{r}'|^3} = m. \tag{1.11}$$

For comparison with the work of Gauss below we wish to have this formula in a completely explicit form with all its coordinate dependence manifest and so we write

$$\mathbf{r} = (x, y, z), \quad \text{and} \quad \mathbf{r}' = (x', y', z') \tag{1.12}$$

giving

$$-\frac{1}{4\pi}\iint \left[(x'-x)^2 + (y'-y)^2 + (z'-z)^2\right]^{-3/2}\big((x'-x)(dy\,dz' - dz\,dy') \\ + (y'-y)(dz\,dx' - dx\,dz') + (z-z')(dx\,dy' - dy\,dx')\big) = m. \tag{1.13}$$

That this is a topological statement is clear and the integer m is actually the *linking number* of the two curves C and C' and so electromagnetic theory has provided with an explicit formula for the linking number and so an early result in knot theory.

This result (1.13) was known to Gauss in *exactly* the form that we have presented it above. Gauss's work on this matter also came from work (in 1833) on electromagnetic theory and is to be found in his *Nachlass* (Estate) cf. [Gauss, 1877] where on p. 605 one finds the formula (which I quote entirely unchanged from the printed version available in [Gauss, 1877] although one should remember that the original is handwritten rather than printed).

ZUR ELECTRODYNAMIK. 605

[4.]

Von der *Geometria Situs*, die LEIBNITZ ahnte und in die nur einem Paar Geometern (EULER und VANDERMONDE) einen schwachen Blick zu thun vergönnt war, wissen und haben wir nach anderthalbhundert Jahren noch nicht viel mehr wie nichts.

Eine Hauptaufgabe aus dem *Grenzgebeit* der *Geometria Situs* und der *Geometria Magnitudinis* wird die sein, die Umschlingungen zweier geschlossener oder unendlicher Linien zu zählen.

Es seien die Coordinaten eines unbestimmten Punkts der ersten Linie x, y, z; der zweiten x', y', z' und

$$\iint \frac{(x'-x)(dydz'-dzdy')+(y'-y)(dzdx'-dxdz'+(z-z')(dxdy'-dydx')}{\left[(x'-x)^2+(y'-y)^2+(z'-z)^2\right]^{\frac{3}{2}}} = V$$

dann ist dies Integral durch beide Linien ausgedehnt

$$= 4m\pi$$

und m die Anzahl der Umschlingungen.

Der Werth ist gegenseitig, d. i. er bleibt derselbe, wenn beide Linien gegen einander umgetauscht werden. 1833. Jan. 22.

This translates[3] to

ON ELECTRODYNAMICS. 605

[4.]

Concerning the *geometria situs*, foreseen by LEIBNITZ, and of which only a couple of geometers (EULER and VANDERMONDE) were allowed to catch a glimpse, we know and have obtained after a hundred and fifty years little more than nothing.

A main task (that lies) on the border between *geometria situs* and *geometria magnitudinis* is to count the windings of two closed or infinite lines.

The coordinates of an arbitrary point on the first line shall be x, y, z; (and) on the second x', y', z', and (let)

$$\iint \frac{(x'-x)(dydz'-dzdy')+(y'-y)(dzdx'-dxdz'+(z-z')(dxdy'-dydx')}{\left[(x'-x)^2+(y'-y)^2+(z'-z)^2\right]^{\frac{3}{2}}} = V$$

[3] I am greatly indebted to Martin Mathieu for providing me with this translation.

then this integral carried out over both lines equals

$$= 4m\pi$$

and m is the number of windings.

This value is shared, i.e., it remains the same if the lines are interchanged.

1833. Jan. 22.

We see on comparing Gauss's formula with (1.13) that they are exactly the same modulo the fact that his integer m is minus our m. Gauss's remarks[4] above also show that he understood the topological nature of his result which he quotes without reference to any electromagnetic quantities; in addition he bemoans the paucity of progress in topology in a manner which makes clear that he thinks that there is much to be discovered eventually.

Maxwell was also aware of Gauss's result and mentions it in [Maxwell, 1904a] when he discusses the conditions for the single valuedness of a function defined by a line integral. It is clear that he realises the need for a topological restriction on the domain of definition of the function. On p. 17 of [Maxwell, 1904a] he says

> There are cases, however, in which the conditions
>
> $$\frac{dZ}{dy} - \frac{dY}{dz} = 0, \quad \frac{dX}{dz} - \frac{dZ}{dx} = 0, \quad \frac{dY}{dx} - \frac{dX}{dy} = 0,$$
>
> which are those of $Xdz+Ydy+Zdz$ being a complete differential, are satisfied throughout a certain region of space, and yet the line-integral from A to P may be different for two lines, each of which lies wholly within that region. This may be the case if the region is in the form of the ring, and if the two lines from A to P pass through opposite segments of the ring .... We are here led to considerations belonging to the Geometry of Position, a subject which, though its importance was pointed out by Leibnitz and illustrated by Gauss, has been little studied. The most complete treatment of this subject has been given by J.B. Listing.[5]

Then on p. 43 of [1904b] Maxwell refers to Gauss's linking number formula and says

> It was the discovery by Gauss of this very integral, expressing the work done on a magnetic pole while describing a closed curve in presence of a closed electric current, and indicating the geometrical connexion between the two closed curves, that led him to lament the small progress made in the Geometry of Position since the time of Leibnitz, Euler and Vandermonde. We have now some progress to report, chiefly due to Riemann, Helmholtz, and Listing.

Maxwell also includes a figure for which the linking number of two oppositely oriented curves is zero.

---

[4] Additional references on the history of knot theory are [Epple, 1995, 1998] of which [Epple, 1998] also discusses the work above of Gauss on knots and, in this connection, prints a fragment from one of Gauss's notebooks which show that Gauss spent some time thinking about what the current literature now calls braids. I am indebted to Ioan James for sending me a preprint of [Epple, 1998].

[5] The work of Listing referred to by Maxwell is [Listing, 1861]. We note that Listing was the first to use the term *Topology* (actually "Topologie", since he wrote in German) in a letter to a friend in 1836, cf. the detailed account of this on pp. 41–42 of [Pont, 1974], cf. also [Listing, 1847].

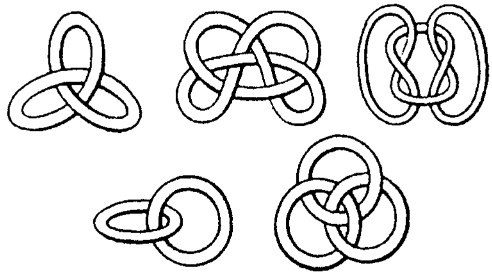


Fig. 2. Some of the vortex tubes considered by Kelvin in [Thomson, 1869].

1.3. *Knots, vortices and atomic theory*

The nineteenth century was to see another discussion of knots and physics before it came to an end. This was in the theory of vortex atoms proposed by Lord Kelvin (alias W.H. Thomson) in 1867; cf. [Thomson, 1867] for his paper on atoms as vortices and [Thomson, 1869, 1875, 1910] for his work on vortices themselves including references to knotted and linked vortices.

Kelvin was influenced by an earlier fundamental paper by Helmholtz [1858] on vortices, and a long and seminal paper of Riemann [1857] on Abelian functions.[6]

His idea really was that an atom was a kind of vortex. He was sceptical about the chemists espousal of the Lucretius atom, in [Thomson, 1867] he says:

> Lucretius's atom does not explain any of the properties of matter without attributing them to the atom itself .... The possibility of founding a theory of elastic solids and liquids on the dynamics of closely-packed vortex atoms may be reasonably anticipated.

and later in the same article

> A full mathematical investigation of the mutual action between two vortex rings of any given magnitudes and velocities passing one another in any two lines, so directed that they never come nearer one another than a large multiple of the diameter of either, is a perfect mathematical problem; and the novelty of the circumstances contemplated presents difficulties of an exciting character. Its solution will become the foundation of the proposed new kinetic theory of gases.

A significant part of the "*difficulties of an exciting character*" referred to by Kelvin above concerned the topological nature of vortices,[7] i.e. that they can be knotted and that several may be linked together (cf. Figure 2).

---

[6] Actually, in [Thomson, 1869], Kelvin specifically refers to Section 2 of this paper which is topological: it discusses multiple connectedness for what we nowadays refer to as Riemann surfaces. This section bears the title: *Lehrsätze aus der analysis situs für die Theorie der Integrale von zweigliedrigen vollständigen Differentialien* or *Theorems from analysis situs for the theory of integrals over total differentials of functions of two variables.*

[7] It seems clear that Kelvin thought of these vortex tubes as knotted tubes of the ether for the opening sentences of [Thomson, 1869], though they do not mention the word ether, describe an ideal fluid of that kind. One reads: *The mathematical work of the present paper has been performed to illustrate the hypothesis, that space is continuously occupied by an incompressible frictionless liquid acted on by no force, and that material phenomena of every kind depend solely on motions created in this liquid.* His belief in the ether lasted much further into the future cf. [Thomson, 1900] which is dated 1900; this of course is to be expected as special relativity still lay five years ahead.

Atiyah [1990] has summarised very well the main points that Kelvin considered to be in favour of his theory. Hence we provide a brief paraphrase of that summary here. The stability of atoms would be be accounted by the stability under deformation of the topological type of knots. The large variety of different knot types can accommodate all the different elements. Vibrational oscillations of knots could be the mechanism for atomic spectral lines.

With regard to this latter point about spectral lines Kelvin even gives [Thomson, 1867], a rough upper bound for the rotation period of a sodium vortex based on the frequency of its celebrated yellow emission line.

Kelvin's contemporary Tait was thereby stimulated to do extensive work on knot theory and to begin work on classifying knots cf. [Tait, 1898] where numerous complex knots are copiously illustrated as well as discussed. However, despite a lot of work, many of his results remained unproven and were christened the *Tait conjectures*. It is clear now that these conjectures were out of range of the mathematical techniques of his day; but many of them were finally disposed of in the 1980's by the work of Jones [1985] which we shall come to in Section 6. This work caused a resurgence in knot theory in mathematics and also coincided with a renewal of the physicists interest in knots. This is a rather special and interesting story cf. Section 6.

1.4. *The advent of Poincaré*

As the nineteenth century drew to a close topological matters were considerably enlivened by the work of Poincaré. This work had some strong connections with physics as we shall now explain.

Poincaré's interest in physics endured throughout his all too short life but began with Newtonian dynamics and, in particular, with the three body problem.

The Swedish mathematician Mittag-Leffler, founder and chief editor of the journal *Acta Mathematica*, was the prime mover in the organisation of an international mathematical competition held to celebrate the 60th birthday of King Oscar II of Sweden and Norway. The King was well disposed towards science and allowed a competition to be announced in 1885. Four problems were proposed but entrants (who were, in theory, anonymous) were also at liberty to choose their own topic.[8] The judges were Hermite, Mittag-Leffler and Weierstrass. Poincaré was declared the winner in 1889 – the King having officially approved the result on January 20th the day before his 60th birthday – there were eleven other entrants.

Poincaré had chosen to work on the first problem and in particular the three body problem.[9] His prize winning memoir[10] bore the title *Sur le problème de trois corps et les équations de la dynamique*.

---

[8] In brief the four problems (which were contributed by Hermite and Weierstrass) were the n-body problem, an analysis of Fuchs theory of differential equations, a problem in nonlinear differential equations and a final (algebraic) problem on Fuchsian functions.

[9] Poincaré actually studied the restricted three body problem. This meant that he took the first mass to be large, the second small and positive and the third negligible. He also took the first two masses to have a circular orbit about their common centre of mass while the third moved in the plane of the circles.

[10] The prize winning entry was to be published in *Acta Mathematica* and this was eventually done in [Poincaré, 1890]. However, the published version of Poincaré's work differs in some important respects from the entry that he submitted. This is because of an error discovered before publication by Phragmén. The mathematical and historical details of this particular story are available in the excellent book [Barrow-Green, 1997].

Before working on this problem Poincaré had worked on the theory of differential equations and had concentrated on obtaining *qualitative* results. A key viewpoint he adopted and exploited was *geometric* – he thought of the solutions as defining geometric objects: e.g., *curves*. He then quickly obtained results of a topological nature.

For example, he studied the singular points of these equations on surfaces of genus p and introduced the notions of saddle points, nodes and foci to classify these singularities – in French he used the terms *cols*, *noeuds* and *foyers*, respectively – then using C, F or N to denote the type of the singularity he proved (cf. Poincaré [1885] and [1880, 1881, 1882, 1885, 1886] that

$$N + F - C = 2 - 2p \tag{1.14}$$

which one recognises immediately as the index of a vector field being equated to the Euler–Poincaré characteristic of the Riemann surface.

As evidence of the stimulus that physics gave to investigations we quote from an analysis Poincaré prepared in 1901 of his own scientific work. This was published after his death (Poincaré [1921]).

> Pour étendre les résultats precedénts aux équations d'ordre supérieur au second, il faut renoncer à la représentation géometrique qui nous a été si commode, à moins d'employer le langage de l'hypergéometrie à n dimensions .... Ce qu'il y a de remarquable, c'est que le troisième et le quatrième cas, c'est à dire ceux qui corresponde à la stabilité, se rencontrent précisément dans les équations générales de la Dynamique .... Pour aller plus loin, il me fallait créer un instrument destiné à remplacer l'instrument géometrique qui me faisait défaut quand je voulais pénétrer dans l'espace à plus de trois dimensions. C'est la principale raison qui m'a engagé à aborder l'étude de l'Analysis situs; mes travaux à ce sujet seront exposés plus loin dans une paragraphe spécial.

that is

> To extend the preceding results to equations of higher than second order, it is necessary to give up the geometric representation which has been so useful to us unless one employs the language of hypergeometry of n dimensions .... What is remarkable is that in the third and fourth case, that is to say those that correspond to stability, are found precisely in the general equations of dynamics .... To go further it was necessary to create a tool designed to replace the geometric tool which let me down when I wanted to penetrate spaces of more than three dimensions. This is the principal reason which led me to take up the study of Analysis situs; my work on this subject will be expounded further down in a special paragraph.

As yet another insight into the way Poincaré was thinking about celestial mechanics we quote the following which is taken from the introduction to his paper *Analysis situs* [Poincaré, 1895]; this is the first of his epoch making papers on topology.

> D'autre part, dans une série de Mémoires insérés dans le Journal de Liouville, et intitulés: Sur les courbes définies par les équations différentielles, j'ai employé l'Analysis situs ordinaire à trois dimensions à l'étude des équations différentielles. Les mêmes recherches ont été poursuivies par M. Walther Dyck. On voit aisément que l'Analysis situs géneralisée permettrait de traiter de même les équations d'ordre supérieur et, en particulier, celles de la Mécanique céleste.

that is

> On the other hand, in a series of memoirs in Liouville's Journal[11] under the title On curves defined by differential equations, I have used ordinary analysis situs in three dimensions in the study of differential equations. The same research has been followed by Mr. Walther Dyck. One easily sees that generalised analysis situs would permit the treatment of equations of higher order and, in particular, those of celestial mechanics.

Poincaré's work opened a new chapter in celestial mechanics; the strong topological content of his papers on differential equations lead directly to his papers [Poincaré, 1892, 1895, 1899a, 1899b, 1900, 1901a, 1901b, 1902a, 1902b, 1904, 1912] on *analysis situs* which gave birth to the subjects of algebraic and differential topology.

1.5. *Poincaré's geometric theorem*

Poincaré left unproved at the time of his death a (now) famous result usually referred to as *Poincaré's geometric theorem.* This theorem has both physical and topological content. Shortly before he died, Poincaré wrote [1912] in order to describe the theorem and the reasons for his believing it to be true.

In the first paragraph (of [Poincaré, 1912]) he says

> Je n'ai jamais présenté au public un travail aussi inachevé; je crois donc nécessaire d'expliquer en quelques mots les raisons qui m'ont déterminé à le publier, et d'abord celles qui m'avaient engagé à l'entreprendre. J'ai démontré il y a longtemps déjà, l'existence des solutions périodiques du problème des trois corps; le résultat laissait cependant encore à désirer; car si l'existence de chaque sorte de solution était établie pour les petites valeurs des masses, on ne voyait pas ce qui devait arriver pour des valeurs plus grandes, quelles étaient celles de ces solutions qui subsistaient et dans quel ordre elles disparaissaient. En réfléchissant à cette question, je me suis assuré que la réponse devait dépendre de l'exactitude ou la fausseté d'un certain théorème de géometrie dont l'énoncé est très simple, du moins dans le cas du problème restreint et des problèmes de Dynamique où il n'y a que deux degrés de liberté.

that is

> I have never presented to the public such an incomplete work; I believe it necessary therefore to explain in a few words the reasons which have decided me to publish it, and first of all those (reasons) which had led me to undertake it. I proved, a long time ago now, the existence of periodic solutions to the three body problem; however the result left something to be desired; for if the existence of each sort of solution were established for small values of the masses, one couldn't see what might happen for larger values, (and) what were those (values) of those solutions which persisted and in what order they disappeared. On reflecting on this question, I have ascertained that the answer ought to depend on the truth or falsity of a certain geometrical theorem which is very simple to state, at least in the case of the restricted (three body) problem and in dynamical problems with only two degrees of freedom.

The geometric theorem is a *fixed point theorem*: it states that a continuous, one to one, area preserving map f from an annulus $0 < a \leqslant r \leqslant b$ to itself has a pair of fixed points (f is also required to have the property that it maps the two boundary circles in

[11] Poincaré uses the term Liouville's Journal to refer to the *Journal de Mathématiques* as found, for example, in [Poincaré, 1881].

opposite senses). When applied to the restricted three body problem it proves the existence of infinitely many periodic solutions.

Topology immediately enters because Poincaré's index theorem can easily be used to show – as Poincaré himself pointed out [Poincaré, 1912]) – that there must be an even number of fixed points; hence it is sufficient to prove that f has at least one fixed point.

In 1913, Birkhoff, who was to be a prime mover and founder of the new subject of *dynamical systems*, proved the theorem cf. [Birkhoff, 1913].

After Poincaré the pursuit of problems with a joint dynamical and topological content was taken up by Birkhoff, Morse, Kolmogorov, Arnold and Moser and many, many others. We shall have something to say about these matters later on in Section 8.

2. A quiescent period

2.1. *Dirac and Schwarzschild*

In the early twentieth century physics was undergoing the twin revolutions of quantum theory and special and general relativity. These revolutions imported much new mathematics into physics but topology, though itself growing at an explosive rate, did not figure prominently in the physics of this story.

Nevertheless two papers on physics of this period, do have a topological content, and are worth noting. In the first case this content is implicit and in the other it is explicit. These papers are, respectively, that of Schwarzschild [1916a, 1916b]) on solutions to Einstein's equations and that of Dirac [1931] on magnetic monopoles.

We shall discuss Schwarzschild's paper in Section 3 but for the moment we want to deal with Dirac's paper because its topological content is explicit from the outset and because it has proved to be so influential.

2.2. *Dirac's magnetic monopoles*

Dirac begins his paper with some comments on the rôle of mathematics in physics which are both philosophical and somewhat prophetic. He says

> The steady progress of physics requires for its theoretical formulation a mathematics that gets continually more advanced. This is only natural and to be expected. What however was not expected by the scientific workers of the last century was the particular form that the line of advancement of the mathematics would take, namely, it was expected that the mathematics would get more and more complicated, but would rest on a permanent basis of axioms and definitions, while actually the modern physical developments have required a mathematics that continually shifts its foundations and gets more abstract. Non-Euclidean geometry and noncommutative algebra, which were at one time considered to be purely fictions of the mind and pastimes of logical thinkers, have now been found to be very necessary for the description of general facts of the physical world. It seems likely that this process of increasing abstraction will continue in the future and that advance in physics is to be associated with a continual modification and generalisation of the axioms at the base of the mathematics rather than with a logical development of any one mathematical scheme on a fixed foundation.

The subject matter of Dirac's paper was both topological and electromagnetic: he found that there were magnetic monopole solutions to Maxwell's equations but that the magnetic charge μ_0 of the monopole had to quantised; in addition there were two striking facts about this quantisation one mathematical and one physical. The mathematical novelty was that the quantisation was not due to the discreteness of the spectrum of an operator in Hilbert space but rather to topological considerations. The physical novelty was that the existence of even one of these monopoles would imply the quantisation of electric charge, something not hitherto achieved.[12]

Dirac considered carefully the phase of a wave function $\psi(x, y, z, t)$ of a particle in quantum mechanics. If A and γ are the amplitude and phase, respectively, then we have

$$\psi = A e^{i\gamma}. \tag{2.1}$$

Once ψ is normalised to unity, in the usual way, there remains a freedom to add a constant to the phase γ without altering the physics of the particle. Dirac wanted to exploit this fact and argue that the absolute value of γ has no physical meaning and only phase *differences* matter physically. In [Dirac, 1913] he wrote

> Thus the value of γ at a particular point has no physical meaning and only the difference between the values of γ at two different points is of any importance.

He immediately introduces a generalisation, saying

> This immediately suggests a generalisation of the formalism. We may assume that γ has no definite value at a particular point, but only a definite difference in value for any two points. We may go further and assume that this difference is not definite unless the two points are neighbouring. For two distant points there will then be a definite phase difference only relative to some curve joining them and different curves will in general give different phase differences. The total change in phase when one goes round a closed curve need not vanish.

Dirac now does two more things: he finds that this change in phase round a closed curve will give rise to an ambiguity unless it takes the same value for all wave functions, and he goes on to equate this phase change to the flux of an electromagnetic field. For a wave function in three *spatial* dimensions, to which Dirac specialises, this flux is just that of a *magnetic field*. Yet closer scrutiny of the situation forces the consideration of the possibility of ψ vanishing (which it will do generically along a line in three dimensions) about which Dirac says

> There is an exceptional case, however, occurring when the wave function vanishes, since then its phase does not have a meaning. As the wave function is complex, its vanishing will require two conditions, so that in general the points at which it vanishes will lie along a line. We call such a line a nodal line.[13]

Dirac now finds that to get something new he must relax the requirement that phase change round a closed curve be the same for all wave functions; he realises that it is possible to have it differ by integral multiples of 2π for different wave functions.

[12] Electric charge, not being the eigenvalue of any basic operator, is not quantised by the mechanism that quantises energy and angular momentum.

[13] In the current literature nodal lines are called Dirac strings.

The final stage of the argument is to compute the flux of the magnetic field **B** through a closed surface S allowing also that nodal lines may lie totally within S or may intersect with S.

Each nodal line is labelled by an integer n which one detects by integrating round a small curve enclosing the line. If n_i are the integers for the nodal lines inside, or intersecting with, S the n_i are then related to the magnetic flux $\int_S \mathbf{B} \cdot \mathbf{ds}$ by

$$\sum_i 2\pi n_i + \frac{2\pi e}{hc} \int_S \mathbf{B} \cdot \mathbf{ds} = 0, \tag{2.2}$$

where h, c and e are Planck's constant, the velocity of light and the charge on the electron, respectively.[14] Now if the nodal lines are closed they always cross S an even number of times and hence contribute zero to $\sum_i 2\pi n_i$ when one takes account of the sign of contributions associated with incoming and outgoing lines. Hence $\sum_i 2\pi n_i$ is only nonzero for those lines having end points within S. Finally Dirac then surrounds just *one* of these end points with a small surface S so that (2.2) immediately implies that magnetic flux emanates from this end point so that it is the location of a *magnetic monopole*. Such an end point will also be a singularity of the electromagnetic field.

Now if a single electric charge produces a field **E** then the size of its charge q is given by

$$q = \frac{1}{4\pi} \int_S \mathbf{E} \cdot \mathbf{ds}, \tag{2.3}$$

where S is a small surface enclosing the charge. So, by analogy, the magnetic charge μ_0 of a magnetic monopole is defined by writing

$$\mu_0 = \frac{1}{4\pi} \int_S \mathbf{B} \cdot \mathbf{ds}. \tag{2.4}$$

Applying this to our small surface S and using (2.2) this gives at once the result that[15]

$$2\pi n + \frac{8\pi^2 e}{hc}\mu_0 = 0 \Rightarrow e\mu_0 = -\frac{nhc}{4\pi}, \quad n \in \mathbb{Z}. \tag{2.5}$$

We see that the magnetic charge μ_0 has a quantised strength with the fundamental quantum being

$$\frac{hc}{4\pi e} \tag{2.6}$$

which we note is inversely proportional to the electric charge e. Dirac noted the key fact that the mere existence of such a monopole means that electric charge is quantised. In [Dirac, 1931] we find the passage

[14] Informally the reason that the RHS of (2.2) is zero is because it is the limit of the change in phase round a closed curve C as C shrinks to zero. C bounds a surface S so that when C has finally shrunk to zero S becomes closed.

[15] When reading [Dirac, 1931] one should be aware that the symbol h denotes Planck's constant divided by 2π; nowadays this quantity is usually denoted by $\hbar$.

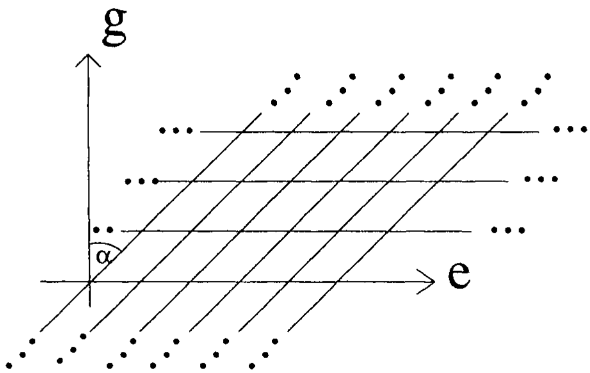


Fig. 3. The lattice of possible magnetic and electric charges.

> The theory also requires a quantisation of electric charge, since any charged particle moving in the field of a pole of strength μ_0 must have for its charge some integral multiple (positive or negative) of e, in order that the wave functions describing the motion may exist.

Dirac's remarkable paper encouraged physicists to consider particles which are simultaneously magnetically and electrically charged – such particles are called *dyons*. The set of possible electric and magnetic charges form a skew lattice covering an entire $\mathbb{R}^2$ (cf. Figure 3). For convenience from now on we denote magnetic charge by g rather than Dirac's μ_0 and in doing so we change to more conventional units of magnetic and electric charge: in these units $\hbar = c = 1$ and the charge q on the electron is given by $q^2/(4\pi) = 1/137$. Dirac's condition (2.5) now reads

$$\frac{eg}{2\pi} = n \in \mathbb{Z}. \tag{2.7}$$

The quantisation condition for dyons is a little more complicated than that for particles with only one type of charge. It has been studied independently by Schwinger [1968] and Zwanziger [1968a, 1968b]. They found that if (e_1, g_1) and (e_2, g_2) represent the electric and magnetic charges of a pair of dyons then

$$\frac{(e_1 g_2 - e_2 g_1)}{2\pi} = n \in \mathbb{Z}$$

and these values we represent on the lattice of Figure 3. If the angle α of Figure 3 is precisely zero then the lattice becomes rectangular rather than skew: this is practically forced if the theory is *CP* invariant since *CP* acts on (e, g) to give $(-e, g)$ (actually $\alpha = \pi/4$ is also allowed). Hence, the size of α is a measure of *CP* breaking which is experimentally found to be small.

For monopoles coming from *non-Abelian* gauge theories serious consideration must be given to a specific topological mechanism for this *CP* breaking. We shall expand on this remark in Section 4.

2.3. *The topology of Dirac's monopoles*

It is very instructive to examine Dirac's work from the point of view of topology. The mathematical setting is that of connections and curvatures on fibre bundles together with the vital calculational tool of characteristic classes; this latter was only at the very beginning of its development when [Dirac, 1931] was published.

The standard physical setting is as follows. In electromagnetic theory the electric and magnetic fields form the components of the Maxwell field tensor $F_{\mu\nu}$ according to[16]

$$\begin{aligned} &E_i = F_{i0}, \quad B_i = \frac{1}{2}\varepsilon_{ijk}F_{jk}, \\ &F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu. \end{aligned} \tag{2.8}$$

However, geometrically speaking the $F_{\mu\nu}$ are also the components of a curvature tensor for the gauge potential A_μ. This suggests immediately that one uses the curvature 2-form F and the connection 1-form A about which we know that

$$\begin{aligned} &F = \frac{1}{2}F_{\mu\nu}\,dx^\mu \wedge dx^\nu, \quad A = A_\mu\,dx^\mu, \\ &F = dA. \end{aligned} \tag{2.9}$$

Now because the electric field $\mathbf{E}$ is expressed as a gradient, while the magnetic field $\mathbf{B}$ is expressed as a curl, then it is natural to associate $\mathbf{E}$ to a 1-form and $\mathbf{B}$ to a 2-form. Hence we define the forms E and B by writing

$$E = E_i\,dx^i, \quad B = \frac{1}{2}F_{ij}\,dx^i \wedge dx^j. \tag{2.10}$$

This means that the curvature 2-form F is expressible as

$$F = dx^0 \wedge E + B \tag{2.11}$$

and Maxwell's equations simply assert the closure of $*F$, i.e.

$$d * F = 0, \tag{2.12}$$

where the $*$ denotes the usual Hodge dual with respect to the standard flat Minkowski metric on $\mathbb{R}^4$.

But recall that Dirac specialises to a three dimensional situation by taking a time independent electromagnetic field. We see at once from (2.11) that, in the absence of time (or at a fixed time), F pulls back to a 2-form on $\mathbb{R}^3$ which is just the magnetic field 2-form B. Thus, denoting for convenience this 2-form on $\mathbb{R}^3$ by $F|_{\mathbb{R}^3}$, we have

$$F|_{\mathbb{R}^3} = B. \tag{2.13}$$

[16] In our notation x^0 denotes the time coordinate, summation is implied for repeated indices and we use the convention that Greek and Latin indices run from 0 to 3 and 1 to 3, respectively.

It is important to note that the magnetic field in the guise of $F|_{\mathbb{R}^3}$ is a curvature on $\mathbb{R}^3$ but the same is not true of the electric field. So particularly for a *static* electromagnetic field the magnetic and electric fields are geometrically *very different*.

Dirac's monopole has to be singular on $\mathbb{R}^3$ so its curvature $F|_{\mathbb{R}^3}$ is a curvature defined on $\mathbb{R}^3 - \{0\}$ with associated connection form $A_i\,dx^i$. Now the gauge invariance of the monopole is the standard possibility of replacing A by $A + df$ where f is a function on $\mathbb{R}^3 - \{0\}$, i.e. the gauge group is the Abelian group $U(1)$.

So, in bundle theoretic terms, we have a connection on a $U(1)$ bundle P, say, over $\mathbb{R}^3 - \{0\}$; but $\mathbb{R}^3 - \{0\}$ is $S^2 \times \mathbb{R}^+$ ($\mathbb{R}^+$ is the positive real axis) and so homotopy invariance means we might as well consider P to be a $U(1)$ bundle over S^2. Such a P has an integral first Chern class $c_1(P)$ given in terms of its curvature form F by the standard formula

$$c_1(P) = \int_{S^2} \frac{F}{2\pi}, \quad c_1(P) \in \mathbb{Z}. \tag{2.14}$$

We can now illustrate everything by doing a concrete calculation: If we use spherical polar coordinates[17] (r, θ, ϕ) and take

$$A = \frac{C}{2}\big(1 - \cos(\theta)\big)\, d\phi, \quad C \text{ a constant}, \tag{2.15}$$

then the connection A has curvature

$$F = \frac{C}{2}\sin(\theta)\, d\theta \wedge d\phi \equiv \frac{C}{r^3}\varepsilon_{ijk}x^i dx^j \wedge dx^k. \tag{2.16}$$

Now F of course is the same as the magnetic field B or $F_{|_{\mathbb{R}^3}}$ introduced above but the integrality of the Chern class of P gives

$$\int_{S^2} \frac{C}{4\pi}\sin(\theta)\, d\theta \wedge d\phi = n \in \mathbb{Z} \;\Rightarrow\; C = n, \tag{2.17}$$

so that the constant C is quantised, and this ensures that P is well defined. This then is Dirac's quantisation condition (2.5) in units where magnetic charges take integral values (i.e. units of size $hc/4\pi e$).

[17] Were we to use Cartesian coordinates (x, y, z) we would obtain

$$A = \frac{C}{2r}\frac{(x\,dy - y\,dx)}{(z + r)}.$$

Note that this expression has a *genuine singularity* at $r = 0$ and a *coordinate singularity* at $z + r = 0$ – i.e. the negative z-axis – this latter singularity corresponds to Dirac's nodal line or the Dirac string. This coordinate singularity can be shifted to the positive z-axis by the gauge transformation $f = \tan^{-1}(y/x)$ yielding the connection

$$A + df = \frac{C}{2r}\frac{(x\,dy - y\,dx)}{(z - r)}.$$

2.4. *Aharonov and Bohm*

After 1931, and the appearance of Dirac's paper, neither physics nor topology stood remotely still but they largely went along separate ways. To pinpoint a significant instance of an interaction between topology and physics we pass forward nearly three decades to 1959 and the paper of Aharonov and Bohm [1959].

The Aharonov–Bohm effect [Aharonov and Bohm, 1959] is a phenomenon in which the nontriviality of a gauge field A is measurable physically even though its curvature F is zero. Moreover, this nontriviality is *topological* and can be expressed as a number n, say, which is a global topological invariant.

To demonstrate this effect physically one arranges that a *non simply connected* region Ω of space has zero electromagnetic field F. We are using the same notation as in the discussion above of Dirac's magnetic monopole, i.e. we have

$$F = dA \tag{2.18}$$

and

$$F = \frac{1}{2} F_{\mu\nu}\, dx^{\mu} \wedge dx^{\nu} \quad \text{and} \quad A = A_{\mu}\, dx^{\mu}, \tag{2.19}$$

where the x_{μ} are local, coordinates on Ω.

Given this F and Ω one can devise an experiment in which one measures a diffraction pattern associated with the parallel transport of the gauge field A round a noncontractible loop C in Ω.

A successful experiment of precisely this kind was done by Brill and Werner [1960]. The experimental setup – cf. Figure 4 – was of the Young's slits type with electrons replacing photons and with the addition of a very thin solenoid. The electrons passed through the slits and on either side of the solenoid and an interference pattern was then detected. The interference pattern is first measured with the solenoid off. This pattern is then found to *change* when the solenoid is switched on even though the electrons always pass through a region where the field F is zero.

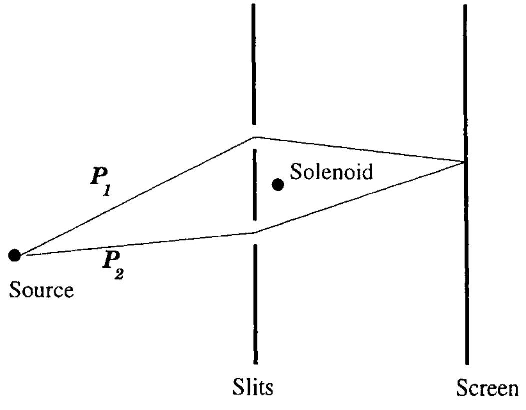


Fig. 4. A schematic Aharonov–Bohm experiment.

This result is quite a dramatic demonstration of the fact that the connection, or gauge field, A is a more fundamental object than the electromagnetic field, or curvature F; all the more so since the key point is a topological one.[18]

A topological explanation is easy to provide and begins with the notion of parallel transport round *closed* curves. The action of parallel transport on multilinear objects, viewed either as vectors, spinors, tensors etc. or sections of the appropriate bundles, is via the operator $PT(C)$ where

$$PT(C) = \exp\left[\int_C A\right].$$

Differential topology provides us immediately with the means to see that $PT(C)$ is nontrivial. The argument goes as follows: $F = dA$ so that the vanishing of F gives us

$$dA = 0 \;\Rightarrow\; A = df \text{ locally.}$$

Hence A determines a de Rham cohomology class $[A]$ and we have

$$[A] \in H^1(\Omega; \mathbb{R}).$$

It is clear from Stokes' theorem that integral $\int_C A$ only depends on the homotopy class of the loop C; in addition the loop C determines a homology class $[C]$ where

$$[C] \in H_1(\Omega; \mathbb{R}).$$

This means that the integral $\int_C A$ (which we can take to be the number n) is just the dual pairing $(\bullet, \bullet)$ between cohomology and homology, i.e.

$$\big([A], [C]\big) = \int_C A.$$

Mathematically speaking we realise the solenoid by a cylinder L so that

$$\Omega = \mathbb{R}^3 - L \;\Rightarrow\; H^1(\Omega; \mathbb{R}) = H^1\big(\mathbb{R}^3 - L; \mathbb{R}\big) = \mathbb{Z}.$$

In the experiment referred to above the loop C is the union of the electron paths P_1 and P_2, cf. Figure 4.

The holonomy group element $PT(C)$ is also of central interest elsewhere. It occurs in the study of the adiabatic periodic change of parameters of a quantum system described by Berry [1984] cf. also Simon [1983]. This phenomenon is known as Berry's phase. A relevant topological application here is an explanation of the quantum Hall effect, cf. Morandi [1988].

The importance of nontrivial flat connections for physics extends to the non-Abelian or Yang–Mills case as we shall see in due course below.

It is now time to discuss topological matters concerning general relativity.

[18] It is possible that this result would not have been a complete surprise to Maxwell because, as is shown in our first quotation from him above, he was well aware of the necessity for topological considerations, and even knew something of their nature, when a function was defined by a line integral.

3. Topology, general relativity and singularities – black holes and the big bang

3.1. *Chandrasekhar and gravitational collapse*

Penrose [1965] made the biggest breakthrough since Chandrasekhar [1931] in understanding the nature of gravitational collapse: He proved the first theorem which showed that singularities of the gravitational field are a generic feature of gravitational collapse. Moreover, Penrose's methods were topological.

Before saying anything more about Penrose's paper we need a brief sketch of some of the salient features of gravitational collapse.

Gravitational collapse is something that is worth investigating for very massive objects such as stars. This simple sounding idea is that, for a sufficiently massive body, the attractive force of gravity may be strong enough to cause it to start to *implode*.

To find something massive enough we have to choose a stellar object such as a star. Now, for a young active star, the burning of the nuclear fuel causes enough outward pressure to counteract all its gravitational inward pressure. However, since the nuclear fuel will eventually be used up this line of thought suggests that one calculate what gravity can do once it is not opposed by the nuclear reactions. In 1931 Chandrasekhar [1931] published his celebrated paper on this matter.

He took a star of mass M to be a relativistic gas at temperature T obeying a relativistic equation of state. With the star's nuclear fuel all spent its cooling and contraction under gravity was opposed by the degeneracy pressure of the electrons produced by Fermi–Dirac statistics. However, Chandrasekhar found that this pressure could not resist gravity if M was greater than about 1.4 solar masses (in standard notation one writes this as $1.4\,M_\odot$ where $M_\odot$ denotes the mass of the Sun). On the other hand for M less than $1.4\,M_\odot$ the star should cool and contract to what is called a white dwarf. Hence, for stars heavier than $1.4\,M_\odot$, unless something special intervened – for example, a mechanism causing matter to be ejected during cooling until the Chandrasekhar limit is eluded – gravitational collapse is predicted.

No one (and this included Einstein) was very comfortable with this result but it resisted all the attempts made to get round it or even to disprove it. White dwarfs as ultimate fates of cooling stars were then supplemented by neutron stars.

Neutrons stars are so dense that their protons and electrons have combined to form neutrons; these neutrons then have a degeneracy pressure which resists the gravitation of the cooling star just as the electrons do in a white dwarf. The same ideas about the maximum mass of white dwarfs apply to neutron stars which then also have a maximum mass, this varies from about $2\,M_\odot$ to $3\,M_\odot$, the precise value depending on one's knowledge of the nuclear force, or strong interactions, at high densities.

Stars which are heavy enough[19] are thought not to end up in the graveyard of white dwarfs or neutrons stars but instead continue their collapse and form black holes.[20]

---

[19] Apparently it may be possible for a large amount of matter to be shed by stars as they collapse: a star may even need a mass M greater than $20\,M_\odot$ in order to be forced to gravitationally collapse. However, there are stars known with mass M ranging up to $100\,M_\odot$ so we do expect gravitationally collapsed stars to exist. For more details cf. Chapter 9 of [Hawking and Ellis, 1973].

[20] Black holes, in the sense of *dark stars* from which light cannot escape, were discussed in Newtonian physics by Michell [1784] and Laplace [1799] (cf. Appendix A of [Hawking and Ellis, 1973] for a translation of Laplace [1799]). They took light to obey a corpuscular theory and computed the size of a star whose escape

The attitude to results about gravitational collapse as such as Chandrasekhar's [1931, 1935] was that they were properties of the unrealistically high degree of symmetry of the solutions: collapse was not expected in the real Universe where such symmetry would not be found. This was also largely the attitude taken to the much later result of Oppenheimer and Snyder [1939]. This was a a paper (in which spherical symmetry was assumed) that produced the new result that a star undergoing gravitational collapse cut itself off from external observation as it contracted through a certain critical radius – the Schwarzschild radius. It contained, too, the facts about time asymptotically slowing to zero for an external observer but not for an observer moving with the star:

> The total time of collapse for an observer co-moving with the star is finite ... an external observer sees the star asymptotically shrinking to its gravitational radius.

The Schwarzschild metric plays a central part in understanding gravitational collapse and we shall now sketch some of its main properties.

3.2. *The Schwarzschild metric*

Schwarzschild [1916a] derived the form of a (spatially) spherically symmetric metric. With spherical polar coordinates (r, θ, ϕ), and time t, it is determined by the line element

$$ds^2 = -\left(1 - \frac{2m}{r}\right) dt^2 + \left(1 - \frac{2m}{r}\right)^{-1} dr^2 + r^2\left(d\theta^2 + \sin^2\theta\, d\phi^2\right). \tag{3.1}$$

This metric is meant to represent the empty space-time outside a spherically symmetric body of mass m.[21] We see at once a singularity at $r = 0$ and one at $r = 2m$. However, the singularity at $r = 0$ is *genuine* – for example, the Riemann curvature tensor diverges there – but the singularity at $r = 2m$ is only a *coordinate singularity* and disappears in an appropriately chosen coordinate system. Incidentally this is precisely analogous to the two singularities we encountered for the Dirac monopole: one at $r = 0$ and one at $z + r = 0$; we found that $r = 0$ was a real singularity but that $z + r = 0$ was only a coordinate singularity.

However, the benign nature of the hypersurface $r = 2m$ was not realised for many years and it was usually misleadingly referred to as the "Schwarzschild singularity". This special value of r is called the *Schwarzschild radius* of the mass m. Schwarzschild [1916b] himself[22] went to the trouble of quoting the value for the Sun: it is very small, namely 3 km.

---

velocity was greater than that of light. For other details of interest cf. the article by Israel in [Hawking and Israel, 1987] and the excellent book for the layman by Thorne [1994].

[21] One might think of the Schwarzschild metric as a solution to the one body problem in general relativity; unfortunately, things are worse than in Newtonian gravity where the three body problem was so hard to solve analytically, in general relativity the two body problem has not (yet!) been solved analytically; this means that accurate approximation methods must be used to treat important problems such as binary stars. As regards the Newtonian three body problem we should add that it was finally solved analytically by Sundman [1909], cf. also Sundman [1907, 1912] and Barrow-Green [1997, p. 187]. Unfortunately, though the Sundman solution is a great triumph, his solution is a convergent series whose convergence rate is incredibly slow: apparently some 10^p terms, with p measuring in the millions, would be needed for practical work.

[22] Since this is a historical article I just add that this presumably was Schwarzschild's last article as he died of an illness while on the Russian front in 1916; he was born in 1876.

This smallness led to the conviction that $r = 2m$ was irrelevant in practice because such a value of r lay deep down in the interior of any realistic body. Hence, comfort was derived from the fact that the Schwarzschild metric was always used to describe the gravitational field in the empty space *outside* the star where r was always much bigger than the Schwarzschild radius. This was fine if a star never started to collapse but not otherwise.

The cosmologist Lemaître did notice in 1933 that $r = 2m$ was not a real singularity but this seems to have gone unnoticed or not been appreciated for a long time. In [Lemaître, 1933] we find the words

> La singularité du champ de Schwarzschild est donc une singularité fictive

i.e.

> The singularity of the Schwarzschild field is therefore a fictitious singularity

It is clear, though, that an ($r = const, t = const$) surface does change its character precisely when r passes through the value $2m$: For $r > 2m$ such a surface is timelike while for $r < 2m$ it is spacelike. This does mean that there is *something* special about the Schwarzschild radius, the question is just what is this something? Penrose was able to provide the answer and use it to make a breakthrough in understanding gravitational collapse. The point is that, for $r < 2m$, one can have what Penrose called a *trapped surface* and these we now consider.

3.3. *Penrose and trapped surfaces*

Gravitational collapse still refused to go away and in the early 1960's with the discovery of gigantic energy sources dubbed quasars the subject again became topical. It was suggested (other more conventional explanations did not seem to fit) that the energy source of a quasar came from the gravitational collapse of an immensely massive object of mass $10^6 M_\odot$–$10^9 M_\odot$. Presumably such a collapse would not be spherically symmetric, for example, one would expect there to be nonzero angular momentum. All this increased the need to study the possibility of gravitational collapse in general, i.e. without without any assumption of a special symmetry, spherical or otherwise.

Fortunately this is precisely where Penrose's topologically obtained result comes to the rescue. Penrose deduced that gravitational collapse to a space-time singularity was inevitable given certain reasonable conditions, and these conditions did *not* require any assumption about symmetry.

In [Penrose, 1965] we find the statements

> It will be shown that, after a certain critical condition has been fulfilled, deviations from spherical symmetry cannot prevent space-time singularities from arising .... The argument will be to show that the existence of a trapped surface implies – irrespective of symmetry – that singularities develop.

Special attention has to be given to providing a definition of a singularity which is both mathematically and physically reasonable. In brief geodesic completeness is used as the

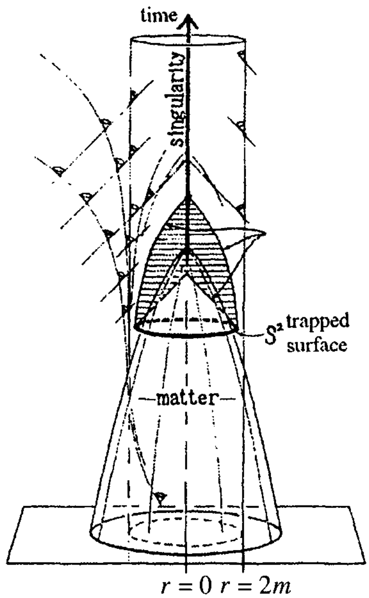


Fig. 5. Gravitational collapse and a trapped surface S^2.

basis for the definition of a singularity[23] of the space-time manifold $\mathcal{M}$; to see the significance of such completeness just consider that if a particle travelled along an *incomplete* timelike geodesic then it could disappear suddenly from $\mathcal{M}$ in a finite time.

It is impossible to give a detailed account here of the arguments so we shall only outline them; for a proper account cf. Hawking and Ellis [1973].

As the matter constituting the star contracts it passes through its Schwarzschild radius $r = 2m$ and after this has happened the matter lies totally within a spacelike sphere S^2 (cf. Figure 5). This S^2 is what is called a trapped surface; technically it is closed, compact, spacelike, two dimensional and has the property that null geodesics which intersect it orthogonally converge in the future.

The space-time manifold $\mathcal{M}$ is the future time development of an initial *noncompact* Cauchy hypersurface.

Figure 5 shows a space-time diagram of the collapse. In perusing the figure the reader should bear in mind that *one spatial dimension is suppressed* and that the circular symmetry of the diagram is there only for aesthetic reasons; the whole point being that no symmetry is assumed. The initial Cauchy hypersurface is represented by the plane at the bottom of the diagram.

[23] In relativity, since the metric is Lorentzian, geodesic completeness exists in three varieties: null, timelike and spacelike. To be singularity free both null and timelike geodesic completeness are demanded of $\mathcal{M}$; spacelike completeness is not required because physical motion does not take place along spacelike curves. In addition to geodesic completeness one also requires a causality condition and a nonnegative energy condition: The causality condition is usually stated as the absence in $\mathcal{M}$ of any closed timelike curves – causes always precede effects. The nonnegative energy condition is that, if $T_{\mu\nu}$ is the energy momentum tensor, then $T_{\mu\nu}t^\mu t^\nu \geqslant 0$ everywhere in $\mathcal{M}$ for all timelike vectors t^μ – in the rest frame of such a t^μ this becomes the statement that the energy $T_{00} \geqslant 0$ which is just another way of saying that gravity is always attractive.

The argument then computes the degree of an appropriate map which shows that null geodesic completeness implies that the future of the trapped surface is *compact*. However this is incompatible with the fact the initial Cauchy hypersurface is *noncompact*; this contradiction forces $\mathcal{M}$ to have a singularity.

Some insight into the importance of a trapped surface can be obtained from a physical discussion: Normally if light is emitted radially outwards from all points on the surface S, say, of a sphere then it creates an outward moving spherical wave with surface S'; furthermore S' has a *bigger area* than S. However, if S is a trapped surface then one finds that S' has a *smaller area* than S; this corresponds to the fact that gravity bends the light back and stops it escaping from the region inside $r = 2m$. Hence, as time progresses, S evolves to a smaller and smaller surface which eventually becomes a singularity, cf. again, Figure 5. The three surface $r = 2m$ is called the (absolute) *event horizon* of the collapse.

The use of the term black hole to describe such singularities is due to Wheeler who coined it in 1968, cf. Thorne [1994]. The entire present day Universe may have originated in a past singularity known as the *big bang* a possibility for which there is considerable experimental evidence nowadays. This has resulted in the big bang being taken very seriously. However, without the use of topological methods to convince one that singularities are generic under certain reasonable conditions the big bang would have been much more difficult to take seriously.

A further important paper on singularities was [Hawking and Penrose, 1970]; the general situation is discussed at great length and in full detail in [Hawking and Ellis, 1973].

To round things off we point out that an important consequence of this work on singularities is that the set of solutions to the general relativistic hyperbolic Cauchy problem, which are destined to evolve into singularities, form a set of positive measure in an appropriate topology.

Just as topology was becoming a permanent bed fellow of relativity it also began to play a rôle in the Yang–Mills or non-Abelian gauge theories, these theories having moved to centre stage in elementary particle theory. This was to be an even more important event for topology as it has led to a genuinely two sided interaction between theoretical physics and mathematics.[24] We begin this story in the next section.

4. Topology and Yang–Mills theory – the latter day explosion

4.1. *The rise of gauge invariance in the 1970's*

There is no doubt that a principal factor in the rise of topology in physics is due to the rise to supremacy[25] of gauge theories in physics. Topology entered, in the main, via gauge theories: physicists learned that gauge theories had a formulation in terms of fibre bundles; they learned too that much useful cohomological data was possessed by these bundles.

[24] There was also some work in the 1960's using algebraic topology to study singularities of Feynman integrals, cf. Hwa and Teplitz [1966], Froissart [1966] and Pham [1967] but this has not continued to any great extent.

[25] This, of course, is another story and it is not our task to tell it here. However it is important for the reader to be aware that the interest in non-Abelian gauge theories was rekindled almost overnight with the vital proof of the renormalisability of non-Abelian gauge theories by 't Hooft [1971] and 't Hooft and Veltman [1972]. This led to the resurrection of earlier papers on the subject and to the so called *standard model* with gauge group $SU(2) \times U(1)$ of weak and electromagnetic interactions and to QCD or *quantum chromodynamics*, with gauge group $SU(3)$, the favoured model for the confined quarks believed to be responsible for the strong interactions.

4.2. *Nielsen, Olesen, Polyakov and 't Hooft*

An early important result of this period – which we may take to be post the papers of 't Hooft [1971] and 't Hooft and Veltman [1972] – is on magnetic monopoles in *non-Abelian* gauge theories. Two independent papers ['t Hooft, 1974] and [Polyakov, 1974] produced the first *non-Abelian monopole* (now referred to as the 't Hooft–Polyakov monopole). An earlier paper by Nielsen and Olesen [1973] on magnetic vortices in superconductors was an important influence: In ['t Hooft, 1974] the author opens with

> The present investigation is inspired by the work of Nielsen et al. [1], who found that quantized magnetic flux lines, in a superconductor, behave very much like the Nambu string [2].

The 't Hooft–Polyakov monopole, like the Dirac monopole, is a static object and lives in $\mathbb{R}^3$; however, unlike the Dirac monopole, it has *no singularity* at the origin and is regular *everywhere* in $\mathbb{R}^3$. Indeed as 't Hooft [1971] says

> Our way for formulating the theory of magnetic monopoles avoids the introduction of Dirac's string [3].

The magnetic charge g of the monopole is topologically quantised and is inversely proportional to q, where q is the electric charge of a heavy gauge boson in the theory. The topology enters through the boundary condition at infinity in $\mathbb{R}^3$. We shall now attempt to elucidate this by supplying some of the details of the mathematical setting.

4.3. *The topology of monopole boundary conditions*

Monopoles are static, finite energy, objects which give the critical points of the energy of an appropriate system of fields defined on a three-dimensional Riemannian manifold M. In fact the usual choice for M is the noncompact space $\mathbb{R}^3$. Analysis on a noncompact M introduces some technical difficulties but these have not proved insurmountable.

The physical system studied consists of a Yang–Mills G-connection A, with curvature F, and a Higgs scalar field ϕ transforming according to the adjoint representation of G. If the Hodge dual with respect to the metric on M is denoted by $*$, then the energy E of the system is given by

$$E = \frac{1}{2}\int_M \left\{-\mathrm{tr}(F \wedge *F) - \mathrm{tr}(d_A\phi \wedge *d_A\phi) + \lambda * \left(|\phi|^2 - C^2\right)^2\right\}, \tag{4.1}$$

where $d_A\phi$ is the covariant exterior derivative of the Higgs field, tr denotes the trace in the Lie algebra $\mathfrak{g}$ of G and $|\phi|^2 = -2\,\mathrm{tr}(\phi^2)$. The field equations for the critical points of this system are difficult to solve explicitly (indeed the 't Hooft–Polyakov monopole is constructed numerically) but many solutions are available in what is called the Prasad–Sommerfield limit (cf. Prasad and Sommerfield [1975]) where the scalar potential term vanishes. The energy is then

$$E \equiv E(A,\phi) = -\frac{1}{2}\int_M \{\mathrm{tr}(F \wedge *F) + \mathrm{tr}(d_A\phi \wedge *d_A\phi)\}$$
$$= \frac{1}{2}\{\|F\|^2 + \|d_A\phi\|^2\}$$
$$= \frac{1}{2}\{\|F \mp *d_A\phi\|^2 \pm 2\langle F, *d_A\phi\rangle\}. \tag{4.2}$$

This shows that the absolute minima of E are attained when the pair (A,ϕ) satisfy

$$F = \mp * d_A\phi \tag{4.3}$$

which is the celebrated Bogomolny equation [Bogomolny, 1976]. The expression[26]

$$\langle F, *d_A\phi\rangle = -\int_M \mathrm{tr}(F \wedge d_A\phi) \tag{4.4}$$

is the absolute minimum and looks like a topological charge.

Now suppose that $M = \mathbb{R}^3$ furnished with the Euclidean metric and also set $G = SU(2)$. For the energy E to converge and to make the field equation problem well posed we must specify boundary conditions at infinity. A standard boundary condition for ϕ is

$$\lim_{r\to\infty} |\phi| \longrightarrow (C + \mathrm{O}(r^{-2})), \tag{4.5}$$

where r is the distance from the origin in $\mathbb{R}^3$. The integral (4.4) can now be nonzero and it does have a topological interpretation which forces it to take discrete values. More precisely, if k is an integer, then

$$\frac{1}{4\pi C}\langle F, *d_A\phi\rangle = k. \tag{4.6}$$

This integer is the magnetic charge and can be thought of as the Chern class of a $U(1)$ bundle over a two sphere which is S^2_∞, the two sphere at infinity; setting $C = 1$ and using Stokes' theorem, we have

$$k = \frac{1}{4\pi}\int_{\mathbb{R}^3} \mathrm{tr}(F \wedge d_A\phi) = \frac{1}{4\pi}\int_{S^2_\infty} \mathrm{tr}(F\phi). \tag{4.7}$$

The condition $|\phi| = 1$ on the boundary defines an S^2 inside the Lie algebra $su(2)$ and $(F\phi)$, when evaluated at infinity, becomes the $U(1)$-curvature of a bundle over S^2_∞ and k is its Chern class. Alternatively, one can write k as the winding number, or degree, of a

[26] Throughout this article the inner product $\langle\omega,\eta\rangle$ between Lie algebra valued p-forms ω and η on a Riemannian manifold M has the standard definition: i.e.

$$\langle\omega,\eta\rangle = -\int_M \mathrm{tr}(\omega \wedge *\eta).$$

map $\widehat{\phi}$: $S^2_\infty \to S^2_{su(2)}$, giving

$$\begin{aligned} &\widehat{\phi}\colon S^2_\infty \longrightarrow S^2_{su(2)}, \quad x \longmapsto \widehat{\phi}(x) = \frac{\phi}{|\phi|}, \quad \text{and} \\ &k = -\frac{1}{2\pi}\int_{S^2_\infty} \operatorname{tr}\big(\widehat{\phi}\, d\widehat{\phi} \wedge d\widehat{\phi}\big). \end{aligned} \tag{4.8}$$

This boundary integer k is the only topological invariant associated with the monopole system; the $SU(2)$ bundle over $\mathbb{R}^3$ is topologically trivial since $\mathbb{R}^3$ is contractible.

4.4. *Dyons and CP breaking*

In Section 2 we promised to return to the subject of dyons and a *topological* mechanism for *CP* breaking. This *CP* breaking requires a non-Abelian monopole since it comes from the presence in the action of a multiple of the second Chern class $c_2(P)$ given by (e and θ are real constants and for convenience we assume that the gauge group is $SU(N)$)

$$\frac{\theta e^2}{16\pi^2}\operatorname{tr}(F \wedge F) = \frac{\theta e^2}{2} c_2(P). \tag{4.9}$$

This has been discussed by Witten [1979] and it is immediate that such a term is *CP* noninvariant. The electric charge of a dyon now involves the θ parameter: one finds that for dyons with magnetic charge $g = 2\pi n_0/e$, $n_0 \in \mathbb{Z}$, their electric charge q obeys the formula

$$q = \left(n - \frac{\theta}{2\pi}\right)e, \quad n \in \mathbb{Z}. \tag{4.10}$$

This has the interesting feature that the *electric* charge of a dyon is not a *rational* multiple of a fundamental electric charge unless the *CP* violating Chern class coefficient θ is zero ($\theta = \pi$ is also allowed but may be too large experimentally). In terms of Figure 3 above it means that the angle α vanishes when θ is zero.

Still further insight into the rôle played by monopoles in quantum field theory has been obtained by combining the electric and monopole charges into a single *complex* parameter $e + ig$; we shall discuss this in Section 8.

4.5. *Gauge theories in four dimensions: Instantons*

Topology came even more to the fore in Yang–Mills theories with the publication by Belavin et al. [1975] of topologically nontrivial solutions to the *Euclidean* Yang–Mills equations in four dimensions. Such solutions have come to be called *instantons*.[27]

Of fundamental importance for these solutions to the Euclidean Yang–Mills equations is that instantons are at the same time *nonperturbative* and *topological*.

[27] Quite a few early papers on the subject used the less attractive term pseudoparticle instead of instanton but luckily this usage was short-lived.

The Euclidean version of a quantum field theory is obtained from the Minkowskian version by replacing the Minkowski time t by it. The relation between the two theories is supposed to be one of analytic continuation in the Lorentz invariant inner products $x_\mu y^\mu$; allowing these inner products to be complex is the simple way to pass from one theory to the other. The existence of such a continuation makes tacit certain assumptions which require proof; significant progress in this technical matter was made in 1973–1975, cf. Osterwalder and Schrader [1973a, 1973b, 1975], and Streater [1975].

The term instanton, though not quite precise, is often generalised to refer to a critical point of finite Euclidean action for any quantum field theory. Such solutions to the equations of motion – for this is what these critical points are – are closely related to quantum mechanical tunnelling phenomena. This property of instantons quickly attracted great interest because tunnelling amplitudes are not calculable perturbatively but require a knowledge of the theory for large coupling as well as small.

This opening of the door into the room of nonperturbative techniques was a noteworthy event and we shall see below that topology was a key ingredient to picking the lock on this door.

A good account of this is to be found in [Coleman, 1979] who showed his pleasure at the progress made in his opening sentences:

> In the last two years there have been astonishing developments in quantum field theory. We have obtained control over problems previously believed to be of insuperable difficulty and we have obtained deep (at least to me) insights into the structure of the leading candidate for the theory of strong interactions, quantum chromodynamics.

In [Coleman, 1979] there is also an account of an important paper ['t Hooft, 1976] which used instantons to solve an outstanding problem known as "the $U(1)$ problem", thereby imbuing the fledgling instantons with considerable status. We shall now give a brief summary of some of the more salient features of an instanton in the Yang–Mills case.

4.6. *Profile of an instanton*

Our life can be made a little easier by choosing a very specific setting: we have a non-Abelian gauge theory with G a compact simple Lie group and action

$$S \equiv S(A) = \|F\|^2 = -\int_M \operatorname{tr}(F \wedge *F) \tag{4.11}$$

with M a closed four dimensional orientable Riemannian manifold and $*$ the Hodge dual with respect to the Riemannian metric on M. Instantons are those A which correspond to critical points of S; however, we shall specialise the term here to mean only *minima* of S.

First we should obtain the Euler–Lagrange equations of motion, i.e. the equation for the critical points. Let A be an arbitrary connection through which passes the family of connections

$$A_t = A + ta. \tag{4.12}$$

Expanding in the vicinity of $t = 0$ gives

$$
\begin{aligned}
S(A_t) &= \big\langle F(A), F(A)\big\rangle + t\frac{d}{dt}\big\langle F(A_t), F(A_t)\big\rangle|_{t=0} + \cdots \quad \text{and} \\
F(A_t) &= F(A) + t(da + A \wedge a + a \wedge A) + t^2 a \wedge a \\
&= F(A) + t d_A a + t^2 a \wedge a \\
\Rightarrow S(A_t) &= \|F(A)\|^2 + t\big\{\big\langle d_A a, F(A)\big\rangle + \big\langle F(A), d_A a\big\rangle\big\} + \cdots \\
&= S(A) + 2t\big\langle F(A), d_A a\big\rangle + \cdots .
\end{aligned} \tag{4.13}
$$

A is a critical point if

$$\frac{dS(A_t)}{dt}\bigg|_{t=0} = 0. \tag{4.14}$$

That is, if

$$
\begin{aligned}
&\langle F(A), d_A a\rangle = 0 \\
&\quad \Rightarrow \langle d_A^* F(A), a\rangle = 0 \\
&\quad \Rightarrow d_A^* F(A) = 0, \quad \text{since } a \text{ is arbitrary.}
\end{aligned} \tag{4.15}
$$

However, $F(A) = dA + A \wedge A$ also satisfies the Bianchi identity $d_A F(A) = 0$ and so we have the pair of equations

$$d_A F(A) = 0, \qquad d_A^* F(A) = 0. \tag{4.16}$$

This is similar to the condition for a form ω to be harmonic, which is

$$d\omega = 0, \qquad d^*\omega = 0. \tag{4.17}$$

It should be emphasised, though, that the Yang–Mills equations are not linear; thus they really express a kind of nonlinear harmonic condition.

The most distinguished class of solutions to the Yang–Mills equations $d_A^* F(A) = 0$ is that consisting of those connections whose curvature is self-dual or anti-self-dual.

To see how such solutions originate we point out that with respect to our inner product on 2-forms d_A^* has the property that

$$d_A^* = -{*}d_A{*} \tag{4.18}$$

so that the Yang–Mills equations become

$$d_A * F(A) = 0. \tag{4.19}$$

Thus if $F = \mp{*}F$ the Bianchi identities immediately imply that we have a solution to the Yang–Mills equations – we have managed to solve a nonlinear second order equation by solving a nonlinear first order equation.

It is also easy to see that these critical points are all *minima* of the action S; here are the details. First we (orthogonally) decompose F into its self-dual and anti-self-dual parts F^+ and F^-, giving

$$\begin{aligned} F &= \frac{1}{2}(F + *F) + \frac{1}{2}(F - *F) = F^+ + F^- \\ &\Rightarrow S = \left\|(F^+ + F^-)\right\|^2 = \left\|F^+\right\|^2 + \left\|F^-\right\|^2 \end{aligned} \tag{4.20}$$

where the crossed terms in the norm contribute zero.

The *topological type* of the instanton is classified by the second Chern class $c_2(F) \in H^2(M; \mathbb{Z})$ of the bundle on which the connection A is defined: Taking G to be the group $SU(N)$ and evaluating $c_2(F)$ on M we obtain the integer

$$c_2(F)[M] = \frac{1}{8\pi^2}\int_M \mathrm{tr}(F \wedge F) \in \mathbb{Z}. \tag{4.21}$$

The *instanton number*, k, is defined to be minus this number so we find that

$$\begin{aligned} k &= -\frac{1}{8\pi^2}\int_M \mathrm{tr}(F \wedge F) = -\frac{1}{8\pi^2}\int_M \mathrm{tr}\{(F^+ + F^-) \wedge (F^+ + F^-)\} \\ &= \frac{\|F^+\|^2 - \|F^-\|^2}{8\pi^2}. \end{aligned} \tag{4.22}$$

The inequality $(a^2 + b^2) \geqslant |a^2 - b^2|$ shows that, for each k, the absolute minima of S are attained when

$$S = 8\pi^2|k| \tag{4.23}$$

and this corresponds to $F^\mp = 0$ or equivalently

$$F = \mp * F \tag{4.24}$$

and we have the celebrated self-dual and anti-self-dual conditions. Changing the orientation of M has the effect of changing the sign of the $*$ operation and so interchanges F^+ with F^-.

Up to now, although we have not mentioned it, for algebraic convenience we have set the coupling constant of the theory equal to unity. But to understand anything nonperturbative the coupling must be present so we now temporarily cease this practice. Denoting the coupling constant by g (the context should prevent any confusion with magnetic charge) the action S is given by

$$S \equiv S(A) = \frac{1}{g^2}\|F\|^2 = -\frac{1}{g^2}\int_M \mathrm{tr}(F \wedge * F). \tag{4.25}$$

Hence if A is an instanton then we immediately have

$$S(A) = \frac{8\pi^2|k|}{g^2}, \quad k \in \mathbb{Z}. \tag{4.26}$$

Finally the corresponding quantum mechanical amplitude is $\exp[-S]$ so that we have

$$\exp\left[-S(A)\right] = \exp\left[-8\pi^2|k|/g^2\right] \tag{4.27}$$

which we see at once is an *inverse* power series in g^2; moreover, topology is uppermost for we note that for this inverse power series to exist the instanton number k must be nonzero.

4.7. *The mathematicians take a strong interest*

The pace of instanton research increased towards the end of the 1970's due in part to a keen interest being taken in the problems by some highly able and gifted mathematicians. As we shall see below this attack on the problems by two distinct groups was to prove highly beneficial to both physics *and* mathematics. In fact some particularly choice fruits of these labours fell into the garden of the mathematicians.

Thus far we have stressed the topological nature of the connections of Yang–Mills theory: the relevant mathematical structure is a fibre bundle and together with this comes cohomological characteristic class data giving discrete numerical invariants such as the instanton number k. However, for instantons, there remains a more prosaic object to study namely the nonlinear partial differential equation for the instanton A itself, i.e. the self-duality equation

$$F = *F \tag{4.28}$$

or, more explicitly,

$$\begin{aligned}&\partial_\mu A^a_\nu - \partial_\nu A^a_\mu + \mathrm{i}gf^{abc}A^b_\mu A^c_\nu \\ &\quad = \frac{1}{2}\varepsilon_{\mu\nu\alpha\beta}\left(\partial^\alpha A^{a\beta} - \partial^\beta A^{a\alpha} + \mathrm{i}gf^{abc}A^{b\alpha}A^{c\beta}\right).\end{aligned} \tag{4.29}$$

A key change of viewpoint on the self-duality equation changed the focus away from differential equations; this was the breakthrough made by Ward [1977].

Ward showed that the solution of the self-duality partial differential equation was equivalent to the construction of an appropriate vector bundle. His paper [1977] gives a brief summary at the beginning

> In this note we describe briefly how the information of self-dual gauge fields may be "coded" into the structure of certain complex vector bundles, and how the information may be extracted, yielding a procedure by which (at least in principle) all self-dual solutions of the Yang–Mills equations may be generated. The construction arose as part of the programme of twistor theory [3]; it is the Yang–Mills analogue of Penrose's "nonlinear graviton" construction [4], which relates to self-dual solutions of Einstein's vacuum equations.

This discarding of the differential equation and its encoding into the transition functions of certain vector bundles immediately made the problem of more interest and accessibility to mathematicians. Atiyah and Ward [1977] showed how the problem was equivalent to one in algebraic geometry; there then followed a complete solution to the problem for $M = S^4$ by Atiyah, Drinfeld, Hitchin and Manin [1978], the situation for other four manifolds is treated in [Atiyah, Hitchin and Singer, 1978].

Atiyah, who was to become a key figure in many subsequent developments of joint interest to mathematicians and physicists describes his introduction to Yang–Mills theories as follows (taken from the preface to Atiyah [1979])

> My acquaintance with the geometry of Yang–Mills equations arose from lectures given in Oxford in Autumn 1976 by I.M. Singer, and I am very grateful to him for arousing my interest in this aspect of theoretical physics.

It was not long before a large body of both mathematicians and physicists were working on a large selection of problems related in some way to Yang–Mills theories. The next breakthrough was in mathematics rather than in physics and we turn to this in the section that follows.

5. The Yang–Mills equations and four manifold theory

5.1. *Donaldson's work*

In the 1980's interest in instantons continued strongly but there was a most striking result proved by Donaldson [1983] which used the Yang–Mills instantons to make a fundamental advance in the topology of four manifolds.

Donaldson's result concerned simply connected compact closed four manifolds M. We shall now give a short account of some of the result's main features so that the reader may be better able to appreciate its significance.

In topology one distinguishes three types of manifold M: topological, piecewise-linear and differentiable (or smooth) which we can denote when necessary by M_{TOP}, M_{PL} and M_{DIFF}, respectively. There are topological obstacles to the existence of PL and DIFF structures on a given topological manifold M. The nature of these obstacles is quite well understood in dimension 5 and higher but, in dimension 4, the situation is quite different and much more difficult to comprehend. It is for this dimension that Yang–Mills theories and Donaldson's work have made such an important contribution.

On the subject of the importance of Yang–Mills theories for obtaining these results Donaldson and Kronheimer [1990] (p. 27) have said the following in favour of Yang–Mills theory.

> These geometrical techniques will then be applied to obtain the differential–topological results mentioned above. It is precisely this departure from standard techniques which has led to the new results, and at present there is no way known to produce results such as these which does not rely on Yang–Mills theory.

5.2. *Donaldson and simply connected four manifolds*

We consider here compact closed four manifolds M. For a simply connected four manifold M, $H_1(M;\mathbb{Z})$ and $H_3(M;\mathbb{Z})$ vanish and the nontrivial homological information is concentrated in the middle dimension in $H_2(M;\mathbb{Z})$. A central object then is the *intersection form* defined by

$$q(\alpha, \beta) = (\alpha \cup \beta)[M], \quad \alpha, \beta \in H_2(M; \mathbb{Z}) \tag{5.1}$$

with $\cup$ denoting cup product so that $(\alpha \cup \beta)[M]$ denotes the integer obtained by evaluating $\alpha \cup \beta$ on the generating cycle $[M]$ of $H_4(M; \mathbb{Z})$ on M. Poincaré duality implies that the intersection form is always *nondegenerate* over $\mathbb{Z}$ and so has $\det q = \mp 1$ – q is then called unimodular. Also we refer to q, as *even* if all its diagonal entries are even, and as *odd* otherwise. A very powerful result of Freedman [1982] can now be called on – the intersection form q very nearly determines the homeomorphism class of a simply connected M, and actually only fails to do so in the odd case where there are still just two possibilities. Further *every* unimodular quadratic form occurs as the intersection form of some manifold.

The relevant theorem is

THEOREM (Freedman [1982]). *A simply connected* 4-*manifold M with even intersection form q belongs to a unique homeomorphism class, while if q is odd there are precisely two nonhomeomorphic M with q as their intersection form.*

An illustration of the impressive nature of Freedman's work is readily available. Recollect that the Poincaré conjecture in four dimensions is the statement that any homotopy 4-sphere, S_h^4 say, is actually *homeomorphic* to the standard sphere S^4. Now S^4 has trivial cohomology in two dimensions so its intersection form q is the zero quadratic form which we write as $q = \emptyset$. But S_h, having the same homotopy type as S^4, has the same cohomology as S^4. So any S_h^4 also has intersection form $q = \emptyset$. But Freedman's result says that for a simply connected M with even q there is only *one homeomorphism class* for M, therefore S_h^4 *homeomorphic* to S^4 and we have established the conjecture. Incidentally this means that the Poincaré conjecture has now been proved for all n except $n = 3$ – the case originally proposed by Poincaré.

Now we come to Donaldson's work which concerns smoothability of four manifolds; one should also note that, when q is a definite quadratic form, a choice of orientation can always render q *positive* definite. Then we have the following theorem

THEOREM (Donaldson [1983]). *A simply connected, smooth* 4-*manifold, with positive definite intersection form q has the property that q is always diagonalisable over the integers to* $q = \mathrm{diag}(1, \ldots, 1)$.

Immediately one can go on to deduce that no simply connected, 4-manifold for which q is even and positive definite can be smoothed! For example, the Cartan matrix for the exceptional Lie algebra e_8 is given by

$$E_8 = \begin{pmatrix} 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 2 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 2 & -1 & 0 & -1 \\ 0 & 0 & 0 & 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 2 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 2 \end{pmatrix}. \tag{5.2}$$

Freedman's result guarantees that there is a manifold M with intersection form $q = E_8 \oplus E_8$. However, Donaldson's theorem forbids such a manifold[28] from existing smoothly. Before Donaldson's work surgery techniques had been extensively used to try to construct smoothly the manifold with intersection form $E_8 \oplus E_8$. We can now see that these techniques were destined to fail.

In fact, in contrast to Freedman's theorem, which allows *all* unimodular quadratic forms to occur as the intersection form of some topological manifold, Donaldson's theorem says that in the positive definite, smooth, case only *one* quadratic form is allowed, namely the identity I.

One of the most striking aspects of Donaldson's work is that his proof uses the Yang–Mills equations. We can only outline what is involved here, for more details cf. Donaldson and Kronheimer [1990], Freed and Uhlenbeck [1984] and Nash [1991].

In brief then let A be a connection on a principal $SU(2)$-bundle over a simply connected 4-manifold M with positive definite intersection form. If S is the usual Euclidean Yang–Mills action S of (4.11) one has

$$S = \|F\|^2 = -\int_M \mathrm{tr}(F \wedge *F). \tag{5.3}$$

Now given one instanton A which minimises S one can perturb about A in an attempt to find more instantons. When this is done the space of all instantons can be fitted together to form a global *moduli space* of finite dimension. For the instanton with $k = 1$ which provides the absolute minimum of S, this moduli space $\mathcal{M}_1$, say, is a noncompact space of dimension 5, with singularities.

We can now summarise the logic that is used to prove Donaldson's theorem: there are very strong relationships between M and the moduli space $\mathcal{M}_1$; for example, let q be regarded as an $n \times n$ matrix with precisely p unit eigenvalues (clearly $p \leqslant n$ and Donaldson's theorem is just the statement that $p = n$), then $\mathcal{M}_1$ has precisely p singularities which look like cones on the space $\mathbf{C}P^2$. These combine to produce the result that the 4-manifold M has the same topological signature $\mathit{Sign}\,(M)$ as p copies of $\mathbf{C}P^2$; now p copies of $\mathbf{C}P^2$ have signature $a - b$ where a of the $\mathbf{C}P^2$'s are oriented in the usual fashion and b are given the opposite orientation. Thus we have

$$\mathit{Sign}\,(M) = a - b. \tag{5.4}$$

Now the definition of $\mathit{Sign}\,(M)$ is that it is the signature $\sigma(q)$ of the intersection form q of M. But since, by assumption, q is positive definite $n \times n$ then $\sigma(q) = n = \mathit{Sign}\,(M)$. So we can write

$$n = a - b. \tag{5.5}$$

[28] The reader may wonder why we did not discuss the four manifold with the simpler intersection form $q = E_8$. This manifold of course exists. It is not smoothable but this fact is due to a much older result of Rohlin [1952] concerning smoothability and the signature of q. Rohlin's theorem only provides a necessary condition for smoothability, this is that the signature of an even q must be divisible by 16. The lack of sufficiency of this condition is shown by the example of $q = E_8 \oplus E_8$ since one can verify that the signature of q in this case is divisible by 16.

However, $a+b=p$ and $p \leqslant n$ so we can assemble this information in the form

$$n = a - b, \qquad p = a + b \leqslant n, \tag{5.6}$$

but one always has $a+b \geqslant a-b$ so now we have

$$n \leqslant p \leqslant n \Rightarrow p = n \tag{5.7}$$

and we have obtained Donaldson's theorem.

6. Physics and knots revisited – the Jones polynomial

6.1. *Three manifolds and Floer, Jones and Witten*

In Section 1 we discussed knots in our material on the nineteenth century. It is now time to return to this subject.

Jones [1985] made a great step forward in knot theory by introducing a new polynomial invariant of knots (and links), now known as the *Jones polynomial* and denoted by $V_L(t)$, where L denotes the knot or link and t is a real variable. Knot invariants of this kind had proved hard to find: the original one was that of Alexander [1928], denoted by $\Delta_L(t)$.

The Jones polynomial originates in certain finite dimensional von Neumann algebras which Jones denotes by A_n. A point of physical interest here is that, as Jones observed in his paper, D. Evans pointed out that some representations of these A_n's had already been constructed in the physics literature in statistical mechanics, the relevant reference (which Jones gives) being [Lieb and Temperley, 1971]. The statistical mechanics concerns the Potts and ice-type models, cf. Baxter [1982]. This leads one to speculate that the combinatorial structure of some models in statistical mechanics has a *topological origin*; this does seem to be borne out by subsequent work.

The Jones polynomial proved powerful enough to decide many of the longstanding Tait conjectures on knots which we referred to in Section 1. The next event of joint topological and physical interest was a result by Witten [1989a, 1989b] which gave a completely new formulation (and generalisation) of the Jones polynomial in terms of a certain kind of quantum field theory, nowadays known as a *topological quantum field theory.*

A vital ingredient in this whole story is the work of Floer [1988a, 1988b] on a new homology invariant of three manifolds constructed from considerations of gauge theory and instantons. We shall meet this work again in Section 8. We mention it now because of its influence on subsequent work. For the moment we just need to inform the reader that Floer considers the critical point behaviour of the function f where f depends on an $SU(2)$ connection A: f is simply the Chern–Simons function obtained by integrating the Chern–Simons secondary characteristic class over a closed three manifold M. If $\mathcal{A}$ denotes the space of all connections A, we have

$$\begin{aligned} &f : \mathcal{A} \longrightarrow \mathbb{R}, \quad A \longmapsto f(A), \quad \text{with} \\ &f(A) = -\frac{1}{8\pi^2} \int_M \operatorname{tr}\left(A \wedge dA + \frac{2}{3} A \wedge A \wedge A\right). \end{aligned} \tag{6.1}$$

Then from a very clever study of the Morse theory of this function f, whose domain is the *infinite dimensional* space $\mathcal{A}$, Floer obtains new homology groups $HF(M)$ known as *Floer* homology groups associated to the three manifold M.

In 1987 Atiyah [1988] speculated that there was a relation between Floer's work and the Jones polynomial: Towards the end of this paper we find the following

> Finally let me list a few of the major problems that are still outstanding in the area .... More speculatively, I would like to end with 4) Find a connection with the link invariants of Vaughan Jones [11].
>
> As circumstantial evidence that this is reasonable I will list some properties shared by Floer homology and the Jones polynomial.
>
> (i) both are subtle 3-dimensional invariants,
> (ii) they are sensitive to orientation of 3-space (unlike the Alexander polynomial),
> (iii) they depend on Lie groups: $SU(2)$ in the first instance but capable of generalisation,
> (iv) there are 2-dimensional schemes for computing these 3-dimensional invariants,
> (v) whereas the variable in the Alexander polynomial corresponds to $\pi_1(S^1)$, the variable in the Jones polynomial appears to be related to $\pi_3(S^3)$, the origin of "instanton numbers",
> (vi) both have deep connections with physics, specifically quantum field theory (and statistical mechanics).

In 1988 Witten [1989a] rose to this challenge and found the relation that Atiyah had suspected existed. The content of Witten [1989b] is described by its author with a certain amount of understatement. He says

> In a lecture at the Hermann Weyl Symposium last year [1], Michael Atiyah proposed two problems for quantum field theorists. The first problem was to give a physical interpretation for Donaldson theory. The second problem was to find an intrinsically three dimensional definition of the Jones polynomial of knot theory. I would like to give a flavour of these two problems.

Our next task is to have a look at the methods that Witten used.

6.2. *Topological quantum field theories*

A topological quantum field theory (also called simply a *topological field theory*) is one which, at first sight, may seem trivial physically: it has an action with no metric dependence. The absence of a metric means that there are no distance measurements or forces and so no conventional dynamics. The Hamiltonian $\mathcal{H}$ of the theory has only zero eigenstates and the Hilbert space of the theory is usually finite dimensional. The theory can, however, be nontrivial: its nontriviality is reflected in the existence of *tunnelling* between vacua.

The particular action chosen by Witten for obtaining the Jones polynomial was the well known Chern–Simons action S given by

$$S = \frac{ik}{4\pi g^2} \int_M \operatorname{tr}\left(A \wedge dA + \frac{2}{3} A \wedge A \wedge A\right), \quad k \in \mathbb{Z}, \tag{6.2}$$

where A is an $SU(2)$ connection or gauge field and M is a closed, compact three dimensional manifold. The partition function for this quantum field theory is $Z(M)$ where

$$Z(M) = \int \mathcal{D}\mathcal{A} \exp\left[-\frac{ik}{4\pi g^2}\int_M \mathrm{tr}\left(A\wedge dA + \frac{2}{3}A\wedge A\wedge A\right)\right]. \tag{6.3}$$

This partition function itself is an invariant – the *Witten invariant* – of the three manifold M; however, at present, we want to study knots: knots enter in the following way. Consider a closed curve C embedded in M so as to form a knot, K, say. Now ones takes the connection A, parallel transports it around C and constructs the holonomy operator $PT(C)$ as described above in the discussion of the Aharonov–Bohm effect in Section 2. This time we have a non-Abelian connection and to obtain a gauge invariant operator we must take the trace of $PT(C)$ giving what is called a *Wilson line*; we denote this by $W(R, C)$ where

$$W(R, C) = \mathrm{tr}\, P \exp\left[\int_C A\right] \tag{6.4}$$

and R denotes the particular representation carried by A. There is a natural correlation function associated with this knot namely

$$\begin{aligned}\langle W(R, C)\rangle = \frac{1}{Z(M)}&\int \mathcal{D}\mathcal{A}\, W(R, C)\\ &\times \exp\left[-\frac{ik}{4\pi g^2}\int_M \mathrm{tr}\left(A\wedge dA + \frac{2}{3}A\wedge A\wedge A\right)\right]\end{aligned} \tag{6.5}$$

and Witten [1989a, 1989b] shows that this determines the Jones polynomial $V_K(t)$ of the knot.

Further if one has not *one* curve C but a number of them, say $C_1, \ldots, C_p$ then one has a p component link L, say, whose Jones polynomial $V_L(t)$ is determined by a multiple correlation function of p Wilson lines given by

$$\begin{aligned}\langle W(R_1, C_1)\cdots W(R_p, C_p)\rangle &= \frac{1}{Z(M)}\int \mathcal{D}\mathcal{A} W(R_1, C_1)\cdots W(R_p, C_p)\\ \times \exp&\left[-\frac{ik}{4\pi g^2}\int_M \mathrm{tr}\left(A\wedge dA + \frac{2}{3}A\wedge A\wedge A\right)\right].\end{aligned} \tag{6.6}$$

A nice thing that happens if we step backwards slightly to the *Abelian* case is that one can recover the Gauss linking number: In the Abelian case A is just a $U(1)$ connection and $S(A)$ becomes only quadratic in A giving (we have set $g = 1$)

$$S(A) = \frac{ik}{4\pi}\int_M \mathrm{tr}(A\wedge dA), \tag{6.7}$$

so that

$$\begin{aligned}\langle W(R_1, C_1)\cdots W(R_p, C_p)\rangle &= \frac{1}{Z(M)}\int \mathcal{D}\mathcal{A}\, W(R_1, C_1)\cdots W(R_p, C_p)\\ \times \exp&\left[-\frac{ik}{4\pi}\int_M \mathrm{tr}(A\wedge dA)\right].\end{aligned} \tag{6.8}$$

The quadratic action, together with the exponential dependence on A of a Wilson line, allows the entire integrand to be written as a Gaussian after completing the square. The calculation of the functional integral rests just on the calculation of a Green's function which, for $M = S^3$ (to which we now specialise), is an elementary computation. The result is that

$$\begin{aligned}&\langle W(n_1, C_1)\cdots W(n_p, C_p)\rangle\\&\quad = \exp\left[\frac{\mathrm{i}}{4k}\varepsilon_{ijk}\sum_{l,m=1}^{p} n_l n_m \int_{C_l} dx^i \int_{C_m} dy^j \frac{(x-y)^k}{|x-y|^3}\right]\end{aligned} \tag{6.9}$$

with x^i and y^j local coordinates on the knots C_l and C_m.

We easily recognise the basic integral

$$\frac{\varepsilon_{ijk}}{4\pi}\int_{C_l} dx^i \int_{C_m} dy^j \frac{(x-y)^k}{|x-y|^3} \tag{6.10}$$

in (6.9) as the linking number of Gauss that we met in Section 1; and we note that we have met it again in a physical context. Incidentally for another physical context in which the linking number appears cf. Wilczek and Zee [1983]; in this paper a connection is made between the spin-statistics properties of particles and topology.[29]

In the non-Abelian case we can also obtain a quadratic functional integral by studying the limit of small coupling g, or, completely equivalently, the limit of large k. For any topological field theory such a limit has considerable significance. When this limit is evaluated for this theory one obtains another differential topological invariant: the Ray–Singer torsion of the connection A on M, cf. Witten [1989a, 1989b].

Numerous topological field theories are now studied in the current literature, we shall meet another one in Section 7; indeed the whole notion of a topological quantum field theory has been axiomatised in [Atiyah, 1989].

Finally we point out some of the new and more general features of the Witten approach to the Jones polynomial: Witten's definition is intrinsically three dimensional and not dependent on any two-dimensional arguments for its validation. The group $SU(2)$ of the connection A is not obligatory – it can be replaced by another Lie group G, say $G = SU(N)$; for the appropriate representation of $SU(N)$ this gives rise to a two variable generalisation of the Jones polynomial cf. Freyd et al. [1985]. There is an immediate generalisation to knots in *any* three manifold M rather than the classical case of knots in S^3 (i.e. compactified $\mathbb{R}^3$). Invariants for three manifolds themselves immediately arise and so the theory is not really just one of knots (i.e. embeddings).

7. Yang–Mills and four manifolds once more

7.1. *Donaldson again: polynomial invariants for four manifolds*

In the 1990's more progress was made in four dimensions with another result of Donaldson [1990]; actually some of these results were announced considerably earlier in 1986 by

[29] On this latter topic there is more work. For some examples cf. Balachandran et al. [1993], Berry and Robbins [1997], Finkelstein and Rubinstein [1968], Mickelsson [1984], Tscheuschner [1989], and references therein.

Donaldson in his Field's medal address (cf. Donaldson [1987]).

In Section 5 we described Donaldson's use of the moduli space $\mathcal{M}_1$ to derive smoothability results about 4-manifolds. The space $\mathcal{M}_1$ only contains instantons with instanton number k equal to one. In addition to this, by using all values of k there exist moduli spaces $\mathcal{M}_k$, $k = 1, 2, \ldots,$ for instantons of any instanton number k. Donaldson's new invariants use all of the $\mathcal{M}_k$ and in the process one obtains powerful differential topological invariants of simply connected 4-manifolds. Donaldson [1990] begins with

> The traditional methods of geometric topology have not produced a clear picture of the classification of smooth 4-manifolds. This gap has been partially bridged by methods using Yang–Mills theory or gauge theory. Riemannian manifolds carry with them an array of moduli spaces – finite dimensional spaces of connections cut out by the first order Yang–Mills equations. These equations depend on the Riemannian geometry of the 4-manifold, but at the level of homology we find properties of the moduli spaces which do not change when the metric is changed continuously. Any two metrics can be joined by a path, so by default, these properties depend only on the differential topology of the 4-manifold, and furnish a mine of potential new differential topological invariants.

The "mine of potential new differential topological invariants", as Donaldson modestly puts it, is a reference to his new polynomial invariants. He goes on, in the same paper, to say

> Here we use infinite families of moduli spaces to define infinite numbers of invariants for simply connected manifolds with b_2^+ odd[30] and greater than 1. These invariants are distinguished elements in the ring:
>
> $$S^*(H^2(X))$$
>
> of polynomials on the cohomology of the underlying 4-manifold X.
>
> Equivalently, they can be viewed as symmetric multi-linear functions:
>
> $$q : H_2(X; \mathbb{Z}) \times \cdots \times H_2(X; \mathbb{Z}) \longrightarrow \mathbb{Z}$$
>
> ... Certainly one of the most striking facts is that we get infinitely many invariants for a single manifold. Discovering to what extent these are independent (i.e. whether there are strong universal relations between them) appears to be an interesting target for future research.

We just want to mention some results that have been obtained with the Donaldson invariants which serve to show that they are nontrivial and important. To be able to do this we must introduce some notation.

Let M be a smooth, simply connected, orientable Riemannian four manifold without boundary and A be an $SU(2)$ connection which is anti-self-dual[31] so that

$$F = -*F. \tag{7.1}$$

Then the dimension of the moduli space $\mathcal{M}_k$ is the integer

$$\dim \mathcal{M}_k = 8k - 3(1 + b_2^+). \tag{7.2}$$

---

[30] The number b_2^+ is defined to be the rank of the positive part of the intersection form.

[31] We take anti-self-dual connections rather than self-dual connections so as to follow Donaldson's sign conventions.

A Donaldson invariant $q_d(M)$ is a symmetric integer polynomial of degree d in the 2-homology $H_2(M;\mathbb{Z})$ of M

$$q_d(M): \underbrace{H_2(M) \times \cdots \times H_2(M)}_{d \text{ factors}} \longrightarrow \mathbb{Z}. \tag{7.3}$$

Given a certain map m (cf. Donaldson [1990, 1996] or Nash [1991])

$$m: H_2(M) \to H^2(\mathcal{M}_k) \tag{7.4}$$

we use m to define by $q_d(M)$ by using de Rham cohomology and differential forms. Setting $d = \dim \mathcal{M}_k/2$ we define $q_d(M)$ by

$$\begin{aligned} &q_d(M): H_2(M) \times \cdots \times H_2(M) \longrightarrow \mathbb{Z}, \\ &a_1 \times \cdots \times a_d \longmapsto \int_{\overline{\mathcal{M}_k}} m(a_1) \wedge \cdots \wedge m(a_d), \end{aligned} \tag{7.5}$$

where $\overline{\mathcal{M}_k}$ denotes a compactification of the moduli space. We see that the $q_d(M)$ are symmetric integer valued polynomials of degree d in $H^2(M)$, i.e. $q_d(M) \in Sym^d(H_2(M)) \subset S^*(H(M))$; also, since $d = \dim \mathcal{M}_k/2 = (8k - 3(1 + b_2^+))/2$, we now understand why Donaldson required b_2^+ to be odd.

Now the Donaldson invariants are, *a priori*, not very easy to calculate since they require detailed knowledge of the instanton moduli space. However, if M is a complex algebraic surface, a positivity argument shows that

$$q_d(M) \neq 0, \quad \text{for } d \geqslant d_0 \tag{7.6}$$

with d_0 some integer – in other words the $q_d(M)$ are all nonzero when d is large enough. Conversely, if M can be written as the connected sum

$$M = M_1 \# M_2, \tag{7.7}$$

where M_1 and M_2 both have $b_2^+ > 0$ then

$$q_d(M) = 0, \quad \text{for all } d. \tag{7.8}$$

The $q_d(M)$ are *differential* topological invariants rather than topological invariants; this means that they have the potential to distinguish homeomorphic manifolds which have distinct diffeomorphic structures. An example where the $q_d(M)$ are used to show that two homeomorphic manifolds are not diffeomorphic can be found in [Ebeling, 1990]. A possible a physical context for this result can be found in [Nash, 1992], cf. too, [Libgober and Wood, 1982] for some earlier work related to [Ebeling, 1990] which was done before the $q_d(M)$ were defined.

The $q_d(M)$ can also be obtained from a topological quantum field theory as we shall now see below.

7.2. *Another topological field theory*

In 1988 Witten showed how to obtain the $q_d(M)$ as correlation functions in a BRST supersymmetric topological field theory. We shall only give a brief statement of facts to give the reader some idea of what sort of action and physical fields are involved; for a full account, cf. Witten [1988].

The action S for the theory is given by

$$S = \int_M d^4x\sqrt{g}\,\mathrm{tr}\left\{\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{4}F^*_{\mu\nu}F^{\mu\nu} + \frac{1}{2}\phi D_\mu D^\mu\lambda + \mathrm{i}D_\mu\psi_\nu\chi^{\mu\nu} - \mathrm{i}\eta D_\mu\psi^\mu \right.$$
$$\left. - \frac{\mathrm{i}}{8}\phi[\chi_{\mu\nu},\chi^{\mu\nu}] - \frac{\mathrm{i}}{2}\lambda[\psi_\mu,\psi^\mu] - \frac{\mathrm{i}}{2}\phi[\eta,\eta] - \frac{1}{8}[\phi,\lambda]^2\right\}, \tag{7.9}$$

where $F_{\mu\nu}$ is the curvature of a connection A_μ and $(\phi, \lambda, \eta, \psi_\mu, \chi_{\mu\nu})$ are a collection of fields introduced in order to construct the right supersymmetric theory; ϕ and λ are both spinless while the multiplet $(\psi_\mu, \chi_{\mu\nu})$ contains the components of a 0-form, a 1-form and a self-dual 2-form, respectively. The significance of this choice of multiplet is that the anti-instanton version of the instanton deformation complex used to calculate $\dim \mathcal{M}_k$ contains precisely these fields. Even though S contains a metric its correlation functions are independent of the metric g so that S can still be regarded as a topological field theory. This can be shown to follow from the fact that both S and its associated energy momentum tensor $T \equiv (\delta S/\delta g)$ can be written as BRST commutators $S = \{Q, V\}$, $T = \{Q, V'\}$ for suitable V and V' – cf. Witten [1988].

With this theory it is possible to show that the correlation functions are independent of the gauge coupling and hence we can evaluate them in a small coupling limit. In this limit the functional integrals are dominated by the classical minima of S, which for A_μ are just the instantons

$$F_{\mu\nu} = -F^*_{\mu\nu}. \tag{7.10}$$

We also need ϕ and λ to vanish for irreducible connections. If we expand all the fields around the minima up to quadratic terms and do the resulting Gaussian integrals, the correlation functions may be formally evaluated. Let us consider a correlation function

$$\langle P\rangle = \int \mathcal{D}\mathcal{F}\exp[-S]\,P(\mathcal{F}), \tag{7.11}$$

where $\mathcal{F}$ denotes the collection of fields present in S and $P(\mathcal{F})$ is a polynomial in the fields. Now S has been constructed so that the zero modes in the expansion about the minima are the tangents to the moduli space $\mathcal{M}_k$; thus, if the $\mathcal{D}\mathcal{F}$ integration is expressed as an integral over modes, all the nonzero modes may be integrated out first leaving a *finite dimensional* integration over $\dim \mathcal{M}_k$. The Gaussian integration over the nonzero modes is a Boson–Fermion ratio of determinants, a ratio which supersymmetry constrains to be ∓ 1 since Bosonic and Fermionic eigenvalues are equal in pairs. This amounts to expressing $\langle P\rangle$ as

$$\langle P\rangle = \int_{\mathcal{M}_k} P_n, \tag{7.12}$$

where P_n is an n-form over $\mathcal{M}_k$ and $n = \dim \mathcal{M}_k$. If the original polynomial $P(\mathcal{F})$ is chosen in the correct way then calculation of $\langle P \rangle$ reproduces evaluation of the Donaldson polynomials.

The next breakthrough in the topology of four manifolds came from physics and was due to Seiberg and Witten [1994a] (cf. Seiberg and Witten [1989a, 1989b] and Witten [1994]) and it is the next topic to which we turn.

7.3. *Physics again: Seiberg–Witten theory and four manifolds*

In [Donaldson, 1996] we find the most upbeat introduction to a review article on the Seiberg–Witten equations; it gives some idea of the excitement and the power of the methods associated with this latest breakthrough.

> Since 1982 the use of gauge theory, in the shape of the Yang–Mills instanton equations, has permeated research in 4-manifold topology. At first this use of differential geometry and differential equations had an unexpected and unorthodox flavour, but over the years the ideas have become more familiar; a body of techniques has built up through the efforts of many mathematicians, producing results which have uncovered some of the mysteries of 4-manifold theory, and leading to substantial internal conundrums within the field itself. In the last three months of 1994 a remarkable thing happened: this research was turned on its head by the introduction of a new kind of differential-geometric equation by Seiberg and Witten: in the space of a few weeks long-standing problems were solved, new and unexpected results were found, along with simpler new proofs of existing ones, and new vistas for research opened up. This article is a report on some of these developments, which are due to various mathematicians, notably Kronheimer, Mrowka, Morgan, Stern and Taubes, building on the seminal work of Seiberg [S] and Seiberg and Witten [SW].

We shall say a little about both the physics and the mathematics relating to the Seiberg–Witten equations; however, we shall make the remarks about the mathematics here but leave the remarks about the physics until Section 9 where they fit in more naturally.

Seiberg and Witten's work allows one to produce another physical theory, in addition to [Witten, 1988], with which to compute Donaldson invariants. In [Witten, 1988], as just described above, Donaldson theory is obtained from a twisted $N = 2$ supersymmetric Yang–Mills theory. Seiberg and Witten produce a duality which amounts to an equivalence between the *strong coupling* limit of this $N = 2$ theory and the *weak coupling* limit of a theory of *Abelian* monopoles. This latter theory is much easier to compute with leading (on the mathematical side) to the advances described in [Donaldson, 1996] and [Witten, 1994].

If we choose an oriented, compact, closed, Riemannian manifold M then the data we need for the Seiberg–Witten equations are a connection A on a line bundle L over M and a "local spinor" field ψ. The Seiberg–Witten equations are then

$$\eth\psi = 0, \qquad F_A^+ = -\frac{1}{2}\overline{\psi}\Gamma\psi, \tag{7.13}$$

where $\eth$ is the Dirac operator and Γ is made from the gamma matrices Γ_i according to $\Gamma = \frac{1}{2}[\Gamma_i, \Gamma_j]\, dx^i \wedge dx^j$. We call ψ a local spinor because global spinors may not exist on M; however, orientability guarantees that a $spin_c$ structure does exist and ψ is the appropriate section for this $spin_c$ structure. We note that A is just a $U(1)$ Abelian connection and so $F = dA$, with F^+ just being the self-dual part of F.

We shall now have a brief look at one example of a new result using the Seiberg–Witten equations. The equations clearly provide the absolute minima for the action

$$S = \int_M \left\{ |\eth\psi|^2 + \frac{1}{2}\left|F^+ + \frac{1}{2}\overline{\psi}\Gamma\psi\right|^2 \right\}. \tag{7.14}$$

If we use a Weitzenböck formula to relate the Laplacian $\nabla_A^* \nabla_A$ to $\eth^*\eth$ plus curvature terms we find that S satisfies

$$\begin{aligned} &\int_M \left\{ |\eth\psi|^2 + \frac{1}{2}\left|F^+ + \frac{1}{2}\overline{\psi}\Gamma\psi\right|^2 \right\} \\ &= \int_M \left\{ |\nabla_A\psi|^2 + \frac{1}{2}|F^+|^2 + \frac{1}{8}|\psi|^4 + \frac{1}{4}R|\psi|^2 \right\} \\ &= \int_M \left\{ |\nabla_A\psi|^2 + \frac{1}{4}|F|^2 + \frac{1}{8}|\psi|^4 + \frac{1}{4}R|\psi|^2 \right\} + \pi^2 c_1^2(L), \end{aligned} \tag{7.15}$$

where R is the scalar curvature of M. The action now looks like one for monopoles – indeed in [Witten, 1994], Witten refers to (7.13) as "the monopole equations". But now suppose that R is *positive* and that the pair (A, ψ) is a solution to the Seiberg–Witten equations: then the LHS is zero and all the integrands on the RHS are positive so the solution must obey $\psi = 0$ and $F^+ = 0$. It turns out that if M has $b_2^+ > 1$ then a perturbation of the metric can preserve the positivity of R but change $F^+ = 0$ to be plain $F = 0$ rendering the connection A flat. Hence, in these circumstances, the solution (A, ψ) is the trivial one. This means that we have a new kind of vanishing theorem in four dimensions.

THEOREM (Witten [1994]). *No four manifold with $b_2^+ > 1$ and nontrivial Seiberg–Witten invariants admits a metric of positive scalar curvature.*

We referred just now to the Seiberg–Witten invariants and unfortunately we cannot define them here. However, we do want to say that they are rational numbers a_i and there are formulae relating the Donaldson polynomial invariants q_d to the a_i.

Many more new results have been found involving, for example, symplectic and Kähler manifolds, cf. Donaldson [1996]; the story, however, is clearly not at all finished.

8. Dynamics and topology since Poincaré

8.1. *Dynamical systems and Morse theory*

In this section we want to return to Poincaré and consider that part of his topological legacy which sprang from his work on dynamics. We shall only be able to look at two areas and these are the theory of *dynamical systems* and *Morse theory*. This is, of necessity, somewhat selective, nevertheless these two subjects do represent mainstream developments which descend directly from Poincaré's work on dynamics and topology.[32] It should also be borne in mind that there is a large overlap between the two subjects.

It is still true that the n-body problem attracts much attention from mathematicians, including those using topological techniques. A few references of interest here are Smale [1970a, 1970b] and Saari and Xia [1996].

8.2. *Dynamical systems*

Poincaré's pioneering work on celestial mechanics prepared the way for the present day subject of dynamical systems with Birkhoff as the actual founder. In this subject one studies an immense diversity of sophisticated mathematical problems usually no longer connected with celestial or Newtonian mechanics.

A very rough idea of what is involved goes as follows: Recall first that the celestial mechanics of n bodies has a motion that is described by a set of differential equations together with their initial data. One then varies the initial data and asks how the motion changes.

Now the modern mathematical setting is to view the orbits of the n bodies as integral curves for their associated differential equations. Then one regards the *qualitative study* of the orbits as being a study of the *global geometry* of the space of integral curves as their initial conditions vary smoothly. Integral curves $\gamma(t)$ are associated with vector fields $V(t)$ via the differential equation

$$\frac{d\gamma(t)}{dt} = V\big(\gamma(t)\big). \tag{8.1}$$

Hence one is now studying the vastly more general subject of the global geometry of the space of flows of a vector field V on a manifold M.

Two notions play a distinguished part in the theory of dynamical systems: closed integral curves and singular points. It is natural to regard two flows on M as *equivalent* if there is a homeomorphism of M which takes one flow into the other; one can also insist that this homeomorphism is smooth, i.e. a diffeomorphism. Finally an equivalence class of flows in the homeomorphic sense is a *topological dynamical system*, and one in the diffeomorphic sense is *a smooth, or differentiable, dynamical system.*

As we explained in Section 1 Birkhoff proved Poincaré's geometric theorem in 1913; a subsequent piece of work of great importance and influence was Birkhoff's proof of what is called his *ergodic theorem* in 1931, cf. Birkhoff [1931].

The subsequent blossoming of ergodic theory can be dated from this time. Ergodic theory originates largely in nineteenth century studies in the kinetic theory of gases. However it has now been axiomatised, expanded, refined and reformulated so that it has links with many parts of mathematics as well as retaining some with physics.

Some dynamical systems exhibit ergodic behaviour, a notable class of examples being provided by *geodesic flow* on surfaces of constant negative curvature. This involves too the study of the flows by a discrete encoding known as symbolic dynamics, use of one dimensional interval maps cf. Bedford et al. [1991]. Classical and quantum chaos, and the distinction between the two, are also studied in this context.

[32] Some more detailed historical material, of relevance here, is that of [Dahan-Dalmédico, 1994, 1996].

A vast body of the theory of dynamical systems concerns *Hamiltonian systems*. These of course have their origin in ordinary dynamics but exist now in a much wider context. To have a Hamiltonian system M must be even dimensional, possess a Hamiltonian function

$$H : M \longrightarrow \mathbb{R} \tag{8.2}$$

and have a closed nondegenerate symplectic form 2-form ω appropriately related to H. The perturbation theory of Hamiltonian systems underwent an enormous development in the 1950's and 1960's with the work particularly of Kolmogorov, Arnold and Moser and the creation of what is known as KAM theory (cf. Broer et al. [1995]).

Gradient dynamical systems were used by Thom [1969, 1971, 1972] in his work on what is now called *Catastrophe theory*. Thom took the system

$$\frac{d\gamma(t)}{dt} = \operatorname{grad} V\big(\gamma(t)\big), \tag{8.3}$$

where V is a potential function. Thom classified the possible critical points of V into seven types known as the seven elementary catastrophes; he then proposed to use these dynamical systems as models for the behaviour of a large class of physical, chemical and biological systems. In many cases the models are not at all adequate, nevertheless, there are some successes. However, the seminal nature of Thom's work is clear though as it is the beginning of the classification theory for singularities. In this connection there are the two results of Arnold [1973, 1978] which closely relate the classification of singularities to the Weyl groups of the various compact simple Lie groups.

8.3. *Morse theory: the topology of critical points*

The aim in Morse theory is to study the relation between critical points and topology. More specifically one extracts topological information from a study of the critical points of a smooth real valued function

$$f : M \longrightarrow \mathbb{R}, \tag{8.4}$$

where M is a compact manifold usually without boundary. For a suitably behaved class of functions f there exists quite a tight relationship between the number and type of critical points of f and topological invariants of M such as the Euler–Poincaré characteristic, the Betti numbers and other cohomological data. This relationship can then be used in two ways: one can take certain special functions whose critical points are easy to find and use this information to derive results about the topology of M; on the other hand, if the topology of M is well understood, one can use this topology to infer the existence of critical points of f in cases where f is too complex, or too abstractly defined, to allow a direct calculation.

Taking a function f the equation for its critical points is

$$df = 0. \tag{8.5}$$

We *assume* that all the critical points p of f are nondegenerate; this means that the Hessian matrix Hf of second derivatives is invertible at p, or

$$\det Hf(p) \neq 0 \quad \text{where } Hf(p) = \left[\partial^2 f/\partial x^i \partial x^j|_p\right]_{n\times n}. \tag{8.6}$$

Each critical point p has an index λ_p which is defined to be the number of *negative* eigenvalues of $Hf(p)$. We can then associate to the function f and its critical points p the Morse series $M_t(f)$ defined by

$$M_t(f) = \sum_{\text{all } p} t^{\lambda_p} = \sum_i m_i t^i. \tag{8.7}$$

The topology of M now enters via $P_t(M)$: the Poincaré series of M. We have

$$P_t(M) = \sum_{i=0}^{n} \dim H^i(M;\mathbb{R})t^i = \sum_{i=0}^{n} b_i t^i. \tag{8.8}$$

The fundamental result of Morse theory is the statement that

$$M_t(f) \geqslant P_t(M) \tag{8.9}$$

from which so many things follow, to mention just one simple example

$$m_i \geqslant b_i \tag{8.10}$$

showing that the number of critical points of index i is bounded below by the Betti number b_i.

Successful applications of Morse theory in mathematics are impressive and widespread; a few notable examples are the proof by Morse [1934] that there exist infinitely many geodesics joining a pair of points on a sphere S^n endowed with any Riemannian metric, Bott's [1956, 1959] proof of his celebrated periodicity theorems on the homotopy of Lie groups, Milnor's construction [1956] of the first exotic spheres, and the proof by Smale [1961] of the Poincaré conjecture for $\dim M \geqslant 5$.

8.4. *Supersymmetric quantum mechanics and Morse theory*

Witten [1982] constructed a quantum mechanical point of view on Morse theory which has proved very influential. It also provides a point of departure for the Floer theory discussed below. In summary Witten gives a quantum mechanical proof of the Morse inequalities on M; however an important extra feature is that the cohomology of M is also explicitly constructed.

Witten takes as Hamiltonian $\mathcal{H}$: the Hodge Laplacian on forms, i.e. one has

$$\mathcal{H} = \bigoplus_{p\geqslant 0} \Delta_p \equiv \bigoplus_{p\geqslant 0} \left(dd^* + d^*d\right)_p. \tag{8.11}$$

The Bosons and Fermions of the supersymmetry are the spaces H^B and H^F formed by even and odd forms while the supersymmetry algebra is generated by two operators Q_1 and Q_2 which are constructed from d and d^*. The appropriate definitions are

$$\begin{aligned} Q_1 &= (d + d^*), & Q_2 &= \mathrm{i}(d - d^*), \\ H^B &= \bigoplus_{p \geqslant 0} \Omega^{2p}(M), & H^F &= \bigoplus_{p \geqslant 0} \Omega^{2p+1}(M). \end{aligned} \tag{8.12}$$

A Morse function f is now incorporated into the model without changing the supersymmetry algebra by replacing d by d_t where

$$d_t = e^{-ft} d e^{ft}. \tag{8.13}$$

It is a routine matter to verify that this conjugation of d by e^{ft} leaves the algebra unchanged. The proof of the Morse inequalities rests on an analysis of the spectrum of the associated Hamiltonian, which is now

$$\mathcal{H}_t = d_t d_t^* + d_t^* d_t = \bigoplus_{p \geqslant 0} \Delta_p(t). \tag{8.14}$$

One needs additional physics to carry out the rest of the work. It turns out that this all comes from the consideration of quantum mechanical tunnelling between critical points of f. One considers grad f as a vector field on M and then studies the integral curves of this vector field, that is the solutions $\gamma(t)$ of the differential equation

$$\frac{d\gamma(s)}{ds} = -\mathrm{grad}\, f(\gamma(s)). \tag{8.15}$$

We can give no more details here but, as has been emphasised by Witten [1982], these ideas are applicable in quantum field theory as well as in quantum mechanics. In that case one has to deal with functions in infinite dimensions and it was not long before a significant result along these lines emerged; this was the work of Floer [1988a, 1988b] which we now examine.

8.5. *Floer homology and Morse theory*

In Section 6, cf. (6.1), we referred to Floer's work and his Morse theoretic study of the function

$$\begin{aligned} &f : \mathcal{A} \longrightarrow \mathbb{R}, \quad A \longmapsto f(A) \quad \text{with} \\ &f(A) = -\frac{1}{8\pi^2} \int_M \mathrm{tr}\left(A \wedge dA + \frac{2}{3} A \wedge A \wedge A\right). \end{aligned} \tag{8.16}$$

We review now some of the details.

The critical points of f are given by

$$df(A) = 0, \tag{8.17}$$

where the exterior derivative is now taken to be acting in the space $\mathcal{A}$ of $SU(2)$ connections on M. If A is such a critical point then we can write $A_t = A + ta$ and obtain

$$f(A_t) = f(A) - \frac{t}{4\pi^2} \int_M \operatorname{tr}\big(F(A) \wedge a\big) + \cdots. \tag{8.18}$$

Hence we can conclude that

$$df(A) = -\frac{F(A)}{4\pi^2} \tag{8.19}$$

and so the critical points of the Chern–Simons function are the *flat connections* on M.

As long as $\pi_1(M) \neq 0$ then flat connections on M are not trivial, since they can have nonzero holonomy round a nontrivial loop on M. The holonomy of each flat connection is an $SU(2)$ element parametrised by a loop on M; in this way it defines a representation of $\pi_1(M)$ in $SU(2)$ and the space of *inequivalent* such representations is the quotient

$$\operatorname{Hom}\big(\pi_1(M), SU(2)\big)/Ad\, SU(2). \tag{8.20}$$

Having found a critical point Morse theory requires us to calculate its index and so we must also calculate the Hessian of f: the snag is that this gives an operator which is unbounded from below rendering the index formally infinite. This is not entirely unexpected since we are working in infinite dimensions.

Floer gets round this very cleverly by realising that he only needs a *relative index* which he can compute via spectral flow and the Atiyah–Singer index theorem. He takes two critical points A_P and A_Q in $\mathcal{A}$ and joins then with a steepest descent path $A(t)$; i.e. a path which obeys the equation

$$\frac{dA(t)}{dt} = -\operatorname{grad} f\big(A(t)\big) \tag{8.21}$$

with grad denoting the gradient operator on the space $\mathcal{A}$. The consequence of all this for the Morse theory construction is that he is able to construct a homology complex and associated homology groups $HF_p(M)$. However, the topology of the situation dictates that the relative Morse index of f is only well defined mod 8. This means that $HF_p(M)$ are graded mod 8 and one only obtains eight homology groups: $HF_p(M),\ p = 0, \ldots 7$; for more details cf. Nash [1991].

Morse theory has also been successfully applied to other problems in Yang–Mills theory. Some important papers are [Atiyah and Bott, 1982] on Yang–Mills theories on Riemann surfaces where an equivariant Morse theory was required, and [Taubes, 1985, 1988] on pure Yang–Mills theory and Yang–Mills theory for monopoles. In all these examples one has to grapple with the infinite dimensionality of the quantum field theory.

8.6. *Knots again*

Vasil'ev [1990a, 1990b] has developed an approach to knot theory using singularity theory. Vasil'ev constructs a huge new class of knot invariants and we shall now give a sketch of what is involved.

A knot is a smooth embedding of a circle into $\mathbb{R}^3$. So a knot gives a map

$$f : S^1 \longrightarrow \mathbb{R}^3 \tag{8.22}$$

so that f belongs to the space $\mathcal{F}$ where $\mathcal{F} = \mathrm{Map}(S^1, \mathbb{R}^3)$. Not all elements of $\mathcal{F}$ give knots since a knot map f is not allowed to self-intersect or be singular. Let Σ be the subspace of $\mathcal{F}$ which contains either self-intersecting or singular maps, then the subspace of knots is the *complement*

$$\mathcal{F} - \Sigma. \tag{8.23}$$

Now any element of Σ can be made smooth by a simple one parameter deformation, hence Σ is a *hypersurface* in $\mathcal{F}$ and is known as the *discriminant*. As the discriminant Σ wanders through $\mathcal{F}$ it skirts along the edge of the complement $\mathcal{F} - \Sigma$ and divides it into many different connected components. Clearly knots in the same connected component can be deformed into each other and so are equivalent (or isotopic).

Now any knot *invariant* is, by the previous sentence, a function which is *constant* on each connected component of $\mathcal{F} - \Sigma$. Hence the task of constructing all (numerical) knot invariants is the same as finding all functions on $\mathcal{F} - \Sigma$ which are constant on each connected component. But topology tells us at once that this is just the 0-cohomology of $\mathcal{F} - \Sigma$. In other words

$$H^0(\mathcal{F} - \Sigma) = \text{The space of knot invariants.} \tag{8.24}$$

Vasil'ev provides a method for computing most, and possibly all, of $H^0(\mathcal{F} - \Sigma)$. There are connections, too, to physics cf. Bar-Natan [1996].

9. Strings, mirrors and duals

9.1. *String theory, supersymmetry and unification of interactions*

String theory has, by now, a fairly long history very little of which we can mention here. We shall, in the main, limit ourselves to remarks which relate in some way to topology.

Topology enters string theory at the outset because a moving string sweeps out a two-dimensional surface and, in quantum theory, all such surfaces must be summed over whatever their topology. This leads to the Polyakov expression for the partition function Z of the Bosonic string which contains a sum over the genera p of Riemann surfaces Σ. One has

$$Z = \sum_{\text{genera}} \int \mathcal{D}g\mathcal{D}\phi \exp\left[-\frac{1}{2}\int_\Sigma \langle\partial\phi, \partial\phi\rangle_g\right] = \sum_{p=0}^{\infty} Z_p, \tag{9.1}$$

where the functional integral is over all metrics g on Σ and the string's position ϕ. We have $\phi \equiv \phi^{\mu}(x^1, x|^2)$, $\mu = 1, \ldots, d$, so the ϕ^{μ} can be thought of as specifying an embedding of Σ in a d-dimensional space-time M. It is well known that the theory is only conformally invariant when $d = 26$: the critical dimension; however, if one includes Fermions then this critical dimension changes to $d = 10$.

A string theory is also a two-dimensional conformal field theory and this latter subject is very important for mathematics as well as physics. It has been axiomatised in a very fruitful and influential way by Segal [1989]. It involves important representation theory of infinite-dimensional groups such as $LG = \text{Map}(S^1, G)$ and $\text{Diff}(S^1)$. We regret that we have been unable to trace its history in this article because of lack of space; its importance is immediately apparent when one reads the literature on string theory as well as that of many statistical mechanical models, it is also a key notion used in calculation and conceptual work in string theory. Finally, closely connected to conformal field theories, are the subjects of Kac–Moody algebras, vertex operator algebras and quantum groups; these all have close connections with physics but, although they have topological aspects, their algebraic properties are more prominent and this, as well as considerations of space, is another reason why we have had to omit them from this essay. Conformal field theory is also a vital ingredient in the surgery argument used in [Witten, 1989a, 1989b] to compute the Jones polynomial and, viewed from this standpoint, conformal field theories can be seen to provide a link between the infinite-dimensional representational theory just mentioned and the topology of two and three manifolds.

String theory came into its own with the incorporation of supersymmetry in the early 1980's (cf. Green, Schwarz and Witten [1979, 1989a]).

Quantum field theories with chiral Fermions sometimes exhibit a pathological behaviour when coupled to gauge fields, gauge invariance may break down, this is referred to as an anomaly. Such anomalies, too, have played a major part in shaping present state of string theory; a key paper here is that of Green and Schwarz [1984] who discovered a remarkable anomaly cancellation mechanism which thereby singles out five distinguished supersymmetric $d = 10$ string theories. Anomalies also have an important topological aspect involving the Atiyah–Singer index theorem for families of Dirac operators. We have not had space to discuss this here, cf. Nash [1991] for more details.

The low energy limit of a string theory is meant to be a conventional quantum field theory which should be a theory that describes all known interactions including gravity. Supersymmetric string theories do succeed in including gravity and so offer the best chance so far for a quantum theory of gravity as well as, perhaps, for an eventual unified theory of all interactions.

The five string theories singled out by the work of Green and Schwarz [1984] are all supersymmetric and contain Yang–Mills fields. These five 10-dimensional theories are denoted by type I, type IIA, type IIB, $E_8 \times E_8$ heterotic and $SO(32)$ heterotic.

Only four of the ten dimensions of space-time M are directly observable; so the remaining six are meant to form a small (i.e. small compared with the string scale) compact six-dimensional space M_6, say. The favoured physical choice for M_6 is that it be a three complex dimensional *Calabi–Yau* manifold – this means that it is complex manifold of a special kind: it is Kähler with holonomy group contained in $SU(3)$. The favoured status of Calabi–Yau manifolds has to do with what is called *mirror symmetry* which we now briefly review as it is of considerable interest to both mathematicians (for example, algebraic geometers) and physicists.

9.2. *Mirror symmetry*

Mirror symmetry refers to the property that Calabi–Yau manifolds come in dual pairs which were conjectured to give equivalent string theories. These dual pairs are then called *mirror manifolds*. Mirror symmetry can also be profitably thought of as providing a *transform*. In other words a difficult problem on one Calabi–Yau manifold may be much easier, but equivalent to, one on its dual. We give an example of a result obtainable this way below; it concerns the number of curves of prescribed degree and genus on a Calabi–Yau manifold.

The term *mirror manifolds* was coined in [Greene and Plesser, 1990]. This paper also contains the first, and as yet only, known construction of such dual pairs.

If M and N are Calabi–Yau manifolds of complex dimension n ($n = 3$ in the string theory cited above), and if $h^{(p,q)}(M)$ denote the Hodge numbers[33] of a complex manifold M, then mirror symmetric pairs satisfy

$$h^{(p,q)}(M) = h^{(n-p,q)}(N). \tag{9.2}$$

The term mirror symmetry originates in the fact that this represents a reflection symmetry about the diagonal in the Hodge diamond formed by the $h^{(p,q)}$'s. This reflection property of the Hodge numbers is not sufficient to ensure that M and N are mirror manifolds, one must prove that the associated conformal field theories are also identical, cf. Greene and Plesser [1990, 1992] for more information.

The introduction (by Greene, Vafa and Warner [1989]) of manifolds which are complete intersections in *weighted* projective spaces is an important part of this story, cf. too, Candelas, Lynker and Schimmrigk [1990]; previously complete intersections had been studied in ordinary projective space, cf., for example, Green, Hübsch and Lütken [1989] and references therein.

A bold prediction of mirror symmetry (following from the equality of the three point functions on a Calabi–Yau M and its mirror) concerned the numbers n of curves of genus g and degree d on a Calabi–Yau manifold; these numbers could be read off from a rather sophisticated instanton calculation, cf. Candelas, Ossa et al. [1991] and the articles in [Yau, 1992] for more details. The startling nature of these predictions can be partly appreciated by browsing Table 1 (this is half of a table appearing in [Bershadsky et al., 1994]). The particular Calabi–Yau manifold is three dimensional and is a certain quotient by $\mathbb{Z}_5 \times \mathbb{Z}_5 \times \mathbb{Z}_5$ of a quintic hypersurface in $\mathbf{CP}^4$ whose equation in homogeneous coordinates is

$$z_1^5 + z_2^5 + \cdots + z_5^5 = 0. \tag{9.3}$$

Mathematical verifications of these spectacular, physically obtained, numbers were at first only available for small values of the degree. However, there has now been a beautiful confirmation of all of them by Givental [1996]; this work also extends the ideas to non-Calabi–Yau manifolds; a key paper, whose results are used in [Givental, 1996], is that of Kontsevich [1995].

[33] Hodge numbers are the dimensions of the various Dolbeault cohomology groups of M, i.e. $h^{(p,q)}(M) = \dim H_{\bar\partial}^{(p,q)}(M) = \dim H^q(M; \Omega^p)$ where Ω^p is the sheaf of holomorphic p-forms on M.

Table 1
Table showing numbers of curves of genus g on a quintic hypersurface as predicted by mirror symmetry

Degree	$g = 0$	$g = 1$
$n = 0$	5	50/12
$n = 1$	2875	0
$n = 2$	609250	0
$n = 3$	317206375	609250
$n = 4$	242467530000	3721431625
$n = 5$	229305888887625	12129909700200
$n = 6$	248249742118022000	31147299732677250
$n = 7$	295091050570845659250	71578406022880761750
$n = 8$	375632160937476603550000	154990541752957846986500
$n = 9$	503840510416985243645106250	324064464310279585656399500
⋮	⋮	⋮
large n	$a_0 n^{-3}(\log n)^{-2} e^{2\pi n\alpha}$	$a_1 n^{-1} e^{2\pi n\alpha}$

The solution methods employed for these problems involving curve counting, or *enumerative geometry*, are closely connected with another development of joint physical and mathematical interest: this is the subject of *quantum cohomology*. Quantum cohomology originates in quantum field theory. One considers a quantum cohomology ring $H_q^*(M)$ which is a natural deformation of the standard cohomology ring $H^*(M)$ of a manifold M. The q in $H_q^*(M)$ is a real parameter which can be taken to zero and, when this is done, one recovers the standard cohomology ring $H^*(M)$; thus $q \to 0$ represents the *classical limit*, one also recognises similarities with the deformations of Lie algebras known as quantum groups. For some physical and mathematical background material on this highly interesting new area cf. Morrison and Plesser [1995] and Kontsevich and Manin [1994] and references therein.

9.3. *Dyons again and the tyranny of dualities*

Several new kinds of duality emerged from about 1994 onwards. Their origin can be traced back to the subject of dyons and in particular to a paper of Montonen and Olive [1977]; on the other hand the present work in the subject is due in great part to the papers of Seiberg and Witten [1979, 1989a] and the insights they offer into strong coupling problems such as the celebrated conundrum of the mechanism for quark confinement.

Montonen and Olive proposed a duality between a theory of magnetic monopoles and one containing gauge fields. They were motivated in part by a semiclassical analysis of dyons and gauge fields. Let g denote the coupling of the gauge field[34] then a dyon with (magnetic, electric) quantum numbers (n, m) has mass M given by

$$M^2 = V^2\left(n^2 + \frac{16\pi^2}{g^4}m^2\right), \tag{9.4}$$

[34] Previously we used g to denote magnetic charge, since this conflicts with our present choice we shall go back to Dirac's notation and use μ for magnetic charge.

where V is the Higgs vacuum expectation value. If we interchange g with $1/g$, n with m, and V and $4\pi V/g^2$ then M is invariant; there is also an exchange of the gauge group G with $\widehat{G}$: a group with weight lattice dual to that of G.

Montonen and Olive astutely observed this invariance and boldly conjectured, with accompanying reasons, that it led to a duality possessed by a full quantum theory of dyons. They proposed that the two quantum field theories passed between by the interchanges above were really dual: i.e. it was only necessary to calculate one of them to obtain full knowledge of the other. This is a very dramatic and attractive conjecture because the interchange of g with $1/g$ is an exchange of strong and weak coupling: i.e. the (intractable) strong coupling limit of one theory ought to be calculable as the (tractable) weak coupling limit of a theory with appropriately altered spectrum and quantum numbers. Osborn [1979] found that an $N = 4$ supersymmetric $SU(2)$ gauge theory was the best candidate for which the conjecture might hold; nevertheless Seiberg and Witten's results are for an $N = 2$ theory.

It transpires that the electric and magnetic charges e, μ and the CP breaking angle θ live more naturally together as the single complex variable

$$e + \mathrm{i}\mu = e_0(n + \tau), \quad \text{where } \tau = \frac{\theta}{2\pi} + \frac{4\pi \mathrm{i}}{g^2}. \tag{9.5}$$

A general point on the dyon lattice of Figure 3 is now given by

$$e_0(m\tau + n), \quad \text{where } m, n \in \mathbb{Z}. \tag{9.6}$$

This addition of the angle θ allows the interchange (or $\mathbb{Z}_2$) symmetry of Montonen and Olive to be promoted to a full $SL(2, \mathbb{Z})$ symmetry under which

$$\tau \mapsto \frac{a\tau + b}{c\tau + d}, \quad \text{with } \begin{cases} a, b, c, d \in \mathbb{Z}, \\ ad - bc = 1. \end{cases} \tag{9.7}$$

Now the conventional string viewpoint of the physics is that this $SL(2, \mathbb{Z})$ symmetry, and the mirror symmetry, are only a low energy manifestation of a richer symmetry of the full string theory. There is now considerable successful work in this direction which involves string dualities known by the symbols S, T and U. Two theories which are S dual have the property that their weak and strong coupling limits are equivalent, two theories which are T dual have the property that one compactified on a large volume is equivalent to the other compactified on a small volume, finally U duality corresponds to a theory compactified on a large, or small, volume being equivalent to another at strong, or weak, coupling, respectively.

Certain pairs of the five basic superstring theories are thought to be dual to one another in this framework. It is conjectured that these five theories are just different manifestations of a single eleven-dimensional theory known as M theory.

9.4. *Black hole postscript*

It has also recently become of interest to study what happens when strings have ends that move on p-dimensional membranes (called Dirichlet branes or D-branes because of the boundary condition imposed at their ends). This has made possible a study of black holes in this string setting. Included in this study is the ability to calculate some of their quantum properties and their entropy (Strominger and Vafa [1996]). The topological nature of space-time here makes contact with the noncommutative geometry of Connes [1994].

10. Concluding remarks

In the twentieth century, some time after the early successes of the new quantum mechanics and relativity, quantum field theory encountered difficult mathematical problems. Efforts to solve these problems led to the birth of the subject of axiomatic quantum field theory. In this approach the main idea was to tackle the formidable problems of quantum field theory head on using the most powerful mathematical tools available; the bulk of these tools being drawn from analysis.

It is now evident that the way forward in these problems is considerably illuminated if, in addition to analysis, one uses differential topology. We have also seen that this inclusion of topology has produced profound results in mathematics as well as physics.

Not since Poincaré, Hilbert and Weyl took an interest in physics has such lavish attention been visited on the physicists by the mathematicians. The story this time is rather different: In the early part of the twentieth century the physicists imported Riemannian geometry for relativity, thereby of course accelerating its rise to be an essential pillar of the body of mathematics; Hilbert spaces were duly digested for quantum mechanics as was the notion of symmetry in its widest possible sense leading to a systematic use of the theory of group representations.

The Riemannian geometry used by physicists in relativity was first used implicitly in a local manner; but it was inevitable that global issues would arise eventually. This, of course, entails topology and so more mathematics has to be learned by the physicist but the singularity theorems of general relativity more than justify the intellectual investment required.

Thus far, then, the gifts, were mainly from the mathematicians to the physicists. For the last quarter of the twentieth century things are rather different, the physicists have been able to give as well as to receive. The Yang–Mills equations have been the source of many new results in three- and four-dimensional differential geometry and topology.

In sum, the growing interaction between topology and physics has been a very healthy thing for both subjects. Their joint futures look very bright. It seems fitting that we should leave the last word to Dirac [1931]

> It seems likely that this process of increasing abstraction will continue in the future and that advance in physics is to be associated with a continual modification and generalisation of the axioms at the base of the mathematics rather than with a logical development of any one mathematical scheme on a fixed foundation.

Bibliography

Aharonov, Y. and Bohm, D. (1959), *Significance of electromagnetic potentials in quantum theory*, Phys. Rev. **115**, 485–491.

Alexander, J.W. (1928), *Topological invariants of knots and links*, Trans. Amer. Math. Soc. **30**, 275–306.

Arnold, V.I. (1973), *Normal forms of function of functions close to degenerate critical points. The Weyl groups* A_k, D_k, E_k *and Lagrangian singularities*, Funct. Anal. Appl. **6**, 254–272.

Arnold, V.I. (1978), *Critical points of functions on manifolds with boundary, the simple Lie groups* B_k, C_k *and* F_4, Russ. Math. Surv. **33**, 99–116.

Atiyah, M.F. (1979), *Geometry of Yang–Mills Fields*, Accademia Nazionale dei Lincei.

Atiyah, M.F. (1988), *New Invariants of 3 and 4 Dimensional Manifolds*, Symposium on the Mathematical Heritage of Hermann Weyl, May 1987, R.O. Wells, ed., Amer. Math. Soc.

Atiyah, M.F. (1989), *Topological quantum field theories*, Inst. Hautes Études Sci. Publ. Math. **68**, 175–186.

Atiyah, M.F. (1990), *The Geometry and Physics of Knots*, Cambridge Univ. Press.

Atiyah, M.F. and Bott, R. (1982), *The Yang–Mills equations over Riemann surfaces*, Phil. Trans. Roy. Soc. London A **308**, 523–615.

Atiyah, M.F., Hitchin, N.J., Drinfeld, V.G. and Manin, Y.I. (1978), *Construction of instantons*, Phys. Lett. **65A**, 185–187.

Atiyah, M.F., Hitchin, N.J. and Singer, I.M. (1978), *Self-duality in four dimensional Riemannian geometry*, Proc. Roy. Soc. London A **362**, 425–461.

Atiyah, M.F. and Ward, R.S. (1977), *Instantons and algebraic geometry*, Commun. Math. Phys. **55**, 117–124.

Balachandran, A.P., Daughton, A., Gu, Z.-C., Sorkin, R.D., Marmo, G. and Srivastava, A.M. (1993), *Spin-statistics theorems without relativity or field theory*, Int. J. Mod. Phys. **A8**, 2993–3044.

Bar-Natan, D. (1996), *Vassiliev and quantum invariants of braids. The interface of knots and physics*, Proc. Sympos. Appl. Math. **51**, 129–144.

Barrow-Green, J. (1997), *Poincaré and the Three Body Problem*, Amer. Math. Soc.

Baxter, R.J. (1982), *Exactly Solved Models in Statistical Mechanics*, Academic Press, New York.

Bedford, T., Keane, M. and Series, C. (eds) (1991), *Ergodic Theory, Symbolic Dynamics and Hyperbolic Spaces*, Oxford Univ. Press.

Belavin, A.A., Polyakov, A.M., Schwarz, A.S. and Tyupkin, Y.S. (1975), *Pseudoparticle solutions of the Yang–Mills equations*, Phys. Lett. **59B**, 85–87.

Berry, M.V. (1984), *Quantal phase factors accompanying adiabatic changes*, Proc. Roy. Soc. London A **392**, 45–57.

Berry, M.V. and Robbins, J.M. (1977), *Indistinguishability for quantum particles: spin, statistics and the geometric phase*, Proc. Roy. Soc. London A **453**, 1771–1790.

Bershadsky, M., Cecotti, S., Ooguri, H. and Vafa, C. (1994), *Kodaira–Spencer theory of gravity and exact results for quantum string amplitudes*, Commun. Math. Phys. **165**, 311–428.

Birkhoff, G.D. (1913), *Proof of Poincaré's geometric theorem*, Trans. Amer. Math. Soc. **14**, 14–22.

Birkhoff, G.D. (1931), *Proof of the ergodic theorem*, Proc. Nat. Acad. Sci. USA **17**, 656–660.

Bogomolny, E.B. (1976), *Stability of classical solutions*, Sov. J. Nucl. Phys. **24**, 861–870.

Bott, R. (1956), *An application of Morse theory to the topology of Lie groups*, Bull. Soc. Math. France **84**, 251–281.

Bott, R. (1959), *The stable homotopy of the classical groups*, Ann. Math. **70**, 313–337.

Brill, D.R. and Werner, F.G. (1960), *Significance of electromagnetic potentials in the quantum theory in the interpretation of electron fringe interferometer observations*, Phys. Rev. Lett. **4**, 344–347.

Broer, H.W., Hoveijn, Takens, F. and van Gils, S.A. (1995), *Nonlinear Dynamical Systems and Chaos*, Birkhäuser.

Candelas, P., Lynker, M. and Schimmrigk, R. (1990), *Calabi–Yau manifolds in weighted* P_4, Nucl. Phys. **B341**, 383–402.

Candelas, P., de la Ossa, X.C., Green, P.S. and Parkes, L. (1991), *A pair of Calabi–Yau manifolds as an exactly soluble superconformal theory*, Nucl. Phys. **B359**, 21–74.

Chandrasekhar, S. (1931), *The maximum mass of ideal white dwarfs*, Astrophys. J. **74**, 81–82.

Chandrasekhar, S. (1935), *The highly collapsed configurations of a stellar mass*, Mon. Not. R. Astron. Soc. **95**, 207–225.

Coleman, S. (1979), *The uses of instantons*, The Ways of Subnuclear Physics, Erice Summer School 1977, A. Zichichi, ed., Plenum Press.

Connes, A. (1994), *Noncommutative Geometry*, Academic Press, New York.

Dahan-Dalmédico, A. (1994), *La renaissance des systèmes dynamiques aux Etats-Unis après la deuxième guerre mondiale: l'action de Solomon Lefschetz*, Rend. Circ. Mat. Palermo (2) Suppl. No 34, 133–166.

Dahan-Dalmédico, A. (1996), *Le difficile héritage de Henri Poincaré en systèmes dynamiques*; *Henri Poincaré: science et philosophie (Nancy, 1994)*, Publ. Henri-Poincaré-Arch., Akademie-Verlag, Berlin, 13–33.

Dirac, P.A.M. (1931), *Quantised singularities in the electromagnetic field*, Proc. Roy. Soc. London A **133**, 60–72.

Donaldson, S.K. (1983), *An application of gauge theory to four dimensional topology*, J. Diff. Geom. **18**, 279–315.

Donaldson, S.K. (1987), *The Geometry of 4-Manifolds*, Proc. of the International Congress of Mathematicians, Berkeley 1986, A.M. Gleason, ed., Amer. Math. Soc.

Donaldson, S.K. (1990), *Polynomial invariants for smooth four manifolds*, Topology **29**, 257–315.

Donaldson, S.K. (1996), *The Seiberg–Witten equations and 4-manifold topology*, Bull. Amer. Math. Soc. **33**, 45–70.

Donaldson, S.K. and Kronheimer, P.B. (1990), *The Geometry of Four Manifolds*, Oxford Univ. Press.

Ebeling, W. (1990), *An example of two homeomorphic, nondiffeomorphic complete intersection surfaces*, Inventiones Math. **99**, 651–654.

Epple, M. (1995), *Branch points of algebraic functions and the beginnings of modern knot theory*, Historia Math. **22**, 371–401.

Epple, M. (1998), *Orbits of asteroids, a braid and the first link invariant*, Math. Intell. **20**, 45–52.

Euler, L. (1736), *Solutio problematis ad geometriam situs pertinentis*, Commentarii Academiae Scientiarum Imperialis Petropolitanae **8**, 128–140.

Finklestein, D. and Rubinstein, J. (1968), *Connection between spin, statistics and kinks*, J. Math. Phys. **9**, 1762–1779.

Floer, A. (1988a), *A relative Morse index for the symplectic action*, Commun. Pure Appl. Math. **41**, 393–407.

Floer, A. (1988b), *An instanton invariant for 3-manifolds*, Commun. Math. Phys. **118**, 215–240.

Freedman, M.H. (1982), *The topology of 4-dimensional manifolds*, J. Diff. Geom. **17**, 357–453.

Freed, D.S. and Uhlenbeck, K.K. (1984), *Instantons and Four-Manifolds*, Springer.

Freyd, P., Yetter, D., Hoste, J., Lickorish, W.B.R., Millet, K. and Ocneanu, A. (1985), *A new polynomial invariant of knots and links*, Bull. Amer. Math. Soc. **12**, 239–246.

Froissart, M. (1966), *Applications of Algebraic Topology to Physics*, Mathematical Theory of Elementary Particles, R. Goodman and I. Segal, eds, MIT Press.

Gauss, C.F. (1877), *Zur mathematischen theorie der electrodynamischen Wirkungen (1833)*, Werke. Königlichen Gesellschaft der Wissenschaften zu Göttingen **5**, 605.

Givental, A.B. (1996), *Equivariant Gromov–Witten invariants*, Internat. Math. Res. Notices **13**, 613–663.

Green, M.B. and Schwarz, J.H. (1984), *Anomaly cancellations in supersymmetric $D = 10$ gauge theory and superstring theory*, Phys. Lett. **149B**, 117–122.

Green, P., Hübsch, T. and Lütken, C.A. (1989), *All the Hodge numbers for all Calabi–Yau complete intersections*, Class. Quant. Grav. **6**, 105–124.

Green, M.B., Schwarz, J.H. and Witten, E. (1987a), *Superstring Theory*, Vol. 1, Cambridge Univ. Press.

Green, M.B., Schwarz, J.H. and Witten, E. (1987b), *Superstring Theory*, Vol. 2, Cambridge Univ. Press.

Greene, B.R. and Plesser, M.R. (1990), *Duality in Calabi–Yau moduli space*, Nucl. Phys. **B338**, 15–37.

Greene, B.R. and Plesser, M.R. (1992), *An introduction to mirror manifolds*, Essays on Mirror Manifolds, S.-T. Yau, ed., International Press, Hong Kong.

Greene, B.R., Vafa, C. and Warner, N.P. (1989), *Calabi–Yau manifolds and renormalization group flows*, Nucl. Phys. **B324**, 371–390.

Hawking, S.W. and Ellis, G.F.R. (1973), *The Large Scale Structure of Space-Time*, Cambridge Univ. Press.

Hawking, S.W. and Israel, W. (eds) (1987), *Three Hundred Years of Gravitation*, Cambridge Univ. Press.

Hawking, S.W. and Penrose, R. (1970), *The singularities of gravitational collapse and cosmology*, Proc. Roy. Soc. London A **314**, 529–548.

Helmholtz, H.L.F. (1858), *Ueber Integrale der hydrodynamischen gleichungen welche den wirbelbewegungen entsprechen*, Jour. für die reine und ang. Math. **55**, 25–55.

Hwa, R.C. and Teplitz, V.L. (1966), *Homology and Feynman Integrals*, Benjamin.

Jones, V.F.R. (1985), *A polynomial invariant for knots via von Neumann algebras*, Bull. Amer. Math. Soc. **12**, 103–111.

Kirchhoff, G. (1847), *Ueber die Auflösung der Gleichungen, auf welche man bei der Untersuchung der linearen Vertheilung galvanisher Ströme gefürt wird*, Ann. der Physik und Chemie **72**, 497–508.

Kontsevich, M. (1995), *Enumeration of rational curves via torus actions*, Progr. Math. **129**, 335–368. *The Moduli Space of Curves*, Texel Island 1994, Birkhäuser, Boston, MA.

Kontsevich, M. and Manin, Y. (1994), *Gromov–Witten classes, quantum cohomology and enumerative geometry*, Commun. Math. Phys. **164**, 525–562.

Laplace, P.S. (1799), *Proof of the theorem that the attractive force of a heavenly body could be so large that light could not flow out of it*, Geographische Ephemeriden, verfasset von Einer Gesellschaft Gelehrten **I**.

Lemaître, G. (1933), *L'univers en expansion*, Ann. Soc. Sci. (Bruxelles) **A53**, 51–85.

Libgober, A.S. and Wood, J.W. (1982), *Differentiable structures on complete intersections – I*, Topology **21**, 469–482.

Lieb, E.H. and Temperley, H.N.V. (1971), *Relations between the percolation and colouring problem and other graph-theoretical problems associated with regular planar lattices: Some exact results for the percolation problem*, Proc. Roy. Soc. London A **322**, 251.

Listing, J.B. (1847), *Vorstudien zur Topologie*, Göttinger Studien, 811–875.

Listing, J.B. (1861), *Der Census räumlicher Complexe oder Veallgemeinerung des Euler'schen Satzes von den Polyedern*, Abhandlungen der königlichen Gesellschaften zu Göttingen **10**, 97–180.

Maxwell, J.C. (1904a), *A Treatise on Electricity and Magnetism*, Vol. I (1873), Ed. 3 (1904), Oxford Univ. Press.

Maxwell, J.C. (1904b), *A Treatise on Electricity and Magnetism*, Vol II (1873), Ed. 3 (1904), Oxford Univ. Press.

Michell, J. (Rev.) (1784), *On the means of discovering the distance, magnitude, etc., of the fixed stars, in consequence of the diminution of their light, in case such a diminution should be found to take place in any of them, and such other data should be procured from observations, as would be further necessary for that purpose*, Phil. Trans. Roy. Soc. London **74**, 35–57.

Mickelsson, J. (1984), *Geometry of spin and statistics in classical and quantum mechanics*, Phys. Rev. **D3**, 1375–1378.

Milnor, J. (1956), *On manifolds homeomorphic to the 7-sphere*, Ann. Math. **64**, 399–405.

Montonen, C. and Olive, D. (1977), *Magnetic monopoles as gauge particles*, Phys. Lett. **B72**, 117–120.

Morandi, G. (1988), *Quantum Hall Effect. Topological Problems in Condensed Matter Physics*, Bibliopolis.

Morrison, D.R. and Plesser, R.M. (1995), *Summing the instantons: quantum cohomology and mirror symmetry in toric varieties*, Nucl. Phys. **B440**, 279–354.

Morse, M. (1934), *Calculus of Variations in the Large*, Amer. Math. Soc. Colloq. Publ.

Nash, C. (1991), *Differential Topology and Quantum Field Theory*, Academic Press, New York.

Nash, C. (1992), *A comment on Witten's topological Lagrangian*, Mod. Phy. Lett. A **7**, 1953–1958.

Nielsen, H. and Olesen, P. (1973), *Vortex line models for dual strings*, Nucl. Phys. **B61**, 45–61.

Oppenheimer, J.R. and Snyder, H. (1939), *On continued gravitational contraction*, Phys. Rev. **56**, 455–459.

Osborn, H. (1979), *Topological charges for $N = 4$ supersymmetric gauge theories and monopoles of spin* 1, Phys. Lett. **83B**, 321–326.

Osterwalder, K. and Schrader, R. (1973a), *Axioms for Euclidean Green's functions*, Commun. Math. Phys. **31**, 83–112.

Osterwalder, K. and Schrader, R. (1973b), *Constructive Quantum Field Theory*, Lecture Notes in Physics vol. 25, Springer, Berlin.

Osterwalder, K. and Schrader, R. (1975), *Axioms for Euclidean Green's functions 2*, Commun. Math. Phys. **42**, 281–305.

Penrose, R. (1965), *Gravitational collapse and space-time singularities*, Phys. Rev. Lett. **14**, 57–59.

Pham, F. (1967), *Introduction à l'Étude Topologique des Singularitiés de Landau*, Mémorial des Sciences Mathématiques, Fasc. 164, Gauthier-Villars, Paris.

Poincaré, H. (1890), *Sur le problème de trois corps et les équations de la dynamique*, Acta Mathematica **13**, 1–270.

Poincaré, H. (1880), *Sur les courbes définies par une équation différentielle*, C. R. Acad. Sc. **90**, 673–675.

Poincaré, H. (1881), *Mémoire sur les courbes définies par une équation différentielle*, J. de Math. **7**, 375–422.

Poincaré, H. (1882), *Mémoire sur les courbes définies par une équation différentielle*, J. de Math. **8**, 251–296.

Poincaré, H. (1885), *Sur les courbes définies par les équations différentielle*, J. de Math. **1**, 167–244.

Poincaré, H. (1886), *Sur les courbes définies par les équations différentielle*, J. de Math. **2**, 151–217.

Poincaré, H. (1892), *Sur l'analysis situs*, C. R. Acad. Sc. **115**, 633–636.

Poincaré, H. (1895), *Analysis situs*, J. Éc. Polyt. **1**, 1–121.

Poincaré, H. (1899a), *Sur les nombres de Betti*, C. R. Acad. Sc. **128**, 629–630.

Poincaré, H. (1899b), *Complément à l'analysis situs*, Rend. Circ. Matem. Palermo **13**, 285–343.

Poincaré, H. (1900), *Second complément à l'analysis situs*, Proc. London Math. Soc. **32**, 277–308.

Poincaré, H. (1901a), *Sur l'analysis situs*, C. R. Acad. Sc. **133**, 707–709.

Poincaré, H. (1901b), *Sur la connexion des surfaces algébriques*, C. R. Acad. Sc. **133**, 969–973.

Poincaré, H. (1902a), *Sur certaines surfaces algébriques; troisième complément à l'analysis situs*, Bull. Soc. Math. Fr. **30**, 49–70.

Poincaré, H. (1902b), *Sur les cycles des surfaces algébriques; quatrième complément à l'analysis situs*, J. Math. Pures et Appl. **8**, 169–214.

Poincaré, H. (1904), *Cinquième complément à l'analysis situs*, Rend. Circ. Matem. Palermo **18**, 45–110.
Poincaré, H. (1912), *Sur un théorème de géométrie*, Rend. Circ. Matem. Palermo **33**, 375–407.
Poincaré, H. (1921), *Analyse des travaux scientifiques de Henri Poincaré, faite par lui même*, Acta Math. **38**, 1–135.
Polyakov, A.M. (1974), *Spectrum of particles in the quantum field theory*, JETP Lett. **20**, 194–195.
Pont, J.-C. (1974), *La Topologie Algébrique des Origines à Poincaré*, Presses Universitaires de France.
Prasad, M.K. and Sommerfield, C.M. (1975), *Exact classical solution for the 't Hooft monopole and the Julia–Zee dyon*, Phys. Rev. Lett. **35**, 760–762.
Riemann, B. (1857), *Theorie der Abel'schen Functionen*, Jour. für die reine und ang. Math. **54**.
Rohlin, V.A. (1952), *New results in the theory of 4-dimensional manifolds*, Dokl. Akad. Nauk USSR **84**, 221–224.
Saari, D.G. and Xia, Z. (eds) (1996), *Hamiltonian Mechanics and Celestial Mechanics*, Amer. Math. Soc.
Schwarzschild, K. (1916a), *Über das Gravitationsfeld eines Masses nach der Einsteinschen Theorie*, Sitzungsberichte Königlich. Preuss. Akad. Wiss., Physik-Math. Kl. 189–196.
Schwarzschild, K. (1916b), *Über das Gravitationsfeld einer Kugal aus inkompressibler Flüssigkeit nach der Einsteinschen Theorie*, Sitzungberichte Königlich. Preuss. Akad. Wiss., Physik-Math. Kl. 424–434.
Schwinger, J. (1968), *Sources and magnetic charges*, Phys. Rev. **173**, 1536–1544.
Segal, G. (1989), *Two Dimensional Conformal Field Theories and Modular Functors*, I. A. M. P. Congress, Swansea, 1988, I. Davies, B. Simon and A. Truman, eds, Institute of Physics.
Seiberg, N. and Witten, E. (1994a), *Electric-magnetic duality, monopole condensation, and confinement in $N = 2$ supersymmetric Yang–Mills theory*, Nucl. Phys. **B426**, 19–52. (Erratum - ibid. **B430**, 485–486.)
Seiberg, N. and Witten, E. (1994b), *Monopoles, duality and chiral symmetry breaking in $N = 2$ supersymmetric QCD*, Nucl. Phys. **B431**, 484–550.
Simon, B. (1983), *Holonomy, the quantum adiabatic theorem and Berry's phase*, Phys. Rev. Lett. **51**, 2167–2170.
Smale, S. (1961), *Generalised Poincaré's conjecture in dimensions greater than four*, Ann. Math. **74**, 391–406.
Smale, S. (1970a), *Topology and mechanics I*, Inventiones Math. **10**, 305–331.
Smale, S. (1970b), *Topology and mechanics II*, Inventiones Math. **11**, 45–64.
Streater, R.F. (1975), *Outline of axiomatic relativistic quantum field theory*, Rep. Prog. Phys. **38**, 771–846.
Strominger, A. and Vafa, C. (1996), *Microscopic origin of the Bekenstein–Hawking entropy*, Phys. Lett. **B379**, 99–104.
Sundman, K.F. (1907), *Recherches sur le problème des trois corps*, Acta Soc. Sci. Fenn. **34**, 1–43.
Sundman, K.F. (1909), *Nouvelles recherche sur le problème des trois corps*, Acta Soc. Sci. Fenn. **35**, 1–27.
Sundman, K.F. (1912), *Memoire sur le problème des trois corps*, Acta Mathematica **36**, 105–179.
Tait, P.G. (1898), *Scientific Papers*, Cambridge Univ. Press. *On knots* I, II, III, 273–437.
Taubes, C.H. (1985), *Min-Max theory for the Yang–Mills–Higgs equations*, Commun. Math. Phys. **97**, 473–540.
Taubes, C.H. (1988), *A framework for Morse theory for the Yang–Mills functional*, Inventiones Math. **94**, 327–402.
Thom, R. (1969), *Topological models in biology*, Topology **8**, 313–335.
Thom, R. (1971), *Modèles Mathématiques de la Morphogenèse*, Acad. Naz. Lincei, Pisa.
Thom, R. (1972), *Stabilité Structurelle et Morphogenèse*, Benjamin.
Thomson, W.H. (Lord Kelvin) (1867), *On vortex atoms*, Proc. Roy. Soc. Edinburg **34**, 15–24.
Thomson, W.H. (Lord Kelvin) (1869), *On vortex motion*, Trans. Roy. Soc. Edinburg **25**, 217–260.
Thomson, W.H. (Lord Kelvin) (1875), *Vortex statics*, Proc. Roy. Soc. Edin. 1875–76 session.
Thomson, W.H. (Lord Kelvin) (1900), *On the duties of ether for electricity and magnetism*, Phil. Mag. **50**, 305–307.
Thomson, W.H. (Lord Kelvin) (1910), *Mathematical and Physical Papers*, Vol. IV, Cambridge Univ. Press.
't Hooft, G. (1971), *Renormalization of massless Yang–Mills fields*, Nucl. Phys. **B33**, 173–199.
't Hooft, G. (1974), *Magnetic monopoles in unified gauge theories*, Nucl. Phys. **B79**, 276–284.
't Hooft, G. (1976), *Computation of the quantum effects due to a four dimensional pseudoparticle*, Phys. Rev. **D14**, 3432–3450.
't Hooft, G. and Veltman, M. (1972), *Regularization and renormalization of gauge fields*, Nucl. Phys. **B44**, 189–213.
Thorne, K.S. (1994), *Black Holes and Time Warps: Einstein's Outrageous Legacy*, Norton.
Tscheuschner, R.D. (1989), *Topological spin-statistics relation in quantum field theory*, Int. J. Theor. Phys. **28**, 1269–1310.
Vasil'ev, V.A. (1990a), *Topology of complements to discriminants and loop spaces*, Adv. Sov. Math. **1**, 9–21.
Vasil'ev, V.A. (1990b), *Cohomology of knot spaces*, Adv. Sov. Math. **1**, 23–69.

Ward, R.S. (1977), *On self-dual gauge fields*, Phys. Lett. **61A**, 81–82.
Wilczek, F. and Zee, A. (1983), *Linking numbers, spin and statistics of solitons*, Phys. Rev. Lett. **51**, 2250–2252.
Witten, E. (1979), *Dyons of charge $e\theta/2\pi$*, Phys. Lett. **86B**, 283–287.
Witten, E. (1982), *Supersymmetry and Morse theory*, J. Diff. Geom. **17**, 661–692.
Witten, E. (1988), *Topological quantum field theory*, Commun. Math. Phys. **117**, 353–386.
Witten, E. (1989a), *Some Geometrical Applications of Quantum Field Theory*, I. A. M. P. Congress, Swansea, 1988, I. Davies, B. Simon and A. Truman, eds, Institute of Physics.
Witten, E. (1989b), *Quantum field theory and the Jones polynomial*, Commun. Math. Phys. **121**, 351–400.
Witten, E. (1994), *Monopoles and four-manifolds*, Math. Res. Lett. **1**, 769–796.
Yau, S.-T. (ed.) (1992), *Essays on Mirror Manifolds*, International Press, Hong Kong.
Zwanziger, D. (1968a), *Exactly soluble nonrelativistic model of particles with both electric and magnetic charge*, Phys. Rev. **176**, 1480–1488.
Zwanziger, D. (1968b), *Exactly soluble nonrelativistic model of particles with both electric and magnetic charge*, Phys. Rev. **176**, 1480–1488.

This Page Intentionally Left Blank

CHAPTER 13

Singularities

Alan H. Durfee

Department of Mathematics, Statistics and Computer Science, Mount Holyoke College, South Hadley, MA 01075, USA

E-mail: adurfee@mtholyoke.edu

1. Introduction

This article recounts the rather wonderful interaction of topology and singularity theory which began to flower in the 1960's with the work of Hirzebruch, Brieskorn, Milnor and others. This interaction can be traced back to the work of Klein, Lefschetz and Picard, and also to the work of knot theorists at the beginning of this century. It continues to the present day, flourishing and expanding in many directions. However, this is not a survey article, but a history; the events of our time are harder to see in perspective, harder to marshal into coherent order, and their very multitude makes it impossible to recount them all. Hence this interaction is followed forward in only a few directions.[1]

The reader may get a sense of the current state of affairs in singularity theory by browsing in the conference proceedings [32, 48]. The focus of this article is singularities of complex algebraic varieties. Real varieties are omitted. Also omitted from this account is the area of critical points of differentiable functions, work initiated by Thom, Mather, Arnold and others; a survey of this subject can be found in the books [3–5].

When two areas interact, ideas flow in both directions. Ideas from topology have entered singularity theory, where algebraic problems have been understood as topological problems and solved by topological methods. (In fact, often the crudest invariants of an algebraic situation are topological.) Conversely, ideas of singularity theory have traveled in the reverse direction into topology. Algebraic geometry supplies many interesting examples both easily and not so easily understood, and these provide a convenient testing ground for topological theories.

[1] That I have attempted to do this at all is due to the prodding of my conscience and a list suggested by W. Neumann of some recent areas where topology has had an effect on singularity theory. He added, though, that "the task becomes immense ... other people would probably come up with almost disjoint lists". The randomness of my efforts here should be readily apparent, and my apologies to those whose work is not mentioned.

HISTORY OF TOPOLOGY
Edited by I.M. James

2. Knots and singularities of plane curves

In the 1920's and 30's there was much activity in knot theory as the new tools of algebraic topology were being applied; the fundamental group of the knot complement was introduced, as were the Alexander polynomial, branched cyclic covers, the Seifert surface, braids, the quadratic form of a knot, linking invariants, and so forth. Many clearly-written wonderful papers were produced on these subjects.

At the same time in algebraic geometry there was interest in understanding complex algebraic surfaces, in particular by exhibiting them as branched covers of the plane. This method is analogous to the method in one dimension lower of projecting a curve to a line. The discriminant locus in the latter case is a set of points and it is easy to understand the branching. For surfaces the branching is more complicated since the discriminant locus is a curve. (The reader is referred to [16] for a detailed historical account of these interactions.)

A method of examining the branching problem for surfaces was proposed by Wilhelm Wirtinger in Vienna, who gave some seminars on this subject beginning in 1905. He divided branch points into two types: At a smooth point of the discriminant curve, the branching group ("Verzweigungsgruppe") of the surface is cyclic, like that of a curve. These points were called "branch points of type I". Singular points of the discriminant curve were called "branch points of type II". He also worked out a simple example.

The classification and the example were recorded by his student Karl Brauner in the beginning of his paper "On the geometry of functions of two complex variables" [6]. Wirtinger's example is the smooth surface in $\mathbf{C}^3$ given by the equation

$$z^3 - 3zx + 2y = 0.$$

When this is projected to the (x, y)-plane, the discriminant curve is

$$x^3 - y^2 = 0.$$

There is one point in the surface over the origin in the (x, y)-plane, two points over the remaining points of the curve $x^3 - y^2 = 0$, and three points over the rest of the plane.

To understand the type II branching of the surface near the origin, a three-sphere $\mathbf{S}^3_r$ of radius r about the origin in the plane was mapped to real three-space by stereographic projection. The image of the intersection of this three-sphere with the discriminant curve was then exhibited as a trefoil knot Γ (Fig. 1). It sufficed to understand the branching of the surface over $\Gamma \subset \mathbf{S}^3_r$. Let A_i, for $i = 1, 2, 3$, be the branching substitution produced by traveling around the loop labelled A_i in the figure. The A_i must satisfy the (now well-known) Wirtinger relation

$$A_0^{-1} A_1 A_0 A_2^{-1} = 1$$

at the left-hand crossing point of the knot projection in the figure. The only possibility for the permutation of the sheets of the covering is thus $A_0 = (12)$, $A_1 = (23)$ and $A_2 = (13)$. Hence the branching group in the neighborhood of $(0, 0)$ is not cyclic (as it is for plane curves), but the symmetric group on three elements.

Brauner concluded "Wir haben aus obigem erkannt, dass es die topologischen Verhältnisse der Kurve Γ sind, welche dieses merkwürdiger Verhalten der Funktionen in der

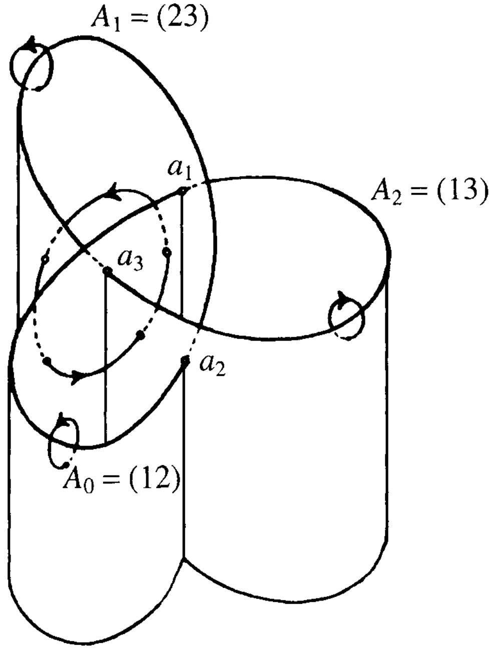


Fig. 1. The discriminant intersected with the sphere ([6], p. 5).

Umgebung der Verzweigungsstellen II. Art bedingen." (We thus have learned that the topological form of the curve Γ determines this remarkable behavior of the function in the neighborhood of a type II branch point.)

There are thus two problems, he said. The first is to determine the topology of the (discriminant) curve in the neighborhood of a singular point, i.e. the knot Γ. The second is to determine the group given by the Wirtinger relations (in modern terminology, the fundamental group of the complement of the knot Γ). These two problems were solved in his paper. He remarked that there are three more problems. The first is to determine the branching group of a function locally in the neighborhood of a point. (This group is of course a quotient of fundamental group of the complement of the knot Γ.) Next, one should determine the global branching group of a function. Finally, given a group, is there a function which has this group as branching group? These problems, he said, would form the subject of two further papers.

He then continued with a systematic study of the links of curve singularities and their fundamental groups. He first looked at the curve

$$ax^n + by^m = 0$$

with the gcd $(n, m) = 1$, parameterizing it by setting $x = \alpha t^m$ and $y = \beta t^n$, where α and β were suitably chosen constants. He wrote the complex number t as $\rho e^{i\phi}$ with ρ and ϕ real and worked out parametric equations for the intersection of the curve with the sphere. Taking its image under the equations for stereographic projection, he observed that the image curve lay on a torus, winding n times in the direction of the meridian and m times in the direction of the equator, and hence was a torus knot.

He then went on to look at two such curves as above and described their linking. He then examined the curve parameterized by $x = t^m(a_m + ta_{m+1} + \cdots)$ and $y = t^n$ and showed that the link is a compound torus knot formed by taking a torus knot on a small tube about the first torus knot and iterating this procedure. He also showed that only a finite number of terms (the characteristic pairs) in the (possibly infinite) power series parameterization of the curve determined the topological type of the knot. He continued by analyzing the case of curves with two branches. Brauner concluded by computing the fundamental group of the complement of these compound torus knots in terms of Wirtinger's generators and relations.

The next work in this area was done by Erich Kähler [24] in Leipzig, who remarks at the beginning of his paper that "Obwohl die betreffenden Fragen zum grössten Teil bereits von Herrn Brauner beantwortet sind, habe ich mir erlaubt den Gegenstand auf dem etwas anschaulicheren Wege ... darzustellen." (Although this question has been for the most part already answered by Mr. Brauner, I have allowed myself to explain it in a somewhat clearer fashion.)

Kähler replaced Brauner's sphere, the boundary of the "round" four-ball $\{|x|^2 + |y|^2 \leqslant r^2\}$ in $\mathbf{C}^2$ by the boundary of the "rectangular" four-ball $\{|x| \leqslant c_1\} \cap \{|y| \leqslant c_2\}$. This is a simplification since a curve tangent to the x-axis (say) intersects this boundary only in $\{|x| \leqslant c_1\} \cap \{|y| = c_2\}$, one of its two sides ($c_1 \ll c_2$). He noted that the two pieces of the boundary could be mapped easily into three-space where they formed a decomposition into two solid tori. He then looked at the curve $y = ax^{m/n}$ and observed that the image of the intersection of this curve with the boundary of the rectangular four-ball is obviously a torus knot or link. He then continued to obtain Brauner's results in easier fashion.

Thus the topological nature of the link could be computed from analytic data. The converse result, that the characteristic pairs could be determined from the topology of the knot, was proved simultaneously by Oscar Zariski at Johns Hopkins University and Werner Burau in Königsberg.

Zariski [61] started with a singular point of the curve X and again derived a presentation of the local fundamental group of its complement. He then found a polynomial invariant $F(t)$ of this group which he later identified as the Alexander polynomial of the knot, and showed that the first Betti number of the k-fold branched cyclic cover of a punctured neighborhood of the origin in $\mathbf{C}^2$ with branch locus X is the number of roots of $F(t)$ which are k-th roots of unity. (This was later recognized to be a purely knot-theoretic result.)

Burau [10], on the other hand, used James Alexander's recent work to compute the Alexander polynomial of compound torus knots. He derived a recursive formula for these polynomials and showed that they were all distinct. He later treated the case when the polynomial had two branches at the origin, i.e. when the link had two components [11].

A survey of the above work was given later by John Reeve [53], who also showed that the intersection number of two branches of a curve at the origin equals the linking number in the three-sphere of their corresponding knots. He gave two proofs. The first, following Lefschetz, notes that the algebraic intersection multiplicity of the curves is their topological intersection multiplicity, which is the linking number of their boundaries. The second proof uses Reidemeister's definition of linking number in terms of the knot projection.

Now let us move forward in time to the present. The computation of knot invariants of the link of a curve singularity becomes increasingly messy as the number of branches of the curve increases. A diagrammatic method for these computations (for the Alexander polynomial,the real Seifert form, the Jordan normal form of the monodromy and so forth) has been developed in [15].

The link of a singularity of a curve has a global analogue, the *link at infinity* K_∞ of a curve $X \subset \mathbf{C}^2$, which is defined to be the intersection of X with a sphere $\mathbf{S}_r^3$ of suitably large radius r. Walter Neumann has shown that if the curve is a regular fiber of its defining equation (i.e. if the map is a locally trivial fibration near this value), then the topological type of the curve is determined by the knot type of $K_\infty \subset \mathbf{S}_r^3$. Also, Neumann and Rudolph have used these techniques to give topological proofs of a result of Abhyankar and Moh (that up to algebraic automorphism, the only embedding of $\mathbf{C}$ in $\mathbf{C}^2$ is the standard one) and similar results of Zaidenberg and Lin [54].

The knot type of the link of a singularity in higher dimensions has received some attention; see for instance [13, 42, 43].

3. Three-manifolds and singularities of surfaces

It is useful at this point to introduce some terminology. An (affine) *algebraic variety* $X \subset \mathbf{C}^m$ is the zero locus of a collection of complex polynomials in m variables. If X is a hypersurface, and hence the zero locus of a single polynomial $f(x_1, x_2, \dots, x_m)$, then a point p is *singular* if $\partial f/\partial x_1 = \cdots = \partial f/\partial x_m = 0$ at p. The set of nonsingular points is a complex manifold of dimension $m - 1$. A point which is not singular is called *smooth.* (The definition of singular point for arbitrary varieties can be found, for example, in [38, Section 2], and similar results hold.)

If $p \in X \subset \mathbf{C}^m$, the *link* of p in X is defined to be

$$K = X \cap \mathbf{S}_\varepsilon^{2m-1},$$

where $\mathbf{S}_\varepsilon^{2m-1}$ is a sphere of sufficiently small radius ε about p in $\mathbf{C}^m$. If p is an isolated singularity of X, then the link is a compact smooth real manifold of dimension one less than the real dimension of X at p. Understanding the topology of the variety X near p is the same as understanding the topology of K and its embedding in the sphere; in fact, X is locally homeomorphic to a cone on K with vertex p [38, 2.20]. (This fact is implicit in the work of Burau and Kähler, but not explicitly stated.) The *local fundamental group* of the singularity is the fundamental group of the link. This is particularly interesting for an isolated singular point of an algebraic surface (complex dimension two) where the link is a three-manifold.

Some time elapsed before the topological investigation of curve singularities chronicled in Section 1 was extended to higher dimensions. In the early 1960's the following result by David Mumford confirmed a conjecture of Abhyankar [40] (see also the Bourbaki talk of Hirzebruch [21]):

THEOREM 3.1. *If p is a normal point of a complex surface X with trivial local fundamental group, then p is a smooth point of X.*

The condition "normal" comes from the algebraic side of algebraic geometry; in particular it implies that the singularity is isolated and that its link is a connected space.

He proved this theorem by resolving the singularity, a technique which in the case of surfaces is old and essentially algorithmic. The process of resolution removes the singular point p from X and replaces it by a collection of smooth transversally-intersecting complex curves $E_1, \ldots, E_r$ so that the new space $\tilde{X}$ is smooth.

He showed that the link could be obtained from the curves E_i by a process called *plumbing*: The tubular neighborhood of E_i in $\tilde{X}$ is identified with a 2-disk bundle over the curve E_i. If E_i and E_j intersect in a point $q \in \tilde{X}$, the two-disk bundles over E_i and E_j are glued together by identifying a fiber over q in one with a disk in the base centered at q in the other. This makes a manifold with corners. If the corners are smoothed (so that the result looks rather like an plumbing elbow joint), the boundary is diffeomorphic to the link.

The *graph* of a resolution of a normal singularity of an algebraic surface is as follows: The i-th vertex corresponds to the curve E_i, labelled by the genus of E_i and the self-intersection $E_i \cdot E_i$. The i-th and j-th vertices are joined by an edge if $E_i \cdot E_j \neq 0$, and the edges are weighted by $E_i \cdot E_j$. The resolution graph thus determines the topological type of the link.

Mumford used Van Kampen's theorem and the plumbing description of the link to give a presentation of the local fundamental group of the singularity and thus prove the theorem.

The local fundamental group of a singularity of an algebraic surface turned out to be a a useful way to classify these singularities. For instance, Brieskorn [9], using earlier work of Prill, showed that if the local fundamental group is finite, then the variety X is locally isomorphic to a quotient $\mathbf{C}^2/G$, where G is one of the well-known finite subgroups of $GL(2, \mathbf{C})$. He listed all such subgroups G together with the resolution graph of the minimal resolution of the corresponding singularity $\mathbf{C}^2/G$.

Philip Wagreich [59], inspired by work of Peter Orlik [46], used Mumford's presentation to find all singularities with nilpotent or solvable local fundamental group. Thus the local fundamental group became closely connected with the local analytic structure of the singularity.

Neumann showed that the topology of the link K determines the graph of the minimal resolution of the singularity. In fact, he showed that $\pi_1(K)$ determines this graph, except in a small number of cases [41] .

Mumford's techniques in a global setting appeared later in work of C.P. Ramanujam [52]:

THEOREM 3.2. *A smooth complex algebraic surface which is contractible and simply connected at infinity is algebraically isomorphic to* $\mathbf{C}^2$.

Ramanujam showed this by compactifying the surface by a divisor with normal crossings, and then using the topological conditions to show that this divisor could be contracted to a projective line. He also showed that the condition of simple connectivity at infinity was essential by producing an example of a smooth affine rational surface X which is contractible but not algebraically isomorphic to $\mathbf{C}^2$. In fact, the intersection of X with a sufficiently large sphere is a homology three-sphere but not a homotopy three-sphere.

Ramanujam's result implies that the only complex algebraic structure on $\mathbf{R}^4$ is the standard one on $\mathbf{C}^2$, so that there are no "exotic" algebraic structures on the complex plane. The search for exotic algebraic structures thus continued in higher dimensions. Ramanujam remarked that the three-fold $X \times \mathbf{C}$ is diffeomorphic to $\mathbf{C}^3$ by the h-cobordism theorem. A cancellation theorem proved later had the corollary that $X \times \mathbf{C}$ is not algebraically isomorphic to $\mathbf{C}^3$. Hence there is an exotic algebraic structure on $\mathbf{C}^3$. Much work followed in this area; for the current state of affairs one can consult [60], for example.

4. Exotic spheres

Egbert Brieskorn, who was spending the academic year 1965–1966 at the Massachusetts Institute of Technology, investigated whether Mumford's theorem extended to higher dimensions. On September 28, 1965, he wrote in a letter to his doctoral advisor Friedrich Hirzebruch that he had examined the three-dimensional variety

$$x_1^2 + x_2^2 + x_3^2 + x_4^3 = 0$$

and its singularity at the origin. He explicitly calculated a resolution of the singular point, then used van Kampen's theorem to show that the link K of this singularity is simply-connected and the Mayer–Vietoris sequence to show that K is a homology 5-sphere. He concluded, using Smale's recent solution of the Poincaré conjecture in higher dimensions, that K is homeomorphic to $\mathbf{S}^5$. Hence Mumford's result did not extend to higher dimensions.

According to Hirzebruch [22, C38], "Dieser Brief von Brieskorn war eine grosse Überraschung" (This letter from Brieskorn was a great surprise). Later letters followed with more squared terms added to the equation above. Brieskorn's final result appeared in [8]: For odd $n \geqslant 3$, the link at the origin of

$$x_0^3 + x_1^2 + x_2^2 + \cdots + x_n^2 = 0 \tag{1}$$

is homeomorphic to the sphere S^{2n-1}.

The attention then shifted to the differentiable structure on this link. To describe the next events, we first need to recall the situation with non-standard or "exotic" differentiable structures on spheres. The first exotic sphere, a differentiable structure on $\mathbf{S}^7$ which is not diffeomorphic to the standard structure, had been discovered only ten years earlier by John Milnor. Further investigations followed by Kervaire and Milnor [27]. By Smale's solution to the higher-dimensional Poincaré conjecture, it was sufficient to look at the set Θ_m of homotopy m-spheres (manifolds homotopy equivalent to the standard sphere S^m). The set Θ_m is an abelian group under connected sum, and Kervaire and Milnor showed that this group is finite ($m \neq 3$).

They also looked at the subgroup $bP_{m+1} \subset \Theta_m$ of homotopy spheres which are boundaries of parallelizable manifolds (manifolds with trivial tangent bundle), and showed that bP_{m+1} is trivial for m even, and finite cyclic for $m \neq 3$ odd.

Its order could be computed as follows: If n is odd, the group bP_{2n} has order one or two. It is generated by the Kervaire sphere which is the boundary of the manifold constructed by plumbing two copies of the tangent disk bundle to S^n. The Kervaire sphere may or may not be diffeomorphic to the standard sphere; the first nontrivial group is bP_{10}. If $\Sigma \in bP_{2n}$ is the boundary of an $(n-1)$-connected parallelizable $2n$-manifold M, whether Σ is diffeomorphic to the standard sphere or the Kervaire sphere depends on the Arf invariant of a geometrically-defined quadratic form on M.

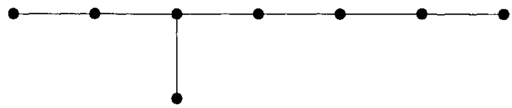


Fig. 2. The E_8 graph.

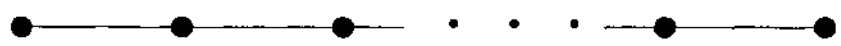


Fig. 3. The A_k graph (k vertices).

If $n \geqslant 4$ is even, the order of bP_{2n} can be calculated in terms of Bernoulli numbers. For example, there are 28 homotopy seven-spheres in $\Theta_7 = bP_8$. Also, the order of $\Sigma \in bP_{2n}$ can be calculated in terms of the signature of the intersection pairing on $H^n(M)$.

The construction of a generator of bP_{2n} for $n \geqslant 4$ even is once again bound up with singularity theory. In a preprint [34] of January, 1959, Milnor had constructed a generator by plumbing according to an even unimodular matrix of rank and index eight. This matrix was not the well-known one associated to the E_8 graph (Fig. 2), though, since its graph had a cycle. He then added a two-handle to make the boundary simply-connected and hence a homotopy sphere. Hirzebruch, however, was familiar with the E_8 matrix from his work on resolution of singularities of surfaces. He constructed a generator of the group bP_{2n} by plumbing copies of the tangent disk bundle to S^n according to the E_8 graph. (For more details, see [22, C30], [21, 20, 35].)

At the same time in the fall of 1965 that Hirzebruch was receiving the letters from Brieskorn, he also received a letter from Klaus Jänich, another of his doctoral students, who was spending the year 1965–1966 at Cornell. Jänich described his work on $(2n-1)$-dimensional $O(n)$-manifolds (manifolds with an action of the orthogonal group). In fact, he had classified $O(n)$-manifolds whose action had just two orbit types with isotropy groups $O(n-1)$ and $O(n-2)$, in particular showing that they were in one-to-one correspondence with the nonnegative integers. (These results were also obtained by W.C. Hsiang and W.Y. Hsiang.)

Hirzebruch noticed the connection between the research efforts of his two students and showed that the link of

$$x_0^d + x_1^2 + \cdots + x_n^2 = 0 \tag{2}$$

for $d \geqslant 2$, $n \geqslant 2$ is an $O(n)$-manifold as above with invariant d, the action being given by the obvious one on the last n coordinates. Since the boundary of the manifold constructed by plumbing copies of the tangent disk bundle of the n-sphere according to the A_{d-1} tree (Fig. 3) also is an $O(n)$-manifold as above with invariant d, these manifolds are identical. Thus the link of the singularity (1) is the $(2n-1)$-dimensional Kervaire sphere; in particular for $n = 5$ it is an exotic 9-sphere.

These results were described in a manuscript "$O(n)$-Mannigfaltigkeiten, exotische Sphären, kuriose Involutionen" of March 1966. (This was not published, since it was supplanted by Hirzebruch's Bourbaki talk [20], and the detailed lecture notes [23] from his course in the winter semester 1966/67 at the University of Bonn.) In a letter [22, C39] of March 29, 1966, Brieskorn reacted to the manuscript with "Klaus Jänich und ich hat-ten von diesem Zusammenhang unserer Arbeiten nichts bemerkt, und ich war vor Freude ganz ausser mir, wie Sie nun die Dinge zusammengebracht haben. Ein schöneres Zusammenspiel von Lehrern und Schülern – wenn ich das so sagen darf – kann man sich doch wirklich nicht denken." (Klaus Jänich and I had not noticed this connection between our work, and I was beside myself with joy to see how you had brought these together. A more beautiful cooperation of student and pupil can one hardly imagine, if I may say so myself.)

At this time the varieties

$$x_0^{a_0} + x_1^{a_1} + \cdots + x_n^{a_n} = 0 \tag{3}$$

($a_i \geqslant 2$) started to receive attention; they are now called "Brieskorn varieties", probably due to the influence of a chapter heading in Milnor's book [38], although they were first examined in this context by Pham and Milnor as well. The corresponding $(2n-1)$-dimensional links

$$K(a_0, a_1, \ldots, a_n) = \left\{x_0^{a_0} + x_1^{a_1} + \cdots + x_n^{a_n} = 0\right\} \cap \mathbf{S}^{2n+1},$$

where $\mathbf{S}^{2n+1}$ is a sphere about the origin, are usually called "Brieskorn manifolds". (The radius of the sphere can be arbitrary since the equation is weighted homogeneous.)

Milnor, who was in Princeton, sent a letter in April of 1966 to John Nash at MIT describing a simple conjecture as to when $K(a_0, a_1, \ldots, a_n)$ is a homotopy sphere: Let $\Gamma(a_0, a_1, \ldots, a_n)$ be the graph with $n+1$ vertices labeled $0, 1, \ldots, n$ and with two vertices i and j joined by an edge if the greatest common divisor (a_i, a_j) is bigger than 1.

CONJECTURE. For $n \geqslant 3$, the link $K(a_0, a_1, \ldots, a_n)$ is a homotopy $(2n-1)$-sphere if and only if the graph $\Gamma(a_0, a_1, \ldots, a_n)$ has

- at least two isolated points, or
- one isolated point and at least one connected component Γ' with an odd number of vertices such that the gcd $(a_i, a_j) = 2$ for all $i \neq j \in \Gamma'$.

Brieskorn then chanced upon an article of Frédéric Pham [50] which dealt with exactly the variety (3) above. In fact, Pham was interested in calculating the ramification of certain integrals encountered in the interaction of elementary particles in theoretical physics. To do this he needed to generalize the Picard–Lefschetz formulas, so let us recall these.

Picard–Lefschetz theory can be summarized as follows (see, for example, [5, 2.1]): Let

$$X_t = \left\{x_0^2 + x_1^2 + \cdots + x_n^2 = t\right\} \subset \mathbf{C}^{n+1}$$

($n \geqslant 1$). Then

(1) The smooth variety X_t for $t \neq 0$ is homotopy equivalent to an n-sphere S^n. (In fact, it is diffeomorphic to the tangent bundle to S^n.)

(2) The homology class of this n-sphere generates the kernel of the degeneration map $H_n(X_t) \to H_n(X_0)$, hence its name of *vanishing cycle*.

(3) The self-intersection of the vanishing cycle is 2 if $n \equiv 0 \bmod 4$, -2 if $n \equiv 2 (\mathrm{mod}\ 4)$ and 0 if $n \equiv 1, 3 (\mathrm{mod}\ 4)$.

(4) Starting at $t = 1$ in the complex plane, traveling once counterclockwise about the origin and returning to the starting point induces a smooth map called the *monodromy* of

X_1 to itself. It is well-defined up to isotopy. Picard–Lefschetz theory gives a description of this map. For example, if $n = 1$ it is a Dehn twist about the one-dimensional vanishing cycle. Picard–Lefschetz theory also describes the induced maps $H_n(X_1) \to H_n(X_1)$ and $H_n(X_1, \partial X_1) \to H_n(X_1)$.

Pham generalized this situation to the case

$$X_t = \{x_0^{a_0} + x_1^{a_1} + \cdots + x_n^{a_n} = t\} \subset \mathbf{C}^{n+1}$$

and found

(1) The smooth variety X_t for $t \neq 0$ is homotopy equivalent to a bouquet $S^n \vee S^n \vee \cdots \vee S^n$ of $(a_0 - 1)(a_1 - 1) \cdots (a_n - 1)$ n-spheres. (This was shown by retracting X_t to a join $Z_{a_0} * Z_{a_1} * \cdots * Z_{a_n}$ where Z_k denotes k disjoint points.)

(2) The homology classes of these n-spheres generate the kernel of the map $H_n(X_t) \to H_n(X_0)$.

(3) An explicit calculation of the intersection pairing on $H_n(X_t)$.

(4) An explicit calculation of the monodromy action on $H_n(X_t)$. (This is induced by rotating each set of points Z_k.)

The article of Pham provided exactly the information Brieskorn needed. (He remarks [7] that "Für den Beweis von [diesen] Aussagen sind jedoch gewisse Rechnungen erforderlich, für die gegenwärtig keine allgemein brauchbare Methode verfügbar ist. Für den Fall der $K(a_0, a_1, \ldots, a_n)$ sind diese Rechnungen aber sämtlich in einem vor kurzem erschienen Artikel von Pham enthalten, und nur die Arbeit von Pham ermöglicht den so mühelosen Beweis unserer Resultate." [Certain calculations, for which there are no general methods at this time, are necessary for the proof of these results. In the case of $K(a_0, a_1, \ldots, a_n)$, however, these calculations are contained in an article of Pham which just appeared, and it is only Pham's work which makes possible such an effortless proof of our results.]) Brieskorn used it to prove a conjecture of Milnor from the preprint [37] about the characteristic polynomial of the monodromy [7, Lemma 4], [38, Theorem 9.1]. He then used this to prove the conjecture above [7, Satz 1], [23, 14.5], [20, Section 2].

Brieskorn also noted that the link $K(a_0, a_1, \ldots, a_n)$, which is $X_0 \cap \mathbf{S}^{2n+1}$, is diffeomorphic to $X_t \cap \mathbf{S}^{2n+1}$ for small $t \neq 0$. This is the boundary of the smooth $(n-1)$-connected manifold $X_t \cap \mathbf{D}^{2n+2}$, which is parallelizable since it has trivial normal bundle. Hence $K(a_0, a_1, \ldots, a_n) \in bP_{2n}$. The information in Pham's paper about the intersection form also led to a formula (derived by Hirzebruch) for the signature of $X_t \cap \mathbf{D}^{2n+2}$. Brieskorn concluded that the link of

$$x_0^{6k-1} + x_1^3 + x_2^2 + x_3^2 + \cdots + x_n^2 = 0 \tag{4}$$

for even $n \geqslant 4$ is k times the Milnor generator of bP_{2n-1} [7, 23].

Through a preprint of Milnor [37], Brieskorn also learned of a recent result of Levine [30] which showed how to compute the Arf invariant needed to recognize whether a link is the Kervaire sphere in terms of the higher-dimensional Alexander polynomial of the knot. The Alexander polynomial for fibered knots is the same as the characteristic polynomial of the monodromy on $H_n(F)$. Hence Brieskorn was able to show [7, Satz 2] that the link of

$$x_0^d + x_1^2 + x_2^2 + \cdots + x_n^2 = 0$$

for $n \geqslant 3$ odd is the standard sphere if $d \equiv \pm 1 \bmod 8$, and the Kervaire sphere if $d \equiv \pm 3 \bmod 8$, thus providing another proof of Hirzebruch's result that the link of the singularity (1) is the Kervaire sphere.

The explicit representation of all the elements of bP_{2n} by links of simple algebraic equations was rather surprising. It provided another way of thinking about these exotic spheres and led to various topological applications.

For example, Nicolaas Kuiper [28] used them to obtain algebraic equations for all non-smoothable piecewise-linear manifolds of dimension eight. (PL manifolds of dimension less than eight are smoothable.) In fact, he started with the complex four-dimensional variety given by Eq. (4) above with $n = 4$. This has a single isolated singularity at the origin. Its completion in projective space has singularities on the hyperplane at infinity, but adding terms of higher order to the equation eliminates these while keeping (analytically) the same singularity at the origin. This variety can be triangulated, giving a combinatorial eight-manifold which is smoothable except possibly at the origin. Since obstructions to smoothing are in one-to-one correspondence with the 28 elements of bP_8, the construction is finished.

Also, the high symmetry of the variety given by Eq. (2) allowed the construction of many interesting group actions on spheres, both standard and exotic [20, Section 4], [23, Section 15]. The actions are the obvious ones: The cyclic group of order d acts by roots of unity on the first coordinate, and there is an involution acting on (any subset of) the remaining coordinates by taking a variable to its negative.

5. The Milnor fibration

About the same time as the above events were happening, Milnor proved a fibration theorem which turned out to be fundamental for much subsequent work. This theorem together with its consequences first appeared in the unpublished preprint [37], which dealt exclusively with isolated singularities. (A full account of this work was later published in the book [38], where the results were generalized to nonisolated singularities. The earlier and somewhat simpler ideas can be found at the end of Section 5 of the book.)

Let $f(x_0, x_1, \ldots, x_n)$ for $n \geqslant 2$ be a complex polynomial with $f(0, \ldots, 0) = 0$ and an isolated critical point at the origin. Let $\mathbf{S}_\varepsilon^{2n+1}$ be a sphere of suitably small radius ε about the origin in $\mathbf{C}^{n+1}$. As before, let $K = \{f(x_0, x_1, \ldots, x_n) = 0\} \cap \mathbf{S}_\varepsilon^{2n+1}$ be the link of $f = 0$ at the origin. The main result of the preprint is the following *fibration theorem*:

THEOREM 5.1. *The complement of an open tubular neighborhood of the link K in $\mathbf{S}_\varepsilon^{2n+1}$ is the total space of a smooth fiber bundle over the circle $\mathbf{S}^1$. The fiber F has boundary diffeomorphic to K.*

The idea of the proof is as follows: If $\mathbf{D}_\varepsilon^{2n+2}$ is the ball of radius ε about the origin and $\delta > 0$ is suitably small, then

$$f : f^{-1}(\mathbf{S}_\delta^1) \cap \mathbf{D}_\varepsilon^{2n+2} \to \mathbf{S}_\delta^1$$

is a clearly a smooth fiber bundle with fiber

$$F' = \left\{f(x_0, x_1, \ldots, x_n) = \delta\right\} \cap \mathbf{D}_\varepsilon^{2n+2}.$$

The total space of this fibration is then pushed out to the sphere $\mathbf{S}_\varepsilon^{2n+1}$ along the trajectories $p(t)$ of a suitably-constructed vector field. This vector field has the property that $|p(t)|$ is increasing along a trajectory, so that points eventually reach the sphere, and also has the property that the argument of $f(p(t))$ is constant and $|f(p(t))|$ is increasing, so that the images of points in $\mathbf{C}$ travel out on rays from the origin. Thus Milnor's proof shows that F is diffeomorphic to F'. The proof also shows that F is parallelizable, since F' has trivial normal bundle.

The fiber F is now called the *Milnor fiber*. He then gives some facts which lead to the topological type of the fiber F and the link K:

(a) The pair $(F, \partial F)$ is $(n-1)$-connected.

(b) The fiber F has the homotopy type of a cell complex of dimension $\leqslant n$. In fact, it is built from the $2n$-disk by attaching handles of index $\leqslant n$.

These assertions follow from Morse theory. In fact, in a lecture at Princeton in 1957 (which was never published), René Thom described an approach to the Lefschetz hyperplane theorems which was based on Morse theory. Thom's approach then inspired Andreotti and Frankel [1] (see also [36, Section 7]) to give another proof of Lefschetz's first hyperplane theorem which used Morse theory, but in a different way: The key observation is that given a n-dimensional complex variety $X \subset \mathbf{C}^m$ and a (suitably general) point $p \in \mathbf{C}^m - X$, then the function on X defined by $|x-p|^2$ for $x \in X$ has nondegenerate critical points of Morse index $\leqslant n$. Thus $H_k(X) = 0$ for $k > n$, which is equivalent to Lefschetz's first hyperplane theorem.

Assertion (b) above follows since the function $|x|$ (or a slight perturbation of it) restricted to F' has critical points of index $\leqslant n$, and assertion (a) follows since the function $-|x|^2$ on F' has critical points of index $\geqslant n$.

By (b), the complement $\mathbf{S}_\varepsilon^{2n+1} - F$ has the same homotopy groups as $\mathbf{S}_\varepsilon^{2n+1}$ through dimension $n-1$. Thus:

(c) The complement $\mathbf{S}_\varepsilon^{2n+1} - F$ is $(n-1)$-connected.

By the Fibration Theorem, $\mathbf{S}_\varepsilon^{2n+1} - F$ is homotopy equivalent to F. Thus

PROPOSITION 5.2. *The fiber F has the homotopy type of a bouquet $S^n \vee \cdots \vee S^n$ of spheres.*

Fact (a) and the above proposition combined with the long exact sequence of a pair show the following:

PROPOSITION 5.3. *The link K is $(n-2)$-connected.*

Milnor used the notation μ for the number of spheres in the bouquet of the first proposition and called it the "multiplicity" since it is the multiplicity of the gradient map of f. However, μ quickly became known as the *Milnor number*. The Milnor number has played a central role in the study of singularities. One reason is that it has analytical as well as a topological descriptions, for example:

$$\mu = \dim_{\mathbf{C}} \mathbf{C}\{x_0, x_1, \ldots, x_n\} / (\partial f/\partial x_0, \partial f/\partial x_1, \ldots, \partial f/\partial x_n)$$

the (vector space) dimension of the ring of power series in $n+1$ variables divided by the Jacobian ideal of the function (see, for example, [47]).

The fact that the Milnor number can be expressed in different ways is extremely useful. For example, the topological interpretation of μ was used by Le and Ramanujam to prove a result which became basic to the study of equisingularity: Suppose that $n \neq 2$. If a family of functions $f_t : \mathbf{C}^{n+1} \to \mathbf{C}$ depending on t with isolated critical points has constant Milnor number, then the differentiable type of the Milnor fibration of f_t is independent of t. The proof uses the topological interpretation of μ to produce a h-cobordism which is thus a product cobordism; hence the restriction $n \neq 2$ [31].

Results similar to the Fibration Theorem and the two propositions have now been obtained in many different situations: complete intersections, functions on arbitrary varieties, polynomials with nonisolated critical points, critical points of polynomials at infinity, and so forth. References to these results can be found in the books and conference proceedings cited at the beginning of this article. Also, there are now many different techniques for computing the Milnor number $\mu = \operatorname{rank} H_n(F)$, the characteristic polynomial of the monodromy $H_n(F) \to H_n(F)$, and the intersection pairing on $H_n(F)$.

The characteristic polynomial of the monodromy turned out to be cyclotomic, and a variety of proofs have appeared of this important fact: the geometric proof of Landman, geometric proofs of Clemens and Deligne–Grothendieck based on resolving the singularity, proofs based on the Picard–Fuchs equation by Breiskorn, Deligne and Katz, and analytic proofs using the classifying space for Hodge structures by Borel and Schmid. For a summary of these and the appropriate references, see [18].

The Milnor number appears in another situation. To describe this we first return to Thom's original observation in his 1957 lecture, as recorded in [2]: Given an n-dimensional complex variety X in affine space and a suitably general linear function $f : X \to \mathbf{C}$, then $|f|^2$ has nondegenerate critical points of Morse index exactly n (except for the absolute minimum). This result is easily proved by writing the function in local coordinates. It forms the basis of Andreotti and Frankel's proof of the second hyperplane theorem of Lefschetz, which says that the kernel of the map on H_{n-1} from a hyperplane section of an n-dimensional projective variety to the variety itself is generated by vanishing cycles.

Thom's original observation was applied in the local context of singularities, where it leads to a basic result in the subject of polar curves relating the Milnor number of a singularity and a plane section. This result has both topological and analytic formulations [57, p. 317]; [29].

6. Brieskorn three-manifolds

The Brieskorn three-manifolds $K(a_0, a_1, a_2)$, the link of

$$x_0^{a_0} + x_1^{a_1} + x_2^{a_2} = 0$$

at the origin, have provided examples figuring in many topological investigations. For example, the local fundamental group of these singularities has proved interesting. As mentioned in Section 3, the surface singularities whose link have finite fundamental group are exactly the quotient singularities. If the surface is embedded in codimension one, and is hence the zero locus of a polynomial $f(x_0, x_1, x_2)$, then these singularities are the well-known *simple singularities*:

$$A_k\colon\ x_0^{k+1} + x_1^2 + x_2^2 = 0,\ (k \geqslant 1)$$

$$D_k\colon\ x_0^{k-1} + x_0x_1^2 + x_2^2 = 0,\ (k \geqslant 4)$$
$$E_6\colon\ x_0^4 + x_1^3 + x_2^2 = 0,$$
$$E_7\colon\ x_0^3 + x_0x_1^3 + x_2^2 = 0,$$
$$E_8\colon\ x_0^5 + x_1^3 + x_2^2 = 0.$$

These equations have appeared, and continue to appear, in many seemingly unrelated contexts [14]. For example, V.I. Arnold showed that they are the germs of functions whose equivalence classes under change of coordinate in the domain have no moduli [3].

More general than Brieskorn polynomials is the class of weighted homogeneous polynomials: A polynomial $f(x_0, x_1, \ldots, x_n)$ is *weighted homogeneous* if there are positive rational numbers $a_0, a_1, \ldots, a_n$ such that

$$f\left(c^{1/a_0}x_0, c^{1/a_1}x_1, \ldots, c^{1/a_n}x_n\right) = cf(x_0, x_1, \ldots, x_n)$$

for all complex numbers c. (Weighted homogeneous polynomials probably first made their appearance in singularity theory in the book of Milnor [38].) Brieskorn singularities are weighted homogeneous, with weights exactly the exponents.

The simple singularities are weighted homogeneous. Milnor [38, p. 80] noted that their weights (a_0, a_1, a_2) satisfy the inequality $1/a_0 + 1/a_1 + 1/a_2 > 1$. He also remarked that the links of the *simple elliptic singularities*

$$\widetilde{E}_6\colon\ x_0^3 + x_1^3 + x_2^3 = 0,$$
$$\widetilde{E}_7\colon\ x_0^2 + x_1^4 + x_2^4 = 0,$$
$$\widetilde{E}_6\colon\ x_0^2 + x_1^3 + x_2^6 = 0,$$

have infinite nilpotent fundamental group. In this case, the sum of the reciprocals of the weights is 1. He conjectured that if $1/a_0 + 1/a_1 + 1/a_2 \leqslant 1$, then the corresponding link had infinite fundamental group, and that this group was nilpotent exactly when $1/a_0 + 1/a_1 + 1/a_2 = 1$.

This conjecture was proved by Peter Orlik [46]. In fact, Orlik and Wagreich [49] had already found an explicit form of a resolution for weighted homogeneous singularities using topological methods based on the existence of a $\mathbf{C}^*$ action, following earlier work by Hirzebruch and Jänich. They also noted that these links are Seifert manifolds [55] and hence could use Seifert's work as well as earlier work by Orlik and others.

Topologists were interested in the question of which homology three-spheres bound contractible four-manifolds (cf. [25, Problem 4.2]). In fact, topological analogues (contractible four-manifolds which are not simply-connected at infinity) of the example of Ramanujam in Section 3 (a contractible complex surface which is not simply connected at infinity) had been found some ten years earlier by Mazur [33] and Poenaru [51]. As Mazur remarks, these examples provide a method of constructing many examples of odd topological phenomena.

It was known (see Milnor's conjecture in Section 4) that $K(a_1, a_2, a_3)$ is a homology three-sphere exactly when the integers a_1, a_2, a_3 are pairwise relatively prime. (As Milnor remarks in [39], this result in this context of Seifert fiber spaces is already in [55].) Links of Brieskorn singularities were particularly easy to study, since a resolution of the singularity exhibited the link as the boundary of a four-manifold, and data from the resolution provided a plumbing description of this manifold which then could be manipulated to eventually get a contractible manifold. For example, Casson and Harer [12] showed that the Brieskorn manifolds $K(2,3,13)$, $K(2,5,7)$ and $K(3,4,5)$ are boundaries of contractible four-manifolds. Much has now happened in this area as can be seen in Kirby's update of his problem list [26].

Brieskorn three-manifolds and their generalizations also provided interesting examples of manifolds with a "geometric structure". Klein proved long ago that the links of the simple singularities listed above are of the form S^3/Γ, the quotient of the group of unit quaternions by a discrete subgroup.

Milnor [39, Section 8] proved by a round-about method that the links of the simple elliptic singularities are quotients of the Heisenberg group by discrete subgroups. He then showed that the links of Brieskorn singularities with $1/a_0 + 1/a_1 + 1/a_2 \leqslant 1$ are quotients of the universal cover of $SL(2, \mathbf{R})$ by discrete subgroups. (Similar results were obtained at the same time by Dolgachev.)

Thus many links admitted a locally homogeneous (any two points have isometric neighborhoods) Riemannian metric and hence provided nice examples of Thurston's eight geometries [58]. These results were extended by Neumann [48]. Later he and Scherk [44] found a more natural way of describing the connection between the geometry on the link and the complex analytic structure of the singularity in terms of locally homogeneous nondegenerate CR structures.

The three-dimensional Brieskorn manifolds have also been central examples in the study of the group Θ_3^H of homology three-spheres. This group is bound up with the question of whether topological manifolds can be triangulated. It was originally thought that this group might just have two elements. However, techniques from gauge theory were used to show that it is actually infinite and even infinitely generated. In particular the elements $K(2,3,6k-1)$ for $k \geqslant 1$ have infinite order in this group, and are linearly independent. Brieskorn manifolds appear in this context because the three-manifolds are boundaries of plumbed four-manifolds upon which explicit surgeries can be performed [17].

Also, the Casson invariant of some types of links of surface singularities in codimension one (including Brieskorn singularities) was proved to be $1/8$ of the signature of the Milnor fiber [45].

7. Other developments

This last section recounts two developments which occurred outside the main stream of events as recounted in the previous sections. They are both applications of topology to algebraic geometry. The first is a theorem of Dennis Sullivan [56]:

THEOREM 7.1. *If K is the link of a point in a complex algebraic variety, then the Euler characteristic of K is zero.*

If the point is smooth or an isolated singular point, then the link is a compact manifold of odd dimension and hence has Euler characteristic zero. The surprising feature of this result is that it should be true for nonisolated singularities as well.

Sullivan discovered this result during his study of combinatorial Stiefel–Whitney classes. He recounts that initially it was clear to him that this result was true in dimensions one and two. He then asked Pierre Deligne if he knew of any counterexamples in higher dimensions, but the latter replied "almost immediately" with a proof based on resolving the singularity. Sullivan then deduced this result in another fashion: Since complex varieties have a stratification with only even-dimensional strata, the link has a stratification with only odd-dimensional strata. He then proved, by induction on the number strata, that a compact stratified space with only odd-dimensional strata has zero Euler characteristic.

Since real varieties are the fixed point set of the conjugation map acting on their complexification, this result has the following consequence for real varieties:

COROLLARY 7.2. *If K is the link of a point in a real algebraic variety, then the Euler characteristic of K is even.*

Sullivan remarks that the result for complex varieties follows from essentially "dimensional considerations", but that the corollary for real varieties is however "geometrically surprising". This result continues to form a basis for the investigation of the topology of real varieties.

The second result is one of Thom. Given a singularity of an arbitrary variety $X_0 \subset \mathbf{C}^m$, one can ask if it can be "smoothed" in its ambient space $\mathbf{C}^m$ in the sense that it can be made a fiber of a flat family $X_t \subset \mathbf{C}^m$ whose fibers X_t for small $t \neq 0$ are smooth. For example, a hypersurface singularity is smoothable in its ambient space since it is the zero locus of a polynomial and hence smoothed by nearby fibers of the polynomial.

The first example of a nonsmoothable singularity was constructed by Thom (see [19]). In fact, Thom showed that the variety $X \subset \mathbf{C}^6$ defined by the cone on the Segre embedding $\mathbf{P}^1 \times \mathbf{P}^2 \to \mathbf{P}^5$ is not smoothable: If it were, the link $K^7 \subset \mathbf{S}^{11}$ of $X \subset \mathbf{C}^6$ at the origin would be null cobordant (as a manifold with complex normal bundle) in $\mathbf{S}^{11}$, but it is not. This is proved by a computation with characteristic classes. (The manifold K^7 is odd-dimensional and hence null cobordant, but not in $\mathbf{S}^{11}$.)

Acknowledgments

I thank F. Hirzebruch for allowing me to use material from a lecture in July 1996 at the Oberwolfach conference in honor of Brieskorn's 60th birthday, and I thank both him and D. O'Shea for comments on a preliminary version of this article. I also thank Harvard University for their hospitality during the year 1997–1998 when this article was written.

Bibliography

[1] A. Andreotti and T. Frankel, *The Lefschetz theorem on hyperplane sections*, Ann. of Math. **69** (1959), 313–713.

[2] A. Andreotti and T. Frankel, *The second Lefschetz theorem on hyperplane sections*, Global Analysis. Papers in honor of K. Kodaira, D. Spencer and S. Iyanaga, eds, Univ. of Tokyo Press and Princeton Univ. Press, Tokyo, Princeton (1969), 1–20.

[3] V.I. Arnold, S.M. Gusein-Zade and A.N. Varchenko, *Singularities of Differentiable Maps*, Vol. I, Birkhäuser, Boston (1985).

[4] V.I. Arnold, S.M. Gusein-Zade and A.N. Varchenko, *Singularities of Differentiable Maps*, Vol. II, Birkhäuser, Boston (1988).

[5] V.I. Arnold (ed.), *Dynamical Systems VI (Singularity Theory I)*, Springer, Berlin (1993).

[6] K. Brauner, *Zur Geometrie der Funktionen zweier komplexen Veränderlichen*, Abh. Math. Sem. Hamburg **6** (1928), 1–54.

[7] E. Brieskorn, *Beispiele zur Differentialtopologie von Singularitäten*, Inventiones Math. **2** (1966), 1–14.

[8] E. Brieskorn, *Examples of singular normal complex spaces which are topological manifolds*, Proc. Nat. Acad. Sci. USA **55** (1966), 1395–1397.

[9] E. Brieskorn, *Rationale Singularitäten komplexer Flächen*, Inventiones Math. **4** (1968), 336–358.

[10] W. Burau, *Kennzeichnung der Sclauchknoten*, Abh. Math. Sem. Hamburg **9** (1932), 125–133.

[11] W. Burau, *Kennzeichnung der Schlauchverkettungen*, Abh. Math. Sem. Hamburg **10** (1934), 285–397.

[12] A. Casson and J. Harer, *Some homology lens spaces which bound rational homology balls*, Pacific J. Math. **96** (1981), 23–36.

[13] A. Durfee, *Knot invariants of singularities*, Proc. Symp. Pure Math 29: Algebraic Geometry, Arcata 1974, R. Hartshorne, ed., Amer. Math. Soc., Providence, RI (1975), 441–448.

[14] A. Durfee, *Fifteen characterizations of rational double points and simple critical points*, Enseign. Math. **25** (1979), 131–163.

[15] D. Eisenbud and W. Neumann, *Three-Dimensional Link Theory and Invariants of Plane Curve Singularities*, Princeton Univ. Press, Princeton, NJ (1985).

[16] M. Epple, *Branch points of algebraic functions and the beginnings of modern knot theory*, Historia Mathematica **22** (1995), 371–401.

[17] R. Fintushel and R. Stern, *Invariants for homology 3-spheres*, Geometry of Low-Dimensional Manifolds I (Proceedings of the Durham Symposium, 1989), S.K. Donaldson and C.B. Thomas, eds, Cambridge Univ. Press, Cambridge (1990), 125–148.

[18] P. Griffiths, *Appendix to the article of A. Landman: On the Picard–Lefschetz transformation for algebraic manifolds acquiring general singularities*, Trans. Amer. Math. Soc. **181** (1973), 123–126.

[19] R. Hartshorne, *Topological conditions for smoothing algebraic varieties*, Topology **13** (1974), 241–253.

[20] F. Hirzebruch, *Singularities and exotic spheres*, Seminaire Bourbaki, 1966/67, No. 314.

[21] F. Hirzebruch, *The topology of normal singularities of an algebraic surface (d'apres Mumford)*, Seminaire Bourbaki, 1962/63, No. 250.

[22] F. Hirzebruch, *Gesammelte Abhandlungen*, Springer, Berlin (1987). (The commentary to paper number n is cited as [Cn].)

[23] F. Hirzebruch and K.H. Mayer, *$O(n)$-Mannigfaltigkeiten, exotische Sphären und Singularitäten*, Springer, Berlin (1968).

[24] E. Kahler, *Über die Verzweigung einer algebraischen Funktion zweier Veränderlichen in der Umgebung einen singulären Stelle*, Math. Zeit. **30** (1929), 188–204.

[25] R. Kirby, *Problems in low dimensional manifold theory*, Algebraic and Geometric Topology (Proc. Symp. Pure Math. 32, vol. 2), Amer. Math. Soc., Providence, RI (1978), 273–312.

[26] R. Kirby, *Problems in low-dimensional topology* (available from http://math.berkely.edu/~kirby) (1996).

[27] M. Kervaire and J. Milnor, *Groups of homotopy spheres: I*, Ann. of Math. **77** (1963), 504–537.

[28] N. Kuiper, *Algebraic equations for nonsmoothable 8-manifolds*, Publ. Math. IHES **33** (1968), 139–155.

[29] D.T. Le, *Calcul du nombre de cycles évanouissants d'une hypersurface complexe*, Ann. Inst. Fourier (Grenoble) **23** (1973), 261–270.

[30] J. Levine, *Polynomial invariants of knots of codimension two*, Ann. of Math. **84** (1966), 537–554.

[31] D.T. Le and C. Ramanujam, *The invariance of Milnor's number implies the invariance of the topological type*, Amer. J. Math. **98** (1976), 67–78.

[32] D.T. Le, K. Saito and B. Teissier (eds), *Singularity Theory*, World Scientific, Singapore (1995). (Proceedings of the Trieste Conference 1991).

[33] B. Mazur, *A note on some contractible 4-manifolds*, Ann. of Math. (2) **73** (1961), 221–228.

[34] J. Milnor, *Differentiable manifolds which are homotopy spheres* (unpublished preprint, Princeton) (1959).

[35] J. Milnor, *Differential topology*, Lectures on Modern Mathematics, T. Saaty, ed., Wiley, New York (1964), 165–183.

[36] J. Milnor, *Morse Theory*, Princeton University Press, Princeton (1966).

[37] J. Milnor, *On isolated singularities of hypersurfaces* (unpublished preprint, Princeton) (1966).

[38] J. Milnor, *Singular Points of Complex Hypersurfaces*, Princeton University Press, Princeton (1968).

[39] J. Milnor, *On the 3-dimensional Brieskorn manifolds $M(p,q,r)$*, Knots, Groups and 3-Manifolds, L. Neuwirth, ed., Princeton Univ. Press., Princeton, NJ (1975), 175–225.
[40] D. Mumford, *The topology of normal singularities of an algebraic surface and a criterion for simplicity*, Publ. Math. IHES **9** (1961).
[41] W. Neumann, *A calculus for plumbing applied to the topology of complex surface singularities and degenerating complex curves*, Trans. Amer. Math. Soc. **268** (1981), 299–344.
[42] W. Neumann, *Complex algebraic plane curves via their links at infinity*, Invent. Math. **98** (1989), 445–489.
[43] W. Neumann and L. Rudolph, *Unfoldings in knot theory*, Math. Ann. **278** (1987), 409–439; and: *Corrigendum* **282** (1988), 349–351.
[44] W. Neumann and J. Scherk, *Links of surface singularities and CR space forms*, Comment. Math. Helvetici **62** (1987), 240–264.
[45] W. Neumann and J. Wahl, *Casson invariant of links of singularities*, Comment. Math. Helvetici **65** (1990), 58–78.
[46] P. Orlik, *Weighted homogeneous polynomials and fundamental groups*, Topology **9** (1970), 267–273.
[47] P. Orlik, *The multiplicity of a holomorphic map at an isolated critical point*, Real and Complex Singularities, Oslo 1976, P. Holm, ed., Sijthoff and Noordhoff International, Alphen aan den Rijn (1977), 405–474.
[48] P. Orlik (ed.), *Singularities*, Proc. Symp. Pure Math. 40, Amer. Math. Soc., Providence, RI (1983).
[49] P. Orlik and P. Wagreich, *Isolated singularities of algebraic surfaces with* $\mathbf{C}^*$ *action*, Ann. of Math. **93** (1971), 205–228.
[50] F. Pham, *Formules de Picard–Lefschetz généralisées et ramification des intégrales*, Bull. Soc. Math. France **93** (1965), 333–367.
[51] V. Poenaru, *Les décompositions de l'hypercube en produit topologique*, Bull. Soc. Math. France **88** (1960), 113–129.
[52] C. Ramanujam, *A topological characterization of the affine plane as an algebraic variety*, Ann. of Math. **94** (1971), 69–88.
[53] J.E. Reeve, *A summary of results in the topological classification of plane algebroid singularities*, Rendiconti Sem. Math. Torino **14** (1954), 159–187.
[54] L. Rudolph, *Embeddings of the line in the plane*, J. Reine Angew. Math. **337** (1982), 113–118.
[55] H. Seifert, *Topologie dreidimensionaler gefäserter Räume*, Acta Mathematica **60** (1932), 147–238.
[56] D. Sullivan, *Combinatorial invariants of analytic spaces*, Proceedings of Liverpool Singularities Symposium I, C.T.C. Wall, ed., Springer, Berlin (1971), 165–168.
[57] B. Teissier, *Cycles évanescents, sections planes, et conditions de Whitney*, Astérisque 7–8: Singularités à Cargèse, F. Pham, ed., Soc. Math. France, Paris (1973), 285–362.
[58] W. Thurston, *Three-dimensional manifolds, Kleinen groups and hyperbolic geometry*, Bull. Amer. Math. Soc. N.S. **6** (1982), 357–381.
[59] P. Wagreich, *Singularities of complex surfaces with solvable local fundamental group*, Topology **11** (1972), 51–72.
[60] M. Zaidenberg, *Lectures on exotic algebraic structures on affine spaces*, Preprint AG/9801075 available from the xxx Mathematics Archive at http://front.math.ucdavis.edu/.
[61] O. Zariski, *On the topology of algebroid singularities*, Amer. J. Math. **54** (1932), 453–465.

CHAPTER 14

One Hundred Years of Manifold Topology

S.K. Donaldson

Department of Mathematics, Stanford University, Stanford, CA 94305, USA

This article makes little claim to any scholarly merit: it is merely a discussion of some episodes in the development of manifold theory which interest the author, partly in the light of contemporary research developments. This discussion might perhaps suggest some themes which could be dealt with more thoroughly in the future: in the present article we cannot come close to doing justice to these large themes. The author would like to mention, particularly to colleagues who know far more about these matters than he does, that the starting point for this article was a lecture (in St. Catherine's College, Oxford) aimed at a general audience, and it is thus more at the level of popular history rather than the scholarly study which the material deserves.

* * *

The study of the topology of manifolds can be regarded as beginning with the renowned series of papers [14] by Poincaré, published between 1895 and 1904. This granted, the subject has just past its centenary. It Poincaré's papers we find the beginnings of homology theory, the fundamental group and the birth of the *classification problem*, notably of course in the Poincaré conjecture on simply connected 3-manifolds. This classification problem; that is the definition of invariants of manifolds and the enumeration of the manifolds with given invariants (in the various categories: topological, smooth, PL, etc.) makes up one pillar in the century of work since then. Other pillars are formed by the interaction between manifold topology *per se* and neighbouring areas of geometry and analysis. The theme which we will focus on in this article is the topology of complex algebraic varieties, particularly complex surfaces. This was clearly one of the main motivations for Poincaré's work: it is the first of the three examples of the application of topological ideas which he mentions in the introduction to [14], and it was the subject of the third and fourth *compléments* in the series.

HISTORY OF TOPOLOGY
Edited by I.M. James

1. The Poincaré–Picard–Lefschetz description of the topology of a complex algebraic variety

The crucial role of the topology of (real) surfaces in the function theory of one complex variable is well-known and goes back to Riemann. In modern language we are dealing with compact Riemann surfaces, which have of course a single topological invariant, the genus. The early point of view was to think of Riemann surfaces presented as branched covers of the Riemann sphere. That is, we consider a branched covering map $f : \Sigma \to \mathbb{CP}^1$, of degree d, with branch points $b_i \in \mathbb{CP}^1$. The pair (Σ, f) can be recovered from data consisting of the configuration of points b_i in the sphere and the *monodromy* homomorphism from $\pi_1(\mathbb{CP}^1 \setminus \{b_i\})$ to the permutation group on d elements. The study of these Riemann surfaces was in large part motivated by the study of contour integrals: in modern language we consider meromorphic one-forms on the surface Σ. The importance of topological ideas stems from the connection between these and the genus of the surface. More precisely the space of holomorphic 1-forms (or *forms of the first kind*) has dimension g. A meromorphic 1-form with zero residue at each singularity is called a *form of the second kind*, and another important fact is that the forms of the second kind, modulo the derivatives of meromorphic functions, form a space of dimension $2g$ (isomorphic to the first real cohomology group of Σ).

Beginning in 1882, E. Picard began the study of meromorphic forms and their integrals on complex algebraic surfaces [13]. This lead him naturally to topological questions which, as mentioned above, were taken further by Poincaré and developed into a comprehensive theory by Lefschetz [11]. The point of view all these authors adopted was to describe a surface as the total space of a family of hyperplane sections: a "Lefschetz pencil". (A detailed modern treatment of Lefschetz' theory has been given by Lamotke [10]). This is the natural generalisation of the description of a Riemann surface as a branched cover. For example, suppose our complex surface S is embedded in projective 3-space and so, leaving out the points at infinity, can be described by the solutions of a polynomial equation $P(x, y, z) = 0$ in three complex variables x, y, z. Then for each fixed $t \in \mathbb{C}$ the equation $P(x, y, t) = 0$ defines a complex curve C_t in the x-y space, and the main idea is to study S via this family of curves. If the polynomial is reasonably generic then C_t will be smooth for all but finitely many values of t, and at these exceptional values the curve will acquire an ordinary double point, with two branches crossing transversally. More precisely, the generic picture is the complex analogue of that of the Morse function in differential topology: the function on the affine part of the surface given by projection to the z coordinate has finitely many critical points, around each of which it is given by a nondegenerate quadratic form in suitable holomorphic local coordinates. In more abstract, and more general, language, we suppose that $L \to S$ is an ample line bundle and that s_0, s_∞ are two sections of L, cutting out curves $C_0, C_\infty \subset S$ which we suppose meet transversally in a finite set A in S. The ratio $f = s_1/s_0$ is a meromorphic function on S which gives a genuine holomorphic map $\bar{f}$ from the blow-up $\overline{S}$ of S at the points A to $\mathbb{CP}^1$. We suppose that this satisfies the Morse condition as above, interpreted in the obvious way over the point at infinity. We also suppose that there is a single critical point lying over each critical value. For $t \in \mathbb{CP}^1$ the fibre $\bar{f}^{-1}(t)$ is a complex curve C_t which can be identified with the curve in S cut out as the zero-set of the section $s_0 - ts_\infty$ of L.

In the picture above there are a finite number, r say, of *critical values* b_i in $\mathbb{CP}^1$ and away from these the map $\bar{f}$ is a fibration. The structure of $\bar{f}$ around a critical value is determined by a "vanishing cycle". Suppose without loss that 0 is a critical value and consider the family of Riemann surfaces C_t as t runs over a small interval $[0, \varepsilon]$ in $\mathbb{R} \subset \mathbb{C}$. The singular fibre C_0 can be thought of as being obtained from the smooth fibre C_ε by collapsing a circle $\delta \subset C_\varepsilon$ to a point. To be completely explicit we may work in local coordinates z_1, z_2 about the critical point in which $\bar{f}$ is given by $z_1^2 + z_2^2$. Then a model for the vanishing cycle is the circle given by the *real* solutions of the equation $z_1^2 + z_2^2 = \varepsilon$. This vanishing cycle δ is well-defined up to isotopy of the generic fibre of $\bar{f}$.

A fundamental ingredient in the Picard–Lefschetz–Poincaré theory is the monodromy action on the homology of the fibres. If t_0, t_1 are points in $\mathbb{CP}^1 \backslash \{b_i\}$ and γ is a path from t_0 to t_1 which avoids the critical values there is a monodromy map, well-defined up to isotopy, $\phi^\gamma : C_{t_0} \to C_{t_1}$ which induces a corresponding map ϕ_*^γ on H_1. It is striking that these early authors do not consider the map ϕ^γ explicitly but pass directly to homology. In any case the key thing is to consider the monodromy of a path γ which loops once around a single critical value and returns to the same base point, for example the circle $|t| = \varepsilon$ in our standard model. For this path the monodromy is the *Dehn twist* of the fibre C_ε about the vanishing cycle δ. That is, we identify a neighbourhood N of δ in C_ε with the cylinder $\mathbb{R}/\mathbb{Z} \times [0, 1]$ and define a map from C_ε to itself to be equal to the identity outside N and by the formula $(\theta, s) \mapsto (\theta + s, s)$ inside N. Again, up to isotopy, this is independent of choices (one needs to check orientations: the map is pinned down by an orientation of C_ε, but an orientation of δ is not needed). This description of the monodromy leads directly to the famous Picard–Lefschetz formula for the action on homology

$$\phi_*^\gamma(\eta) = \eta - \langle \eta, \delta \rangle \delta,$$

for $\eta \in H_1(C_\varepsilon)$.

We can now state the Picard–Lefschetz description of the homology of the surface S. Let us suppose for definiteness that all the critical values are contained in a large disc in $\mathbb{C}$ and that the imaginary parts of the critical values are all different. Choose two base points t_-, t_+ on the real line, with $t_- \ll 0$, $t_+ \gg 0$ and fix paths γ_i^+, γ_i^- from the critical value b_i to t_-, t_+, respectively, which are made by slightly bending the horizontal lines $\mathrm{Im}(z) =$ constant through the critical values. By transporting the vanishing cycles along these lines we get a collection of curves δ_i^+ in $C_+ = C_{t_+}$ and another set δ_i^- in $C_- = C_{t_-}$. Now define

$$A : H_1(C_-) \to \mathbb{Z}^r, \qquad B : \mathbb{Z}^r \to H_1(C_+)$$

by

$$A(\eta) = \big(\langle \delta_i^-, \eta \rangle\big), \qquad B(\lambda_i) = \sum_i \lambda_i \delta_i^+.$$

Then the composite $B \circ A : H_1(C_+) \to H_1(C_-)$ is zero and the homology of the complex

$$0 \to H_1(C_-) \to \mathbb{Z}^r \to H_1(C_+) \to 0$$

is given by

$$\ker A = H_3(S) = H_3(\overline{S}), \qquad \ker B/\mathrm{Im}\, A = \overline{C}^\perp / \overline{C},$$
$$\mathrm{coker}\, B = H_1(S) = H_1(\overline{S}).$$

Here we mean by $\overline{C}$ the homology class of the fibres C_t in $H_2(\overline{S})$: this has self-intersection 0 so is contained in its orthogonal complement $\overline{C}^{\perp} \subset H_2(\overline{S})$ with respect to the intersection form. We should also recall that

$$H_2(\overline{S}) = H_2(S) \oplus \mathbb{Z}^a,$$

the extra summand being generated by the fundamental classes of the exceptional curves in the blow-up. (At the level of differential topology, the blow-up $\overline{S}$ is the connected sum of S with a copies of the orientation-reversed projective plane, but again this point of view comes much later.) A variant of this construction is to use the fact that the vanishing cycles naturally define classes in $H_1(C_+\backslash A)$ and to consider maps:

$$0 \to H_1(C_-) \to \mathbb{Z}^a \to H_1(C_+\backslash A) \to 0$$

defined as before. This gives a complex which computes the "primitive part" of $H_2(S)$, which is just the orthogonal complement with respect to the intersection form of the curve C_0.

If we choose a large semi-circle in the upper half plane joining t_+, t_- we can transport the cycles δ_i^+ from C_+ to get a second set of cycles, which we still denote by δ_i^+ in C_-. These can be expressed in terms of the original set δ_-^i using the monodromy formula: if the critical values are ordered so that the imaginary part of b_i decreases with i we have:

$$\delta_1^+ = \delta_1^-,\ \delta_2^+ = R_1(\delta_2^-),\ \delta_3^+ = R_2R_1(\delta_3^-),\ \ldots,$$

where R_i is the endomorphism of $H_1(C_-)$ given by $R_i(\eta) = \eta - \langle\delta_i^-, \eta\rangle\delta_i^-$. (Then one can check that the equation $B \circ A = 0$ follows from the condition $R_nR_{n-1}\ldots R_2R_1 = 1$ which holds because the monodromy around a large circle is trivial.) So in sum the Picard–Lefschetz–Poincaré theory gives a complete description of the homology of the surface in terms of the collection of vanishing cycles δ_i^- in the fibre C_-.

By itself, the description above does not seem to lead to any general results about the homology of complex surfaces, but it gives a framework within which such results can be obtained by a deeper study. The first such result stated by Lefschetz is a particular case of what is called the "hard Lefschetz theorem" for the first homology of S. This can be expressed in various equivalent ways (cf. [10, 7]). In terms of cohomology (with real coefficients), if we let $h \in H^2(S)$ be the fundamental class of one of the curves C_t (i.e. the first Chern class of L) then the result asserts that:

$$(\alpha_1, \alpha_2) \mapsto \langle\alpha_1 \cup \alpha_2 \cup h, [S]\rangle,$$

defines a nondegenerate skew form on $H^1(S)$. In particular the first Betti number of S is even. Lefschetz observes from this ([11, p. 18]) that there are compact oriented 4-manifolds which are not homeomorphic to algebraic surfaces, in contrast to the case of complex curves. (Of course we would find it simpler to use the fact that the second Betti number of S is nonzero.) In the framework of the vanishing cycles description above, the hard Lefschetz assertion translates into the following: let $V \subset H_1(C_-; \mathbb{R})$ be the subspace generated by the vanishing cycles δ_i^- and $V^{\perp}$ be its annihilator under the intersection form on H_1. The subspace $V^{\perp}$ consists of the "invariant cycles", fixed by the monodromy around all loops. According to our general statement, $H^1(S)$ can be identified with $V^{\perp}$ and hard Lefschetz is the assertion that the intersection form on $H_1(C_-)$ restricts to a nondegenerate form on $V^{\perp}$, or equivalently that $V \cap V^{\perp} = 0$ – the invariant cycles are complementary to the vanishing cycles.

Lefschetz' direct topological argument to prove that $V \cap V^{\perp} = 0$ is notoriously hard to follow. However, he also gave ([11, p. 62]) a transcendental argument to prove the assertion. According to results of Picard [13], there is for every nonzero invariant cycle $\eta \in V^{\perp}$ a closed one-form σ of the second kind on S (that is, a meromorphic 1-form with zero residues) whose integral around η is 1. But then if η lies in V it bounds in S, so the integral of σ around η must be zero, a contradiction. The essential point here is that a closed one form of the second kind defines a cohomology class, even though it has singularities. The fact that

$$H^1(S; \mathbb{C}) = \frac{\text{closed one-forms of 2nd kind}}{d\text{-meromorphic functions}},$$

seems likely to also have been known to Picard–Lefschetz, although I have not been able to trace a precise statement.

It would be extremely interesting to reconstruct, if possible, a direct topological proof of the hard Lefschetz theorem. In this direction we would like to point out that it seems that one *cannot* prove the result using only the general properties of the pencil description. More precisely we consider "topological Lefschetz fibrations". By this we mean a smooth map $d : X \to \mathbb{CP}^1$, where X is some differentiable 4-manifold, such that g is a submersion outside a finite set of points in X and the structure around these critical points is modelled on the complex case. The whole discussion of vanishing cycles, etc. goes through to this situation, but the analogue of the hard Lefschetz theorem is *not* true: we can find 4-manifolds X which admit such a topological fibration but with odd first Betti number. For example, suppose that $\overline{S} \to \mathbb{CP}^1$ is a genuine complex fibration as considered above, and that $H^1(S)$ is nonzero, so we have a nontrivial decomposition

$$H_1(C) = V \oplus V^{\perp}$$

of the homology of a generic fibre C. Take an element in $H_1(C)$ which does not lie in either of the subspaces V, $V^{\perp}$ and which can be represented by an embedded circle τ in the Riemann surface C. The Dehn twist in τ is a diffeomorphism α of C and using the Picard–Lefschetz formula one sees that $V^{\perp} \cap \alpha_*(V^{\perp})$ is the codimension one subspace of $V^{\perp}$ consisting of classes whose intersection with τ is zero. In particular, the dimension of $V^{\perp} \cap \alpha_*(V^{\perp})$ is *odd* (since that of $V^{\perp}$ is even). Now form a topological Lefschetz pencil X by taking the fibre sum of two copies of $\overline{S}$, gluing together the boundaries of the tubular neighbourhoods of copies of C, but using the nontrivial diffeomorphism α to identify the two copies. We get a topological Lefschetz fibration with twice the number of critical values as in the original pencil $\overline{S}$ and the invariant subspace $V_X^{\perp}$ is $V^{\perp} \cap \alpha_*(V^{\perp})$, so the first Betti number of X is odd.

One can counter this example by various points. One salient point is that Lefschetz' argument in ([11, Chapter II, Section 13]) uses the fact that in the complex case one may suppose the pencil forms part of a larger linear system: that is, he uses a third section of the line bundle L or in geometrical terms the existence of a holomorphic branched covering map from S to $\mathbb{CP}^2$. But in any case the discussion above perhaps highlights further the interest in understanding the topological meaning of the hard Lefschetz theorem, a theme we will return to in Section 3 below.

One of the features that distinguishes Lefschetz' work from his predecessors' is that he went on to consider algebraic varieties of all dimensions. The general set-up is the same: if Z is an n-dimensional complex projective manifold one considers a blow-up $\overline{Z}$ of Z along an axis A and a "meromorphic Morse function"

$$\bar{f}:\overline{Z} \to \mathbb{CP}^1$$

with a finite number m of critical points. (In the case of Riemann surfaces we are just considering branched covers with simple branch points.) The generic fibre, or hyperplane section, Y is an $(n-1)$-dimensional complex manifold and the vanishing cycles are now embedded $n-1$ spheres in Y. Then the homology groups $H_p(Z)$, $H_p(Y)$ are isomorphic for $p \leqslant n-2$ (the "hyperplane theorem") and there is a complex

$$0 \to H_{n-1}(Y) \to \mathbb{Z}^r \to H_{n-1}(Y\backslash A) \to 0$$

which computes $H_{n-1}(Z)$, its dual $H_{n+1}(Z)$, and the "primitive part" of the middle-dimensional homology of Z. This higher-dimensional theory also gives results about surfaces. By starting with a projective space and applying the hyperplane theorem to a sequence of intersections one deduces that any complex surface which is a complete intersection in some projective space must be simply connected: so for example no product of Riemann surfaces $\Sigma_1 \times \Sigma_2$ apart from $S^2 \times S^2$ can be a complete intersection: a purely algebro-geometric application of the topological theory.

In the 1930's and 1940's Hodge [8] developed a completely different approach to the homology of complex projective varieties. Hodge's main theorem, that the cohomology of a compact Riemannian manifold can be represented by harmonic forms, applies to general manifolds, but the most important applications arise in the case of complex manifolds. These include proofs of the hard Lefschetz and "primitive decomposition" theorems (which express the whole homology of a complex algebraic variety as a sum of copies of the primitive parts, as encountered above, see [10, 7]). The key notion in this theory is the existence of a Kähler form ω on a complex projective manifold Z, whose de Rham cohomology class is the Poincaré dual of the class of a hyperplane section. Thus Z is simultaneously a complex, symplectic and Riemannian manifold and all three structures play a role. The primitive decomposition, from Hodge's point of view, arises from the fact that the operation of wedge product with ω, which corresponds at the level of cohomology to intersecting cycles with a hyperplane, maps harmonic forms to harmonic forms. In a similar vein, the special properties of harmonic forms on a Kähler manifold were shown by Deligne et al. [2] to lead to further topological constraints: all the Massey products on such a manifold are trivial. Recall that these are higher order operations, defined on cohomology classes whose primary cup products vanish. In the simplest case of classes defined by holomorphic forms this follows from the fact that the wedge product of holomorphic forms is again holomorphic and if the product vanishes in cohomology it does so pointwise. More generally, the real homotopy type of such a manifold is "formal"; entirely determined by the cohomology ring and fundamental group.

2. Whitney discs and high-dimensional manifold topology

In this section we change tack and discuss a result of Whitney [21] which is a milestone in the development of geometric topology (and, incidentally falls almost exactly mid-way through the century of development of that subject). The problem Whitney was considering was the embedding of a compact oriented n-manifold M^n in Euclidean space, and the result he proved was that M^n can be embedded in $\mathbb{R}^{2n}$. Recall that the much easier result in this direction is that M^n can be embedded in $\mathbb{R}^{2n+1}$: this follows from general position arguments, indeed a generic map from M^n to $\mathbb{R}^{2n+1}$ is an embedding. In a similar way one can see that a generic map from M^n to $\mathbb{R}^{2n}$ is an immersion, with a finite number of double points where two sheets cross transversally. Using the orientations we can attach a sign to each of these intersection points. The immersion may be chosen to have normal Euler number zero and then the algebraic sum of the intersections vanishes. For example, suppose there are just two double points p_+, p_- with opposite signs: the problem Whitney addressed is to show that the map can be modified to cancel these two intersection points. The argument used in an essential way a hypothesis on the dimension: $n \geqslant 3$. (The main result on embeddings is true for all n, but we use a different argument for $n = 1, 2$: indeed we know that the circle embeds in $\mathbb{R}^2$ and any oriented surface in $\mathbb{R}^3$.)

The significance of Whitney's argument is that homological conditions which are rather clearly necessary (algebraic intersection zero) are shown by a deeper analysis to be actually *sufficient* to obtain a geometrical statement: removal of double points. With minor variants, Whitney's argument adapts to other problems involving submanifolds, or pairs of submanifolds, always provided suitable dimension restrictions hold, and it is this which accounts for the division of manifold theory into "high-dimensional" and "low-dimensional" topology. We will sketch a version of Whitney's argument for the case of a pair of middle-dimensional submanifolds $P^n, Q^n \subset V^{2n}$, where all manifolds are oriented, P and Q are connected and V is simply connected. (For a complete treatment, in the PL setting, see [17].) If P and Q are in general position they have a finite number of transverse intersections: we would like to deform P and Q to remove these intersections. More precisely, we seek an isotopy f_t of V with f_0 the identity and with $f_1(P)$ disjoint from Q. The obvious necessary condition is that the homological intersection number $P \cdot Q$ is zero, and the main result is that if $n \geqslant 3$ this condition is also sufficient. Suppose again for simplicity that there are just two intersection points v_+, v_- with opposite signs. Then we choose an embedded path γ_P from v_+ to v_- in P and another one γ_Q in Q. The loop $\gamma_P \gamma_Q^{-1}$ bounds a map of the disc D into V since V is simply connected. Now the condition $n \geqslant 3$ is used in three different ways.

1. Since $\dim V > 4$, a generic map from the 2-dimensional disc D to V is an embedding.

2. Since $\dim V > 2 + \dim P$, a generic map of the disc does not meet P in its interior, likewise for Q.

3. Given an embedded disc D as above, we wish to choose a framing of the normal bundle to D in V, $(e_1, \dots, e_{2n-1}, f_1, \dots, f_{2n-1})$ such that over γ_P the e_i give a framing for the normal bundle of γ_P in P and likewise for the f_j over γ_Q. The obstruction to being able to make such a normal framing lies in $\pi_1(SO(2n-2))/\pi_1(SO(n-1) \times SO(n-1))$, which vanishes for $n \geqslant 3$.

Having made these three steps, the normal framing allows us to identify a neighborhood of D in V with a universal model: the product of a lens-shaped region in $\mathbb{R}^2$ with $\mathbb{R}^{n-1} \times \mathbb{R}^{n-1}$, where P and Q correspond to the two boundary arcs of the lens multiplied by the different $\mathbb{R}^{n-1}$ factors. In this model it is then easy to write down the desired isotopy, supported in a small neighbourhood of D; pulling one arc of the lens boundary across the other. The edifice of high-dimensional manifold theory (developed principally in the 1950's and 1960's) is in large part built on this Whitney argument. (Of course this edifice relies in an essential way on other developments lying closer to algebraic topology and homotopy theory, namely cobordism, bundle theory and characteristic classes; there is also the huge body of work on the relation between different categories of manifolds, but this article does not attempt to be a thorough account – apart from anything else because of the author's meager knowledge of all these subjects.) On the one hand, there is the system of ideas revolving around handlebody decompositions. On the other hand, there is the theory of surgery. Both of these constructions can be seen in the framework of Morse theory. Thus, if $f : W^{n+1} \to \mathbb{R}$ is a Morse function we can consider for each regular value $\tau \in \mathbb{R}$ the level set $V^n_\tau = f^{-1}(\tau)$ and the manifold-with-boundary $W^{n+1}_\tau = f^{-1}((-\infty, \tau])$, with $\partial W_\tau = V_\tau$. Let τ_0 be a critical value with just one associated critical point, of index p say – that is, the Hessian of f at the critical point has a maximal negative subspace of dimension p. Then as the parameter τ increases through the critical value τ_0 we have the following, locally standard, descriptions.

1. The level set V_τ changes by a *surgery* on a $(p-1)$ sphere: so $V_{\tau+\varepsilon}$ is obtained from $V_{\tau-\varepsilon}$ by cutting out a tubular neighbourhood $D^{n-p+1} \times S^{p-1} \subset V_{\tau-\varepsilon}$ and replacing it by $S^{n-p} \times D^p$ which has the same boundary:

$$\partial\big(D^{n-p+1} \times S^{p-1}\big) = \partial\big(S^{n-p} \times D^p\big) = S^{n-p} \times S^{p-1}.$$

2. The manifold-with-boundary W_τ changes by attaching a p-handle: so $W_{\tau+\varepsilon}$ is obtained from $W_{\tau-\varepsilon}$ by attaching the handle $D^p \times D^{n+1-p}$ along a neighbourhood of the $(p-1)$-sphere considered before in the boundary $V_{\tau-\varepsilon}$.

Given any closed $n+1$ manifold W we can always choose a Morse function on W and get a description of the manifold as a handlebody, by successively attaching handles as above. In fact, we get two such descriptions, one using f and one using $-f$. (At the level of homology this gives a proof of the Poincaré duality theorem, essentially the same as that via dual cell complexes.) One fundamental problem is to simplify the description as far as possible, subject to the constraints given by the homology groups. If $n \geqslant 5$ and the manifold is simply connected one of the main results of Smale [18, 19] asserts that this can be done. Suppose for simplicity that the homology groups of W have no torsion, then Smale constructs a handle decomposition/Morse function with the minimal number b_p of handles/critical points of index p, where b_p is the p-th Betti number of W. The crucial point is this: if we start with a Morse function having excess critical points one reduces to the case when there are critical points a, b of adjacent indices $p, p+1$ with $f(a) < f(b)$ (and with no other critical values in the interval $[f(a), f(b)]$) which "cancel" from the point of view of homology. The meaning of this is that if we choose an intermediate level V_τ with $\tau \in (f(a), f(b))$ the p-sphere in V_τ used to attach the handle belonging to b has homological intersection number ± 1 with the $(n-p)$-sphere used to attach the handle belonging to a when we "turn the picture upside down" and replace f by $-f$ in our handle decomposition. Now in this dimension range the Whitney disc argument allows us to deform these spheres inside the manifold V_τ to have a single transverse intersection point, and then one can go back to change f, in the neighbourhood of a path from a to b, to remove these excess critical points.

A particular case of the discussion above is when W has the homology of the $(n+1)$-sphere: Smale obtains then a Morse function with only two critical points from which a proof of a version of the generalised Poincaré conjecture follows immediately: a differentiable manifold homotopy equivalent to S^m, $m \geqslant 6$, is homeomorphic to S^m. (The result is true for $m = 5$, with a more complicated proof.) A variant for manifolds-with-boundary gives the h-cobordism theorem: if V_0, V_1 are simply-connected n-manifolds, $n \geqslant 5$, and W is a cobordism from V_0 to V_1 which retracts onto both V_0 and V_1, then W is diffeomorphic to a cylinder and V_0 and V_1 are diffeomorphic. The point of this, of course, is that the existence of any cobordism from V_0 to V_1 is a well-understood problem, following Thom, and among all the cobordisms the h-cobordisms can be readily detected algebraically, via their homology.

Similar issues arise in the other fundamental construction of surgery. If P is a $(p-1)$-sphere embedded in an n-manifold V with trivial normal bundle then one can perform a surgery as above, cutting out a neighbourhood $S^{p-1} \times D^{n-p+1}$ of P in V and replacing it by $S^p \times D^{n-p}$. (The construction may depend on a choice of a specific trivialisation of the normal bundle in certain dimension ranges.) Under suitable conditions the new n-manifold will have smaller homology than the original one. By Poincaré duality it suffices to work in dimensions $p - 1 \leqslant n/2$. Suppose we know that a particular homology class can be represented by some map of the $(p-1)$-sphere (i.e. lies in the image of the Hurewicz homomorphism), then in the critical case of the middle dimension when $p - 1 = n/2$ we can use the Whitney argument to obtain an embedded sphere. (For smaller p this comes immediately from general position.) The normal bundle condition can be attacked via bundle theory and characteristic classes.

This sketch scarcely does justice to the vast body of work on high-dimensional manifold topology, but perhaps at least gives an inkling of the reasons for the differences between low and high dimensions. As an illustration of the many concrete classification results which have been obtained let us quote a result of Smale. Any compact, 2-connected, 6-manifold is diffeomorphic to either S^6 or a connected sum of copies of $S^3 \times S^3$. This statement is the obvious generalisation one might hope for starting from the classification of surfaces (although of course the proof is much harder). More generally, Wall [20] gave a classification of $(m-1)$-connected $2m$-manifolds for $m \geqslant 3$. Optimistically, one might expect that something similar happens in the intermediate case when $m = 2$, that is, of simply connected 4-manifolds, but – as we shall describe in the next section – the picture there is quite different.

3. Four dimensions and symplectic topology

The algebraic topology of a simply connected 4-manifold X is straightforward: everything is determined by the intersection form on $H_2(X)$, a nondegenerate quadratic form over the integers. The naive guess would be that the manifold classification problem should follow the algebraic classification of forms. One of the first authors to discuss this issue explicitly was Milnor [12]. The examples furnished by complex algebraic surfaces were prominent from this time; in particular Milnor discussed the topology of what are now called $K3$ *surfaces*. A $K3$ surface is, by definition, a compact complex surface with first

Betti number zero and with trivial canonical bundle. All such are diffeomorphic and have intersection form $E_8 \oplus E_8 \oplus H \oplus H \oplus H$, where E_8 is the negative definite rank 8 form associated to the roots of the corresponding Lie algebra and H is the standard even rank 2 form $(x, y) \mapsto 2xy$. The bundle theory relevant to 4-manifolds, and questions such as the existence of plane fields and almost complex structures, was developed by Hirzebruch and Hopf. One easy result is that a simply-connected 4-manifold has a spin structure if and only if the intersection form is even. A much deeper fact is Rohlin's theorem [16]: the signature of a spin 4-manifold is divisible by 16 whereas, as a matter of algebra, the signature of an even form need only divisible by 8 – witness the form E_8. For example, the $K3$ surface is spin and has signature -16. There cannot be a smooth simply connected 4-manifold with intersection form E_8, nor, more generally, any sum of an odd number of copies of E_8 and copies of H.

Four-dimensional manifolds can be studied via handle decompositions and a detailed "calculus" for manipulating these was developed by Kirby and his school, through which experts achieve a remarkable ability to visualise 4-dimensional topology. However, in this dimension where the Whitney argument does not apply, general results were hard to come by. Substantial progress came in the early 1980's when, on the one hand, Freedman [5] pushed techniques from high-dimensional topology through to achieve a complete classification of simply connected *topological* 4-manifolds up to *homeomorphism*. As well as the general machinery of the h-cobordism theorem, etc., Freedman's theory used a novel infinite construction, starting with work of Casson, to surmount the difficulties caused by the unwanted intersections of Whitney discs. Freedman's result (incorporating a refinement of Quinn) can be stated as follows: for every odd quadratic form Q there are exactly two homeomorphism classes of 4-manifolds, X_Q, Y_Q say, where $X_Q \times \mathbb{R}$ can be given a smooth structure but $Y_Q \times \mathbb{R}$ cannot; while for every even form there is just one homeomorphism class Z_Q, where $Z_Q \times \mathbb{R}$ can be given a smooth structure if and only if the signature of Q is divisible by 16. A particular case of Freedman's result is the proof of the *topological* version of the 4-dimensional Poincaré conjecture.

On the other hand, starting in the early 1980's, new techniques for studying *smooth* 4-manifolds were found which depended in an essential way on differential geometry and analysis. A landmark in the mid 1990's was the introduction of the Seiberg–Witten equations [22] which seem, at the time of writing, to have brought these developments – as far as the foundations go – into fairly final form. There are a number of surveys of these developments (for example, [3, 4, 6]), so we prefer not to rehearse the details here, but suffice it to say that for any 4-manifold X the Seiberg–Witten theory defines a collection of "basic classes" $\kappa \in H^2(X)$ (along with coefficients n_κ). These are differentiable invariants but definitely not homeomorphism invariants: there are now in the literature hosts of examples of (simply connected) smooth 4-manifolds with the same intersection forms, hence homeomorphic according to Freedman, which are distinguished differentiably by their Seiberg–Witten invariants. The invariants are defined by counting in a suitable sense the solutions of a partial differential equation over X involving a connection on an auxiliary bundle and a spinor field: the equation requires the choice of a Riemannian metric on the 4-manifold, although the "number" of solutions does not in the end depend on this choice, somewhat in the mould of the Hodge theory. Similarly, arguments with these equations give obstructions, beyond Rohlin's theorem, to the realisation of forms by smooth manifolds, i.e. to smoothing many of Freedman's manifolds.

The Seiberg–Witten invariants, viewed as preferred cohomology classes defined by the smooth structure of the 4-manifold, are slightly reminiscent of characteristic classes in classical manifold theory and there are indeed important connections with the first Chern class $c_1(TS)$ in the case of a complex algebraic surface S: the first Chern class is a basic class (with multiplicity 1) and, outside a small range of examples which can be listed explicitly, there are no other basic classes of S. The $K3$ surface is an important example in all of this theory: it appears in a sense as the "simplest" 4-manifold, despite its apparently complicated topology. This is connected with the fact that the $K3$ surface admits a hyper-Kähler metric, and with its very high degree of symmetry: the diffeomorphisms of the $K3$ surface realise an index 2 subgroup of the group of isometries of the intersection form.

Important results of Taubes extend the discussion of the invariants of complex algebraic surfaces to general symplectic 4-manifolds. If (X, ω) is a symplectic 4-manifold there is a unique homotopy class of compatible almost-complex structures so we have a first Chern class c_1 and again this is a basic class with multiplicity 1, so in particular the Seiberg–Witten invariants of X are nontrivial. In this case, however, there can be many additional basic classes. According to Taubes these are related to the "Gromov invariants" defined by counting the complex curves in X with respect to an almost-complex structure.

Despite all this progress, the classification of smooth 4-manifolds (including the 4-dimensional smooth Poincaré conjecture) remains a mystery; likewise for the circle of questions involving the relation between the three classes

$$\text{Complex algebraic surfaces} \subset \text{Symplectic 4-manifolds} \subset \text{4-manifolds}.$$

In general, one does not know if the phenomena detected by the Seiberg–Witten equations are the complete story or whether there is more to say. We can illustrate some of these points by returning to the Lefschetz pencils of Section 1. We have seen that a complex algebraic surface can be described by such a pencil and that this leads to a collection of vanishing cycles δ_i^- in a Riemann surface C_-. These are unique modulo an equivalence relation generated by certain "moves" that one can describe. This data, a collection of circles in a Riemann surface, is closely related to the data specifying a handle decomposition of the complex surface. The vanishing cycles determine the monodromy of the pencil and hence the whole structure of the fibration. More generally, one may consider topological Lefschetz fibrations, determined by any collection of loops δ_i in a Riemann surface subject to the condition that the product of their Dehn twists is the identity. As we have seen in Section 1 there are examples of these which do not come from complex surfaces, but according to unpublished work of Gompf, the total space always admits a symplectic structure. Thus one point of view on the comparison between symplectic 4-manifolds and complex algebraic surfaces is to say that one is asking what are the special properties of the patterns of vanishing cycles coming from genuine algebro-geometric fibrations. The hard Lefschetz theorem gives one such special property, but there must be many more. The difficulty of these problems arises from the fact that while the Picard–Lefschetz theory has to do with the homology classes of the vanishing cycles we are now asking about the full isotopy classes which are much less tractable.

When we go to higher dimensions the differential topology of algebraic varieties comes under firm control, using the techniques discussed in Section 2. From our present point of view this amounts to saying that there is little extra differential topological information, beyond homology, in the patterns of vanishing cycles which arise: they can be pulled around by the Whitney argument. The same is not however true when we consider *symplectic topology*, in which "new" phenomena – loosely speaking, beyond algebraic topology – occur in all dimensions. One can show that the vanishing cycles are naturally *Lagrangian submanifolds* of the fibre Y_t and it is reasonable to expect that questions about the symplectic topology of the total space Z have to do with the problem of removing excess intersection points of these Lagrangian vanishing cycles by *symplectic* isotopies. This type of Lagrangian intersection problem has been the subject of tremendous developments in the past 20 years, along with the whole subject of symplectic topology. A particularly important case occurs when one submanifold is the graph of a symplectic map and the other is the diagonal, so the intersection points are the *fixed points* of the map. Much of the progress in this area springs from the "Arnold conjecture" made in [1]. For simplicity consider a compact symplectic manifold M with trivial first homology; then the conjecture asserts that if f is a symplectic map which is symplectically isotopic to the identity then the number of fixed points of f (assuming these are transverse) cannot be less than the sum of the Betti numbers of M. By contrast, for a general map which is merely isotopic to the identity the best lower bound is given by the Euler characteristic, which can be strictly smaller. A general result in this direction was proved by Floer, and the problem was one of the main motivations for his development (in the late 1980's) of the theory of Floer homology, which also bears on more general Lagrangian intersections (see, for example, the papers and references in [9]). To sum up we can say that while the past century has seen vast developments in our understanding of manifolds there are still fundamental questions about which almost nothing is known, and a notable feature of many current developments is that they are enriched by contact with other branches of geometry and analysis in the spirit of Poincaré original conception. Indeed, to bring our story full circle, we can observe that the origins of Arnold's conjecture, and, hence, of much of the current activity in symplectic topology, go back to the last paper published by Poincaré [15], in which – motivated by the search for closed orbits in Hamiltonian systems – he discusses fixed points of symplectic (i.e. area-preserving) maps from an annulus to itself.

Bibliography

[1] V. Arnold, *Classical Mechanics*, Springer, Berlin (1978).

[2] P. Deligne, P. Griffiths, J. Morgan and D. Sullivan, *Real homotopy theory of kahler manifolds*, Invent Math. **29** (1975), 245–274.

[3] S. Donaldson, *The Seiberg–Witten equations and 4-manifold topology*, Bull. Amer. Math. Soc. **33** (1996), 45–70.

[4] S. Donaldson and P. Kronheimer, *The geometry of four-manifolds*, Oxford Univ. Press, Oxford (1990).

[5] M. Freedman, *The topology of four-manifolds*, J. Differential Geom. **17** (1982), 357–454.

[6] R. Friedman and J. Morgan, *Algebraic surfaces and four-manifolds: Some conjectures and speculations*, Bull. Amer. Math. Soc. **18** (1988), 1–19.

[7] P. Griffiths and J. Harris, *Principles of Algebraic Geometry*, Wiley, New York (1978).

[8] W. Hodge, *Harmonic Integrals*, Cambridge Univ. Press, Cambridge (1940).

[9] Hofer, Taubes, Weinstein and Zehnder (eds), *The Floer Memorial Volume*, Birkhäuser, Basel (1995).

[10] K. Lamotke, *The topology of complex projective varieties after S. Lefschetz*, Topology **20** (1981), 15–51.

[11] S. Lefschetz, *L'Analyse Situs et la Géometrie Algébrique*, Gauthier-Villars, Paris (1924, 2nd ed. 1950).

[12] J. Milnor, *On simply-connected four-manifolds*, Symp. Int. Top. Algebraica, Mexico (1958).

[13] E. Picard and G. Simart, *Théorie des Fonctions Algébriques de Deux Variables, Vol. I*, Paris (1897).

[14] H. Poincaré, *Analysis Situs (with five complements)* See *Oeuvres de Henri Poincaré, T. VI*, Gauthier-Villars, Paris (1955) and J. Ecole Polytech. **1** (1895), 1–121.

[15] H. Poincaré, *Sur un théoreme de géometrie*, Rendiconti del Circolo matematico di Parlemo **33** (1912), 375–407.

[16] V. Rohlin, *New results in the theory of 4-dimensional manifolds*, Dokl. Akad. Nauk USSR **84** (1952), 221–224 (in Russian).

[17] C. Rourke and B. Sanderson, *Introduction to Piecewise-Linear Topology*, Springer, Berlin (1972).

[18] S. Smale, *The generalised Poincaré conjecture in dimensions greater than four*, Ann. of Math. **74** (1961), 391–406.

[19] S. Smale, *A survey of some recent developments in differential topology*, Bull. Amer. Math. Soc. **69** (1963), 133–145.

[20] C. Wall, *Classification of* $(n-1)$*-connected* $2n$*-manifolds*, Ann. of Math. **75** (1962), 163–189.

[21] H. Whitney, *The self-intersections of a smooth* n*-manifold in* $2n$*-space*, Ann. of Math. **45** (1949), 220–246.

[22] E. Witten, *Monopoles and 4-manifolds*, Math. Res. Lett. **1** (1994), 769–796.

This Page Intentionally Left Blank

CHAPTER 15

3-Dimensional Topology up to 1960

C.McA. Gordon

Department of Mathematics, The University of Texas at Austin, Austin, TX 78712-1082, USA
E-mail: gordon@math.utexas.edu

1. Introduction

In this paper we discuss the development of 3-dimensional topology, from its beginnings in the 1880's, up until roughly 1960. The decision to stop at 1960 was more or less arbitrary, and indeed we will sometimes briefly describe developments beyond that date. Our account is very much in the nature of a survey of the literature (an internal history, if you will), an approach which is feasible because the literature is so finite. (This continues to be true through the 1960's, when the number of people working in 3-dimensional topology was still relatively small. During the last twenty years or so, not only has the actual literature grown tremendously, but the number of major themes in the subject has also increased.)

The early papers that deal with 3-manifolds are few: Poincaré's *Analysis Situs* [79] and his fifth complement to that paper [81], Heegaard's dissertation [44], Tietze's Habilitationschrift [110], and the paper of Dehn [22], is almost a complete list up to the end of the First World War, and one or two short papers of Alexander, together with Kneser's paper [56], then take us through the next decade. The 1930's saw an increase in activity, with the work of Reidemeister, Seifert, Seifert and Threlfall, and others, in Germany, and, in England, J.H.C. Whitehead, but this ended with the Second World War, and not much more appeared until the 1950's, when we find Moise's proof of the existence and uniqueness of triangulations [65], Papakyriakopoulos' proof of Dehn's lemma and the sphere theorem [72], and, at the end of the decade, Haken's use of normal surfaces to solve the knot triviality problem [42].

This last is an instance where it would be artificial to try to separate knot theory from 3-dimensional topology (the discussion of Dehn surgery in [22] is another), but in general we have ignored papers that deal specifically with knots, such as Dehn's 1914 paper [23]. Another topic that we do not discuss is "wild" topology.

My warmest thanks go to Hélène Barcelo for help with Poincaré's French, to Gerhard Burde for the useful information he gave me about the German school and its literature, and especially to Cynthia Hog-Angeloni, for generously providing English translations of several German papers. Thanks also to Wolfgang Metzler and Alan Reid for helpful comments on the original manuscript.

HISTORY OF TOPOLOGY
Edited by I.M. James

2. Beginnings

Possibly the first attempt at a systematic approach to the study of 3-manifolds is contained in a short note by Walther Dyck in the Report of the 1884 Meeting, held in Montreal, of the British Association for the Advancement of Science [27]. He says his goal is to classify 3-manifolds:

> The object is to determine certain characteristical numbers for closed threedimensional spaces, analogous to those introduced by Riemann in the theory of his surfaces, so that their identity shows the possibility of its 'one–one geometrical correspondence'.

He offers the following method of construction of 3-manifolds:

> We cut out of our space[1] $2k$ parts, limited by closed surfaces, each pair being respectively of [genus] $p_1, p_2, \ldots, p_k$. Then, by establishing a mutual one–one correspondence between every two surfaces, we close the space thus obtained.

The "characteristical numbers" he had in mind are the genera $p_1, p_2, \ldots, p_k$ of the surfaces and "the manner of their mutual correspondence".

Presumably Dyck's construction was suggested by the fact that any closed orientable surface can be obtained by removing an even number of disjoint disks from S^2 and identifying the resulting boundary components in pairs. Perhaps it was also this analogy that led him to make the rather rash claim that

> we can form all possible threedimensional spaces by [this] procedure.

Dyck notes that his construction gives a 3-manifold containing nonseparating surfaces, and closed curves "which can neither be transformed into each other, nor be drawn together into one point".

To illustrate his remark about closed curves, Dyck gives as examples the two 3-manifolds obtained by removing a pair of solid tori from S^3 and identifying the resulting boundaries, firstly, so that meridians are identified with meridians and latitudes with latitudes,[2] and secondly, so that meridians are identified with latitudes and vice versa. (These manifolds are, respectively, the connected sum of two copies of $S^1 \times S^2$, and $S^1 \times S^2$.) He points out that in the first case a meridian of one of the solid tori cannot be shrunk to a point, while in the second case it can.

Why was Dyck interested in 3-manifolds? He says his motivation was "... certain researches on the theory of functions, ...", and also mentions the theory of Abelian integrals. Poincaré is more explicit. In the introduction to his 1895 *Analysis Situs* paper [79], (which we discuss below), he gives three examples to justify his interest in manifolds of dimension greater than 2. (Note that when Poincaré talks about n-dimensional *Analysis Situs* he means the study of $(n-1)$-manifolds in $\mathbb{R}^n$.)

> The classification of algebraic curves into genera depends, after Riemann, on the topological classification of real closed surfaces. An immediate induction makes us understand that the classification of algebraic surfaces and the theory of their birational

[1] Dyck explains that by this he means S^3, the one point compactification of $\mathbb{R}^3$.

[2] These suggestive terms for the two obvious isotopy classes of curves on the boundary of a solid torus became standard. Somewhere along the way, however, (possibly first in [121]), "latitude" mistakenly became "longitude". In this article we will revert to the original terminology.

transformations is intimately connected with the topological classification of real closed surfaces in 5-dimensional space.

Again, ..., I have used ordinary 3-dimensional *Analysis Situs* in the study of differential equations. The same researches have been pursued by Dyck. One sees easily that generalized *Analysis Situs* would allow one to treat in the same way equations of higher order, and, in particular, those of celestial mechanics.

Jordan has determined analytically the groups of finite order contained in the linear group of n variables. Klein has earlier, by a geometric method of rare elegance, solved the same problem for the linear group of two variables. Could one not extend Klein's method to the group of n variables, or even to an arbitrary continuous group?

Finally, Heegaard, in the preface to his dissertation [44], explains the motivation for his investigations:

The theory of functions with one independent variable is very closely connected with the theory of algebraic curves. The geometry of such a curve becomes therefore of fundamental importance.

He recalls that one approach to this was "the topological examinations of the Riemann surfaces that represent the algebraic curve". He goes on to say:

The transformations of algebraic surfaces play an analogous role in the theory of functions of two variables,

but regrets that although there has been some attempt to generalize "the Riemann–Betti theory of connectivity numbers" to higher-dimensional manifolds (mentioning Picard, Poincaré and Dyck in this connection), "a completely satisfactory account is nowhere to be found". Therefore, he says, before embarking on this approach, "we need a theory of correspondence of manifolds of dimension greater than 2".

These are some of the considerations that provided the impetus for the study of manifolds of dimension greater than 2. It was only natural that the first case, of dimension 3, should receive special attention.

3. Poincaré's Analysis Situs

Three-dimensional topology was really born in Poincaré's foundational paper [79], published in 1895, (the results were announced in 1892 [77]), where we find it inextricably linked with the origins of topology in general. Paper [79] introduces manifolds, homeomorphism, homology, Poincaré duality, and the fundamental group, and in it 3-manifolds appear as examples, both to illustrate these general concepts and also with which to test the strength of the topological invariants (the Betti numbers and the fundamental group) that Poincaré has defined.

The first mention of 3-manifolds in [79] is to illustrate Poincaré's definition of the Betti numbers of a manifold V. Having defined homology in V in terms of m-submanifolds bounding $(m+1)$-submanifolds, he then explains that homologies can be combined in the same way as ordinary equations, and defines the m-th Betti number $\beta_m(V)$ to be the maximal number of linearly independent m-dimensional submanifolds of V.[3] To "clarify

[3] Actually Poincaré works with $P_m = \beta_m + 1$, but we will adopt the modern convention (which in [95] is attributed to Weyl).

these definitions", Poincaré considers a submanifold V of $\mathbb{R}^3$ bounded by n disjoint closed surfaces $S_1, \ldots, S_n$, and asserts that

$$\beta_2(V) = n - 1, \qquad \beta_1(V) = \frac{1}{2}\sum_{i=1}^{n} \beta_1(S_i),$$

mentioning as particular examples the region bounded by a sphere, the region between two spheres, the region bounded by a torus, and the region between two tori. In these formulae we see early hints of Poincaré–Lefschetz duality.

More important for 3-dimensional topology is Poincaré's description of 3-manifolds as being obtained by identifying faces of 3-dimensional polyhedra. Interestingly, Poincaré regards the 3-manifold V itself as being embedded in $\mathbb{R}^4$, but points out that, if it can be decomposed into pieces that are homeomorphic to polyhedra in $\mathbb{R}^3$, in such a way that the intersections of the pieces correspond to faces of the polyhedra, then

> ... the knowledge of the polyhedra P_i and the way their faces are identified provides us, in ordinary space, with an image of the manifold V, and this image suffices for the study of its properties from the point of view of *Analysis Situs*.

He then gives the following five explicit examples of face identifications of a single polyhedron P, four with P being the cube, and one with P an octahedron.

(1) Opposite faces of the cube are identified with no rotation, i.e. by reflection in the parallel plane midway between them.

(2) Two pairs of opposite faces of the cube are identified with an anticlockwise rotation through $\pi/2$, and the third pair with a clockwise rotation through $\pi/2$.

(3) Opposite faces of the cube are identified with an anticlockwise $\pi/2$ rotation. (There is a misprint in the identification of the second pair of faces, but this is clearly what is intended.)

(4) Two pairs of opposite faces of the cube are identified with no rotation, and the third pair with a rotation through π.

(5) Opposite faces of a regular octahedron are identified by reflection in the center of the octahedron.

Poincaré returns to these examples later, but let us note here that (1) is the 3-torus T^3, (3) is quaternionic space, (4) is the T^2-bundle over S^1 with monodromy $\left[\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right]$, and (5) is 3-dimensional real projective space $\mathbb{RP}^3$.

Poincaré explains that a space constructed from a polyhedron P in this way will be a 3-manifold if and only if the link of every vertex is a sphere, and shows, using Euler's formula, how this can be checked from the manner of identification of the faces of P. In particular, this shows that all the above examples except (2) are indeed manifolds.

The most far-reaching discussion in [79] for 3-dimensional topology, however, begins with Poincaré considering the idea of obtaining a 3-manifold as the quotient of a properly discontinuous action of a group G on $\mathbb{R}^3$, relating this to the previous definition by pointing out that such a manifold can be described by identifying suitable faces on the boundary of a fundamental domain. He says:

> The analogy with the theory of Fuchsian groups is too obvious to labour; I will restrict myself to a single example.

Despite this remark, it seems that Poincaré was not aware of any examples of hyperbolic 3-manifolds, although he had already, in his 1883 memoir on Kleinian groups [78] (see [105] for an English translation), described the action of $PSL_2(\mathbb{C})$ on the upper half-space model of hyperbolic 3-space.

At any rate, his "single example" is in fact the infinite family of examples M_A, one for each matrix $A = \left[\begin{smallmatrix}\alpha & \beta\\ \gamma & \delta\end{smallmatrix}\right] \in SL_2(\mathbb{Z})$, the corresponding group G_A being the group of affine transformations of $\mathbb{R}^3$ generated by:

$$\begin{aligned}(x, y, z) &\longmapsto (x+1, y, z),\\ (x, y, z) &\longmapsto (x, y+1, z),\\ (x, y, z) &\longmapsto (\alpha x + \beta y, \gamma x + \delta y, z+1).\end{aligned}$$

Thus M_A is the T^2-bundle over S^1 with monodromy induced by the linear map $A : \mathbb{R}^2 \to \mathbb{R}^2$.

After describing these examples, Poincaré begins the next section with the sentence:

> We are thus led to the notion of the fundamental group of a manifold.

He introduces this with a discussion of how the values of a multi-valued function on a manifold V at a point may change when the point describes a loop in V. Thus the function undergoes a "substitution", the set of which, when we consider all possible loops, forms a group. He then defines the *fundamental group* of V as the group of (based) homotopy classes of loops in V, and states that a group of the first type will always be a quotient of this fundamental group.

Poincaré was very much aware of the importance of the fundamental group. In the 1882 announcement [77] of some of the results which were to appear in [79], he says:

> The group G may thus serve to define the form of the surface [4] and may be called the group of the surface. It is clear that if two surfaces can be transformed one into the other by way of continuous deformation, their groups are isomorphic. The converse, although less evident, is nonetheless true, for closed surfaces, so *that which defines a closed surface from the point of view of* Analysis situs, *is its group.*

By the time he wrote [79], this last claim had been downgraded to a question:

> It would be very interesting to treat the following questions:
> 1. Given a group G defined by a certain number of fundamental equivalences, can it give rise to a closed n-dimensional manifold?
> 2. How can one construct this manifold?
> 3. Are two manifolds of the same dimension, which have the same group G, always homeomorphic?
>
> These questions would require difficult studies and long developments. I will not speak of them here.

No doubt Poincaré would have been able to answer his third question if he had thought about it a little longer: examples such as S^4 and $S^2 \times S^2$ would surely have occurred to him. But he had other things to do, and, having put the question aside, he apparently did not return to it. It turns out that there are even nonhomeomorphic 3-manifolds with the same group, namely lens spaces (see Section 7). Nevertheless, in dimension 3 Poincaré's

[4] Recall that by a "surface" Poincaré means an n-dimensional manifold in $\mathbb{R}^{n+1}$.

question is very much to the point: conjecturally, any closed, irreducible 3-manifold, which is not a lens space, is determined by its fundamental group.

Poincaré shows how to derive a presentation for the fundamental group of a 3-manifold obtained by identifying faces of a polyhedron P: there is a generator ("fundamental closed path") for each pair of identified faces, namely the loop defined by joining, by a pair of arcs, a base point in the interior of P to corresponding points in the two faces, and a relation ("fundamental equivalence") for each edge in the manifold, which sets the product of the generators corresponding to the faces around that edge equal to the identity. From the fundamental group, the first Betti number may be calculated simply by abelianizing:

> When one has thus formed the fundamental equivalences, one may deduce the fundamental homologies, which differ only in that the order of the terms is immaterial. The knowledge of these homologies immediately lets one know the Betti number P_1.

Applying this to his earlier examples, he obtains the following presentations for the fundamental groups:

(1) $\langle a, b, c\colon ab = ba,\ ac = ca,\ bc = cb\rangle$; $\beta_1 = 3$.

(3) $\langle a, b, c\colon a^2 = b^2 = c^2,\ a^4 = 1,\ c = ab\rangle$; $\beta_1 = 0$.

He notes that this is a group of order 8, which acts on $\mathbb{R}^4$ (this action is quaternionic multiplication, if we identify the group with $\{\pm 1, \pm i, \pm j, \pm k\}$) so as to leave invariant the cube with faces $x_i = \pm 1$, $1 \leqslant i \leqslant 4$. For this reason he suggests that it might be called the *hypercubic group*.

(4) $\langle a, b, c\colon bc = cb,\ ca = ab,\ b^{-1}a = ac\rangle$; $\beta_1 = 1$.

(5) $\langle a\colon a^2 = 1\rangle$; $\beta_1 = 0$.

Turning to the examples M_A, Poincaré notes that here $\pi_1(M_A) \cong G_A$. He then computes the Betti numbers, showing that

$$\beta_1(M_A) = \begin{cases} 3, & \text{if } A = I, \\ 2, & \text{if trace } A = 2 \text{ and } A \neq I, \\ 1, & \text{otherwise.} \end{cases}$$

There then follows a detailed proof, by a direct group-theoretic argument, that $G_A \cong G_{A'}$ if and only if A and A' are conjugate in $GL_2(\mathbb{Z})$. (The proof distinguishes the three cases, A hyperbolic, elliptic, or parabolic. It is interesting to note that in terms of Thurston's eight 3-dimensional geometries [109], these cases correspond to M_A having a geometric structure modelled on Sol, E^3, and Nil, respectively; see [92, Theorem 5.5].) In particular, Poincaré concludes that there are infinitely many distinct closed 3-manifolds with the same Betti numbers.

Poincaré also remarks that the fundamental groups of his examples (3) and (5) are finite, of orders 8 and 2, respectively, while the group of the 3-sphere is trivial. Thus no two of these manifolds are homeomorphic, but, on the other hand, since their groups are finite, their Betti numbers are zero. In view of this, Poincaré suggests that

> It would seem natural to restrict the meaning of the term *simply connected* and to reserve it for manifolds with trivial fundamental group.

Poincaré wrote five complements to *Analysis Situs*. The first two were in response to the criticisms of [79] by Heegaard, who pointed out, among other difficulties, that Poincaré's duality theorem for the Betti numbers appeared to be false, citing as an example the manifold obtained by gluing together two solid tori in such a way that a meridian of one is identified with a curve that winds twice latitudinally and once meridionally on the other (in other words, real projective space $\mathbb{RP}^3$). Heegaard points out that every 2-cycle in $\mathbb{RP}^3$ bounds, but there is a 1-cycle which does not. The problem is, of course, torsion: Poincaré's definition of the Betti numbers "allows division" [82, Section XVI], in contrast to Betti's definition. At any rate, one of the consequences of this was that Poincaré realized that one could work with homology "without division", and obtain additional invariants, which he called *torsion coefficients*.

In the second complement, [80], Poincaré computes the torsion coefficients of the 3-dimensional manifolds described in [79]. In particular, his examples (3) (quaternionic space) and (5) (real projective space) have the same Betti numbers and torsion coefficients (i.e. the same first homology group, namely $\mathbb{Z}_2$), but have nonisomorphic fundamental groups. Curiously, Poincaré does not explicitly mention this, nor does he note that the manifolds M_A also provide examples of this phenomenon (although not so obviously).

Poincaré does not mention 3-manifolds again in [79], and in the first four complements, their only brief appearance (in the second) is the one we have just mentioned. He returns to them in a big way, however, in his fifth complement, which we discuss in Section 5.

We conclude by remarking that it has become conventional to accuse Poincaré of being obscure and sometimes lacking in rigor, but anyone who does so should reflect that things could have been worse. At the end of the introduction to [79] he says:

> ... my only regret is that [this memoir] is too long; but when I have wanted to restrain myself, I have lapsed into obscurity; I have preferred to be considered a little talkative.

4. The Heegaard diagram

Although the term *Heegaard diagram* eventually acquired a quite specific meaning, Heegaard's original definition of a "diagram" was considerably more general. This is given in his 1898 dissertation [44]. (Because of the influence this work had on Poincaré, a French translation was published in 1916 [45]. An English translation of part of the dissertation has recently been made by A.H. Przybyszewska; see [83]. For a very interesting account of Heegaard's life, see [67].)

In order to investigate the topology of manifolds of dimension greater than 2, Heegaard decides not to take as his model the Riemann–Betti theory of connectivity numbers (as Dyck, Poincaré and Picard had done), but instead to try to generalize the *puncture method* of Petersen "which I recalled from lectures", in which one "puncture[s] the Riemann surface and bring[s] it by continuous deformation into normal form". He explains further:

> The question that we first meet is this: how is one to cut a closed manifold to make it simply connected? To solve this problem we use the following procedure: the manifold is punctured, i.e. a 3-cell neighborhood of a point is removed. Thus a boundary is created, which is enlarged by a continuous deformation so as to remove more and more of the given manifold. We continue in this way until certain parts of the boundary meet others, stopping the deformation in these places when the distance between the parts that are meeting has become infinitely small. In this way we are led to a *diagram* consisting of a system of manifolds of lower dimension than the given one, or rather the neighborhood of this system, i.e. a manifold which is infinitely small in the n-th dimension. The system of lower-dimensional manifolds which constitutes the boundary of the diagram is called the *nucleus*.

Thus the nucleus is an $(n-1)$-dimensional spine of the manifold, and the diagram is a neighborhood of the nucleus, with the cell decomposition of the nucleus as part of the data. In other words, a diagram is essentially a handle decomposition.

Specializing to the 3-dimensional case, Heegaard starts with a 3-manifold obtained by identifying the faces of a polyhedron, the nucleus being the 2-complex resulting from the identifications on the boundary of the polyhedron. The diagram then consists of neighborhoods of the 0-cells, 1-cells, and 2-cells; these neighborhoods he calls *junction spheres*, *strings*, and *plates*.

The boundary of the union of the junction spheres and strings is a surface (which Heegaard allows to be non-orientable) with "connectivity number" $2p+1$,[5] say; Heegaard then states that, if the manifold is closed, there must be p plates, whose "fastening bands" do not disconnect the surface.

Addressing the problem of trying to reduce a diagram to a normal form, Heegaard notes that, in addition to isotopy of the fastening bands, a diagram may be subjected to certain moves, which, expressed in modern terminology, are: 1-handle sliding, 2-handle sliding, and eliminating a cancelling pair of handles. Thus Heegaard has intuitively arrived at the correct equivalence relation between such handle decompositions of 3-manifolds. Although "a lot of simplifications can be done by means of these moves", Heegaard nevertheless concludes that "the problem of reducing the diagram into a normal form is probably very difficult". In this of course Heegaard is also completely correct. Although the search for a "normal form" for 3-manifolds, analogous to that for surfaces, continues to be mentioned in the literature as the ultimate goal, we see that it has quickly become clear that any such normal form, if it exists, will be considerably more complicated than in the 2-dimensional case.

Heegaard next gives some simple examples of diagrams of 3-manifolds. Starting with the case of genus 1, he gives a brief discussion of the simple closed curves on a torus standardly embedded in $\mathbb{R}^3$, noting that in addition to a meridian λ (which bounds a disk "inside" the torus), and a latitude β (which bounds a disk "outside" the torus), there are also curves $[n\beta \pm \lambda]$ and $[\beta \pm n\lambda]$, defined in the obvious way. However, he states that "the complete classification is quite difficult".

Heegaard's next example is the diagram of genus p in which the fastening bands are the meridians of the string surface (the meridian–latitude terminology is extended in the obvious way to handlebodies of arbitrary genus); this manifold is the connected sum of p copies of $S^1 \times S^2$. Regarding it as the double of a handlebody, Heegaard observes that it embeds in $\mathbb{R}^4$ (by embedding the handlebody in $\mathbb{R}^3$, pushing its interior into upper half 4-space, and doubling), and also that it can be obtained by removing $2p$ disjoint 3-cells from S^3 and identifying the resulting boundaries in pairs (recall Dyck's construction [27]), in a way analogous to Klein's normal form for surfaces. Finally, Heegaard describes a genus 3 diagram of the 3-torus T^3, which he defines as the boundary of a neighborhood (the *hull*) of a torus T^2 embedded in $\mathbb{R}^4$, and notes that a similar diagram may be obtained for the hull of any surface in $\mathbb{R}^4$; this will be an orientable S^1-bundle over the surface.

Heegaard observes that associated with a given diagram, there is a second diagram, corresponding to the dual handle decomposition:

> There is a sort of dual connection between the two diagrams: the strings in one of them correspond to the plates in the other, and vice versa.

---

[5] In the sense of Betti, i.e. with respect to homology "without division"; see Section 3.

Finally, he points out that a diagram expresses the manifold as the union of two solid handlebodies, with their boundaries identified in some way, and for this:

> ... it is sufficient to know [on one boundary] the system of nondisconnecting annular cuts which corresponds to the curves β on the string-surface of the other, and the system which corresponds to the curves λ.

In fact there is a certain amount of redundancy here: the manifold is actually determined by the images on the boundary of one handlebody of the meridian curves λ of the other handlebody. Thus a *Heegaard diagram* eventually came to mean two *complete systems* of curves on a closed, orientable surface F of genus p, where a complete system is a disjoint union of p simple loops whose union does not separate F; see, for example, [97].

The difficulties in using Heegaard diagrams to get "normal forms" for 3-manifolds became increasingly clear. The classification of genus 1 diagrams is relatively easy, and is done in [39], but the inherent complexity of diagrams of higher genus, even of S^3, was explicitly pointed out by Frankl [34] and Reidemeister [84]. Specifically, they gave examples of Heegaard diagrams of S^3, consisting of a complete system of curves $K_1, \ldots, K_p$ on the boundary of a handlebody V of genus p (where $p = 3$ and 2, respectively), such that the manifold X obtained by adding a 2-handle to V along K_1 is not a handlebody. In Reidemeister's case ($p = 2$), X is the complement of the trefoil knot, and he points out that any knot that arises in this way will have the property that its group has a presentation with two generators and one relation. (In modern terminology, the knots in question are precisely those with *tunnel number* 1.)

In addition to the problem of analyzing different diagrams of the same underlying *Heegaard splitting*, i.e. the pair (M, F), where the *Heegaard surface* F separates M into two handlebodies, there is also the problem of analyzing different splittings of the same manifold. That this was a problem, even for S^3, was pointed out by Reidemeister in [84] (see Section 5), and Alexander gives the following discussion of these matters in his elegant paper [9] in the Proceedings of the 1932 International Congress of Mathematicians:

> One or two general remarks about the classification of manifolds according to Heegaard's program may, perhaps, be worth making. The problem divides itself naturally into two parts: (i) to determine in how many essentially different ways two canonical regions[6] of genus p can be matched together to form a manifold; (ii) to determine in how many essentially different ways a canonical region can be traced in a manifold. The first part of the problem does not seem hopelessly difficult; it is closely related to the problem of the number of essentially different one–one mappings of one surface of genus p on another. As to the second part of the problem, I have a strong suspicion that if S and S^1 are two canonical surfaces of the same genus in a manifold M then there is always a continuous deformation of the manifold M into itself carrying the surface S into the surface S^1. It would be interesting to have a proof of this hypothetical theorem even for the case where the manifold M is a hypersphere. The theorem for a general manifold M seems to be reducible to this special case.

With hindsight, this seems overly optimistic, with regard to both parts (i) and (ii). It is in fact true that all Heegaard surfaces of S^3 of a given genus are isotopic; this was proved by Waldhausen in 1968 [114]. However it is false for arbitrary 3-manifolds; the first examples, for connected sums of lens spaces, were given by Engmann [31], and, for irreducible 3-manifolds, by Birman, González-Acuña, and Montesinos [14].

[6] By a "canonical region" Alexander means a handlebody, and by a "canonical surface", a Heegaard surface.

Regarding part (i) of Alexander's comments, it does seem to be the case that it was the desire to understand 3-manifolds by means of their Heegaard diagrams that provided the initial motivation for the study of automorphisms of surfaces, by Poincaré (see Section 5), Dehn, Goeritz, and others.

A Heegaard splitting of genus p may be *stabilized* in a trivial way to give a splitting of genus $p+1$; this is the inverse of the handle cancellation observed by Heegaard. Reidemeister [84] and Singer [100] showed that any two Heegaard splittings of a given 3-manifold are stably equivalent, i.e. become isotopic after each is stabilized some number of times. This result was subsequently used by Reidemeister [85] to define certain linking invariants of 3-manifolds.

Reidemeister's proof of the stable equivalence theorem is rather sketchy, while Singer's, although quite detailed, contains a gap. The first correct published proof seems to be the one given by Craggs in [19]. For an interesting account of the proofs of Reidemeister and Singer, their difficulties, and how they can be made rigorous, see Siebenmann [99].

We have seen that Heegaard's diagrams for n-manifolds were motivated by the topological classification of 2-manifolds. Another 2-dimensional phenomenon that prompted the investigation of its higher dimensional analog was the fact that every (closed, orientable) 2-manifold is homeomorphic to a Riemann surface, that is, a branched covering of the 2-sphere. This led to the study of 3-dimensional *Riemann spaces*, in other words, branched coverings of the 3-sphere, the branch set being some link.

Heegaard discusses this in his dissertation [44, Section 13]. Assuming that the branching index around each branch curve in the manifold is 2 (or 1), he shows how to construct a diagram of the 3-manifold from the covering data. (In an earlier section, Section 8, he has done this for Riemann surfaces.) In Section 14 he gives some examples, of n-sheeted coverings M of S^3 with branch set L:

(1) $n = 2$, $L =$ unknot: $M \cong S^3$.
(2) $n = 2$, $L =$ 2-component unlink: $M \cong S^1 \times S^2$.
(3) $L = \nu$-component unlink: $M \cong \#_{\nu-n+1} S^1 \times S^2$ ("a sphere with $\nu - n + 1$ handles").
(4) $n = 3$, $L =$ trefoil: $M \cong S^3$.
(5) $n = 2$, $L =$ trefoil: $M \cong L(3, 1)$.
(6) $n = 2$, $L =$ Hopf link: $M \cong L(2, 1) \cong \mathbb{RP}^3$.

Heegaard then applies these considerations to the subject that motivated his whole investigation, namely the study of complex algebraic surfaces. (Recall that the title of his dissertation is "Preliminary studies towards a topological theory of connectivity of algebraic surfaces".) If p is a singular point of such a surface X, and M is the 3-manifold that is the intersection of X with the 5-sphere boundary of a neighborhood of p in $\mathbb{C}^3$, then Heegaard observes that the corresponding neighborhood of p in X is homeomorphic to the cone on M. After giving some example where $M \cong S^3$, he shows that for, e.g., the curve $z^2 = x^2 - y^2$, and p the origin, the manifold M is homeomorphic to $\mathbb{RP}^3$, and so X is not a manifold near p.

Tietze, in [110, Section 18], also gives a discussion of branched coverings of S^3, explicitly mentioning Heegaard's example (4) above, and the example: $n = 3$, $L =$ Hopf link: $M \cong L(3, 1)$. He says:

> ... it is not known if each closed, orientable 3-manifold is homeomorphic to a "Riemann space" of this kind.

This was answered by Alexander in [3], in all dimensions: he showed that every closed, orientable n-manifold is a branched covering of the n-sphere. He concludes this short paper with the following remarks:

> In the 3-dimensional case, a Riemann space obtained by the above construction contains, in general, a network of branch lines at each of which two or more sheets coalesce. It is easy to show that, without modifying the topology of the space, the branch system may be replaced by a set of simple, nonintersecting closed curves such that only two sheets come together at a curve. The curves may, however, be knotted and linked.
>
> Three-dimensional Riemann spaces have been discussed by Heegaard and Tietze, but neither of these mathematicians seems to have been aware of their complete generality.

5. Poincaré's fifth complement

Poincaré introduces this remarkable paper [81] with the words:

> I have often had occasion to apply my thoughts to *Analysis Situs*;... I now return to this same topic, convinced that one will be able to succeed only by repeated efforts, and that the subject is important enough to merit such efforts.

He goes on to say:

> The final result that I have in view is the following. In the second complement I have shown that to characterize a manifold, it is not enough to know the Betti numbers, but that certain coefficients which I have called torsion coefficients play an important role.
>
> One may then ask if the consideration of these coefficients suffices; if a manifold all of whose Betti numbers and torsion coefficients are trivial is simply-connected in the proper sense of the word, that is to say, homeomorphic to the hypersphere.
>
> We can now answer this question....

As we have remarked above, in Section 3, Poincaré already had in hand examples of 3-manifolds with the same homology groups but different fundamental groups. However, here he proposes the more specific question of whether a homology sphere is homeomorphic to the sphere. This question had certainly occurred to Poincaré earlier; in fact his second complement concludes with the erroneous announcement that the answer is "yes" [80, p. 308].

The example, of a homology 3-sphere with nontrivial fundamental group, which answers the question, comes at the end of the fifth complement. The rest of the paper is taken up with considerations most of which are not logically necessary for the proof that this example has the desired properties, but which may be described as Poincaré's attempts to set up a theory of (his version of) Heegaard diagrams of 3-manifolds.

The most natural setting in which to express Heegaard's definitions in modern terminology is that of piecewise linear topology. By contrast, Poincaré chose to work in a smooth setting. In a remarkably far-sighted discussion, in [81, Section 2], he considers a Morse function on an m-dimensional manifold V, classifies the critical points in terms of their index, and analyzes the effect on the topology of V of passing through a critical point. (For Poincaré, V is embedded in some Euclidean space $\mathbb{R}^k$, and the Morse function corresponds to a 1-parameter family of $(k-1)$-dimensional "surfaces" $\varphi(t)$, whose intersections with V express V as the union of a 1-parameter family of $(m-1)$-dimensional submanifolds $W(t)$, possibly with singularities.)

In Section 5, specializing to dimension 3, Poincaré considers handlebodies. Specifically, he shows that if V is a 3-manifold with boundary, such that the nonsingular level surfaces $W(t)$ are connected, orientable, and increase their genus at each critical point, then there are p disjoint disks in V such that cutting V along these disks results in a 3-ball. (Thus V is a handlebody of genus p.) He thereby proves that such a manifold V is determined up to homeomorphism by p, the genus of ∂V. He also shows that, if $K_1, \ldots, K_p$ is a set of meridians for V, then $\ker(\pi_1(\partial V) \to \pi_1(V))$ is equal to the normal closure $\langle [K_1], \ldots, [K_p] \rangle$, i.e. the set of products of conjugates of $[K_1]^{\pm 1}, \ldots, [K_p]^{\pm 1}$. (We will return to this in Section 11.)

In the next section, Section 6, Poincaré considers a 3-manifold V in which $W(t)$ is a connected, orientable surface, which reduces to a point at $t = 0$ and $t = 1$, steadily increases in genus at each critical point from $t = 0$ to $t = \frac{1}{2}$, and then steadily decreases in genus from $t = \frac{1}{2}$ to $t = 1$. Then V is the union of two handlebodies V' and V'', whose common boundary is the genus p surface $W = W(\frac{1}{2})$, and on W we see meridians $K_1', \ldots, K_p'$ for V' and $K_1'', \ldots, K_p''$ for V''. Thus we find Poincaré arriving at the concept of a Heegaard diagram by a rather different route. (Although there is no mention of it, it is hard to imagine that Poincaré was not influenced here to some extent by Heegaard's work. Certainly he was familiar with Heegaard's dissertation (recall that it was Heegaard's comments on [79] that prompted Poincaré to write his first two complements to that paper), and Heegaard had even sent him a summary of his dissertation in French [67, Section 6].)

Note, however, that Poincaré does not claim that every closed 3-manifold has a Heegaard splitting. From Poincaré's point of view, this would entail showing that one could rearrange the handles in the handle decomposition determined by the Morse function so that the 1-handles preceded the 2-handles. These considerations may be related to his false assertion in the previous section, [81, p. 90], that every closed surface in $\mathbb{R}^3$ bounds a handlebody, since (he says) it bounds a manifold "susceptible to the same [method of] generation as V" (i.e. so that there are only 1-handles). Ironically, he says that this is very surprising, as

> the various sheets of the surface might be shuffled among themselves in a complicated fashion and might form knots which it is impossible to untie without leaving 3-dimensional space.

Continuing his discussion of a closed 3-manifold V with a Heegaard splitting (V', V''), Poincaré shows that every loop in V can be homotoped into the Heegaard surface W, i.e. $\pi_1(W) \to \pi_1(V)$ is onto, and that any element in $\ker(\pi_1(W) \to \pi_1(V))$ is a product of elements in $\ker(\pi_1(W) \to \pi_1(V'))$ and $\ker(\pi_1(W) \to \pi_1(V''))$. Thus

$$\pi_1(V) \cong \pi_1(W) / \langle [K_1'], \ldots, [K_p'], [K_1''], \ldots, [K_p''] \rangle.$$

Poincaré deduces (by abelianizing) that $H_1(V)$ is the quotient of $H_1(W)$ by the subgroup generated by the homology classes of the two sets of meridians, and hence that a $2p \times 2p$ presentation matrix for $H_1(V)$ may be obtained by taking as its rows the coefficients in the expressions of the K_i' and K_i'' as linear combinations of some standard basis $C_1, \ldots, C_{2p}$ for $H_1(W)$. Letting Δ denote the determinant of this matrix, Poincaré observes that if $|\Delta| > 1$ then the Betti number ("relative to homologies by division") of V is 0; if $|\Delta| = 1$ then both the Betti number and torsion coefficients vanish; and if $\Delta = 0$ then the Betti number is greater than zero.

Focusing on the case $\Delta = \pm 1$, Poincaré says that here one can ask if V is simply connected ("in the proper sense of the word", i.e. homeomorphic to S^3), and goes on:

> We shall see, and this is the principal goal of the present work, that it is not always so, and for this we will restrict ourselves to giving one example.

There follows a description of Poincaré's famous homology 3-sphere with nontrivial fundamental group. This manifold is often referred to nowadays as the Poincaré dodecahedral space, although the construction implied by this name in fact came later. Poincaré defines the manifold V in terms of a genus 2 Heegaard splitting (V', V''), with meridians K_1', K_2' and K_1'', K_2'', where K_1'' and K_2'' are explicitly drawn as unions of arcs on the 4-punctured sphere obtained by cutting the genus 2 Heegaard surface W along K_1' and K_2'. Taking a standard system of curves C_1, C_2, C_3, C_4 on W, with $C_1 = K_1'$, $C_3 = K_2'$, Poincaré writes down the elements of $\pi_1(W)$ represented by K_1'' and K_2'', in terms of C_1, C_2, C_3, C_4, and in this way obtains the following presentation for $\pi_1(V)$

$$\langle a, b\colon a^4ba^{-1}b = 1,\ b^{-2}a^{-1}ba^{-1} = 1\rangle.$$

Abelianizing gives the relations

$$3a + 2b = 0, \qquad -2a - b = 0,$$

for which $|\Delta| = 1$, showing that V is a homology sphere.

On the other hand, adjoining to the above presentation the relation $(a^{-1}b)^2 = 1$, Poincaré obtains the presentation

$$\langle a, b\colon (a^{-1}b)^2 = a^5 = b^3 = 1\rangle$$

of the icosahedral group. Since this group is nontrivial, he concludes that $\pi_1(V)$ is also nontrivial.

Finally, Poincaré says:

> There remains one question to consider:
>
> Is it possible that the fundamental group of V can be trivial, and V still not be simply connected?
>
> *In other words,*[7] is it possible to draw [on W] simple closed curves K_1'' and K_2'', so that [$\pi_1(V)$ is trivial], and that meanwhile [there do not exist pairs of meridians C_1', C_2' and C_1'', C_2'', for V' and V'', respectively, such that $|C_i' \cap C_j''| = \delta_{ij}$]?
>
> But this question would lead us too far afield.

This is the famous Poincaré conjecture. Note, however, that although the general question (is a simply-connected (in the modern sense) 3-manifold homeomorphic to S^3?) is implicit here, in fact the question that Poincaré asks is quite specific: is there a 3-manifold with a Heegaard diagram of genus 2 that is simply-connected and not homeomorphic to S^3?

(As Reidemeister points out in [84, footnote on p. 193], Poincaré's formulation of the question implicitly assumes that any genus 2 Heegaard splitting (V', V'') of S^3 has the property that there exist meridians C_1', C_2' for V' and C_1'', C_2'' for V'' such that $|C_i' \cap C_j''| = \delta_{ij}$, i.e. is equivalent to the standard genus 2 splitting. This turns out to be true, but it was not established until 1968, by Waldhausen [114].)

---

[7] My italics, C. McA. G.

It is clear that in order to arrive at his example of a nonsimply-connected homology sphere, and also in investigating his question, Poincaré must have done a good deal of experimentation with Heegaard diagrams, of genus 2 and presumably higher genus also. In particular, he must have come across many nonstandard diagrams of S^3, and realized that they did indeed represent S^3. Thus he must have been aware that such diagrams can be quite complicated. (The pitfalls here are illustrated by the discussion of Poincaré's example in the Dehn–Heegaard Enzyklopädie article [25]. There, the authors attempt to show that Poincaré's manifold is not homeomorphic to S^3 by a geometric argument, by considering the curves on a standard genus 2 Heegaard surface for S^3 that bound disks in one of the handlebodies. The proof, however, is not valid; in fact the diagram given in their paper is actually a diagram of S^3, as Dehn himself realized soon afterwards [21].)

Poincaré's detailed study of curves on surfaces, in [81, Sections 3 and 4], is also clearly motivated by Heegaard diagram considerations. In Section 3 he shows that, if F is a closed orientable surface, then an automorphism of $H_1(F)$ is induced by an automorphism of F if and only if it preserves the intersection form, and deduces that an element of $H_1(F)$ is represented by a simple loop if and only if it is indivisible. In Section 4 he gives an algorithm, using hyperbolic geometry, for deciding whether or not a loop on F is homotopic to a simple loop, and whether or not two loops are homotopic to disjoint loops.

So his remark: "But this would lead us too far afield", should probably be interpreted as indicating that his investigations of Heegaard diagrams (perhaps specifically of genus 2) were inconclusive, and that, realizing the difficulty of the problem, he decided not to pursue the matter further.

Ironically, although of course the (general) Poincaré conjecture is still open, the genus 2 case was established in 1978, with the proof of the Smith conjecture [66, p. 6].

Having decided that his question would "lead [him] too far afield", Poincaré never returned to the study of 3-dimensional manifolds. In a handful of papers, he truly created the field of topology, and 3-dimensional topology in particular. His achievements are all the more remarkable when one considers how relatively little of his time he devoted to the subject, despite being convinced of its importance. In his analysis of his own scientific works [82], for example, written in 1901, out of a total of 99 pages he devotes just over three to his work in topology; in Hadamard's 85 page account of Poincaré's mathematical work [40], topology gets two pages; and of Poincaré's over 500 publications, a mere dozen or so deal with topology, with 3-dimensional topology featuring in only two or three.

For other accounts of Poincaré's work in 3-dimensional topology see [26, 113].

6. Homology 3-spheres

Poincaré's example of a homology 3-sphere not homeomorphic to S^3 generated a good deal of interest, and the construction of other such 3-manifolds (called *Poincaré spaces* by Dehn in [22]), was for some time an identifiable theme in the literature.

The first general construction was given by Dehn in [21]. The main purpose of this short note was to point out the error in describing Poincaré's example in the Dehn–Heegaard Enzyklopädie article [25], but Dehn also took the opportunity to observe that if two copies of the complement of a knotted open solid torus in S^3 are glued together along their boundaries in such a way that a meridian of each one is identified with a latitude of the other, then the resulting manifold is a homology sphere. On the other hand, Dehn states that such a manifold cannot be homeomorphic to S^3, since it contains a torus (namely the common boundary of the two knot complements) which does not bound a solid torus on either side. Although it is indeed true that every torus in S^3 bounds a solid torus, this was not proved until later, by Alexander [4].

Another construction of Poincaré spaces was given by Dehn in his landmark 1910 paper [22]. Here he shows that, again starting with the complement of a solid toral neighborhood of a knot K in S^3, a solid torus may be attached to it in infinitely many ways (naturally indexed by the integers) so as to obtain a homology 3-sphere. Using his *Gruppenbild*, which was introduced in the same paper, he shows that for K a $(2, q)$-torus knot, the manifolds obtained by this construction all have nontrivial fundamental group (apart from the trivial attachment yielding S^3). In fact, except for a single attachment on the complement of the trefoil, the group is always infinite. In these cases the Gruppenbild is derived from a tessellation of the hyperbolic plane: the group modulo its (infinite cyclic) center is a hyperbolic triangle group. In the one exceptional case, Dehn constructs the Gruppenbild from the 1-skeleton of the dodecahedron, and shows that the group is finite, of order 120. He concludes (incorrectly, as was pointed out in [107, p. 68]) that it is isomorphic to the "icosahedral group extended by reflection". (The latter group maps onto $\mathbb{Z}_2$, while the former, being the fundamental group of a homology sphere, is perfect.) In fact the group in question is the binary icosahedral group, the inverse image of the icosahedral group under the 2-fold covering $S^3 \to SO(3)$. Curiously, Dehn makes no mention here of Poincaré's example, M_{Poin}, or the possible relation between it and his manifold M_{Dehn}.

The third member of this trio, the *spherical dodecahedral space*, M_{dodeca}, say, seems to have been first mentioned by Kneser, in a footnote to his 1929 paper [56, p. 256]. Kneser describes this manifold as the quotient of a fixed point free geometric action of the binary icosahedral group on S^3, and notes that it comes from a tiling of S^3 by 120 cells. He also states that M_{dodeca} is homeomorphic to M_{Dehn}.

In the course of their determination of all 3-dimensional spherical space forms, Seifert and Threlfall [107] also describe M_{dodeca}, and show that it can be obtained from a regular dodecahedron by identifying opposite faces by a rotation through $2\pi/10$. From this they derive a presentation of $\pi_1(M_{\mathrm{dodeca}})$, and show that it may be transformed to both Poincaré's presentation for $\pi_1(M_{\mathrm{Poin}})$ and Dehn's presentation for $\pi_1(M_{\mathrm{Dehn}})$; thus all three manifolds have the same fundamental group. Seifert and Threlfall also show, again from its description as a dodecahedron with face identifications, that M_{dodeca} shares another property with M_{Poin}, namely, that it has a Heegaard splitting of genus 2.

In the text of their paper [107], Seifert and Threlfall state that they do not know whether or not any two of the manifolds M_{Poin}, M_{Dehn}, and M_{dodeca} are homeomorphic. However, in a note added in proof, they mention Kneser's reference to M_{dodeca} in [56], and say that he has shown them how to identify the complement of a certain closed curve in M_{dodeca} with the complement of the trefoil, and hence show that M_{dodeca} and M_{Dehn} are homeomorphic.

Finally, in [118], Seifert and Weber showed that all three manifolds are homeomorphic, by showing that they are all *fibered spaces* in the sense of Seifert [95], and using the results of that paper [95, Theorem 12]. We will discuss this in more detail in Section 9.

In [57], Kreines gave an example of a homology sphere obtained by identifying faces of a tetrahedron; it is also homeomorphic to M_{dodeca}.

The procedure of attaching a solid torus to the complement of a neighborhood of a knot has become known as *Dehn surgery*, and has been the focus of a lot of attention in recent years.

As well as being a source of Poincaré spaces, Dehn regarded his construction as giving a way of showing that a knot K is nontrivial: if the fundamental group of one of the homology spheres obtained by Dehn surgery on K is nontrivial, then it follows that $\pi_1(S^3 - K) \not\cong \mathbb{Z}$, and so K is nontrivial. The converse, i.e. if K is nontrivial, then any manifold obtained by nontrivial Dehn surgery on K has nontrivial fundamental group, is known as the Property P conjecture, and is still unsettled.

7. Lens spaces

As we have seen, Poincaré constructed an infinite family of distinct 3-manifolds with the same Betti numbers, but in retrospect the simplest such family is the lens spaces. These were first defined by Tietze [110], as the simplest possible examples of 3-manifolds obtained by identifying faces of a polyhedron. Namely, the equator of a 3-ball is divided into p equal segments, so that the upper and lower hemispheres become p-sided polygons. These hemispherical faces are then identified by a rotation through $2\pi q/p$, where $0 \leqslant q < p$ and $(p, q) = 1$, giving a 3-manifold $L(p, q)$. If a corner is introduced along the equator of the 3-ball it assumes the lens-shaped appearance that gave these manifolds their name, the term *lens space* being introduced by Seifert and Threlfall in their paper on 3-dimensional spherical space forms [107].

Tietze notes that $L(p, q)$ may also be described as the manifold with a genus 1 Heegaard diagram consisting of a curve on the boundary of a solid torus which winds around p times latitudinally and q times meridionally ($[q\lambda + p\beta]$ in Heegaard's notation; as we have seen, the cases $L(2, 1)$ and $L(3, 1)$ were explicitly considered by Heegaard). For this reason, the lens spaces were originally referred to as *torus manifolds* [39, 56]. Tietze also observes that $L(p, q)$ is p-fold covered by $L(1, 0) \cong S^3$, and has fundamental group $\mathbb{Z}_p$, so that here we have orientable 3-manifolds with finite nontrivial fundamental group, in contrast to the situation in dimension 2.

Finally, Tietze points out that the lens spaces provide interesting examples in the context of the main problem of topology, namely the determination of necessary and sufficient conditions for two manifolds to be homeomorphic. For, Poincaré having shown that a 3-manifold is not determined by its Betti numbers and torsion coefficients, it is now natural to ask if it is determined by its fundamental group. (Earlier in his paper, Tietze had given rigorous proofs that all the then known topological invariants of a closed, orientable 3-manifold are determined by its fundamental group.) Tietze suggests that the lens spaces are potential counterexamples, and in particular raises the question of whether $L(5, 1)$ and $L(5, 2)$ (the first pair of lens spaces with the same fundamental group which are not obviously homeomorphic) are in fact homeomorphic.

In 1919 Alexander [2] showed that indeed they are not, although he seems to be unaware that the question had been raised by Tietze.

Alexander's proof is homological, and goes as follows. The lens space $L(5, 1)$ has a Heegaard diagram consisting of a solid torus A and a (5,1)-curve ℓ on ∂A, i.e. it is the union of A with another solid torus whose meridian winds around A five times latitudinally and once meridionally. Similarly, $L(5, 2)$ is defined by a solid torus A' and a (5,2)-curve ℓ' on $\partial A'$. If there were a homeomorphism from $L(5, 2)$ to $L(5, 1)$, we could assume that it takes A' into the interior of A. Then $H_1(A - A') \cong \mathbb{Z} \oplus \mathbb{Z}$, generated by a meridian a' of

A' and a latitude b of A. If θ denotes the winding number of A' in A, then with respect to this basis

$$[\ell] = \theta a' + 5b, \quad \text{and} \quad [\ell'] = (5k+2)a' \pm 5\theta b$$

for some k, the $-$ sign allowing the possibility that the homeomorphism is orientation-reversing.

Since ℓ' bounds a disk in the complement of A', we must have

$$(5k+2)a' \pm 5\theta b = m\big(\theta a' + 5b\big), \quad \text{for some } m.$$

This readily gives $\theta^2 \equiv \pm 2 \pmod 5$, a contradiction.

More generally, Alexander's proof shows that if $L(p,q)$ and $L(p,q')$ are homeomorphic then

$$qq' \equiv \pm r^2 \pmod p, \quad \text{for some } r,$$

the sign being $+$ or $-$ according as the homeomorphism preserves or reverses orientation.

Alexander's argument was eventually formalized into the definition of the linking form $T_1(M) \times T_1(M) \to \mathbb{Q}/\mathbb{Z}$ of a 3-manifold M, where $T_1(M)$ is the torsion subgroup of $H_1(M)$. This was done in [7, 8, 85, 88] and [96]. In particular, Seifert's paper [96] gives a set of local invariants which, in the odd order case, completely classify such forms.

The condition $qq' \equiv \pm r^2 \pmod p$ did not seem to be a sufficient condition for homeomorphism, however; for example $L(7,1)$ and $L(7,2)$ appeared to be topologically distinct. The combinatorial classification of lens spaces, i.e. their classification up to *PL* homeomorphism, was finally achieved by Reidemeister in 1935 [86], using his *torsion* invariant. (This invariant was formalized and generalized to higher dimensions by Reidemeister's student Franz [35].) The result is that $L(p,q)$ and $L(p,q')$ are *PL* homeomorphic if and only if either $q \equiv \pm q' \pmod p$, or $qq' \equiv \pm 1 \pmod p$, where as usual the $\pm$ sign corresponds to the orientation character of the homeomorphism. (The sufficiency of the condition is straightforward.)

This became a classification up to homeomorphism with the proof of the Hauptvermutung by Moise in 1952 [65]. Meanwhile, Fox had outlined an approach to the topological classification, which involved considering the Alexander polynomials of knots in lens spaces, which would not require the Hauptvermutung; see [30, Problem 2]. This was implemented later by Brody [15]. (The fact which replaces the Hauptvermutung in this proof is the topological invariance of simplicial homology.)

This is a convenient place to say that although Moise's result that every 3-manifold can be triangulated, in an essentially unique way, is clearly of fundamental importance, we will not discuss it further. We remark that a simpler proof was given later by Bing [13].

The lens spaces were also natural subjects for investigations of a more algebraic topological nature. In this vein, Rueff showed [91] that there exists a degree 1 map $L(p,q) \to L(p,q')$ if and only if $qq' \equiv r^2 \pmod p$, for some r. The homotopy classification of lens spaces was obtained by Whitehead [124]: $L(p,q)$ and $L(p,q')$ are homotopy equivalent if and only if $qq' \equiv \pm r^2 \pmod p$, for some r. In particular, they are orientation-preservingly homotopy equivalent if and only if their linking forms are isomorphic. Franz [36] showed that the homotopy class of a map $L(p,q) \to L(p,q')$ is determined by the homomorphism it induces on the fundamental group, together with its degree. Together with Rueff's result, this also gives the classification up to homotopy type.

We have seen that Tietze suspected that the lens spaces provide examples of distinct 3-manifolds with isomorphic fundamental groups. He also drew attention to another apparent source of this phenomenon, at least for manifolds with boundary. In [110, pp. 96, 97] he considers the exteriors M and M' of two split links L and L' in S^3, where L consists of two copies of the right-handed trefoil K, and L' consists of a copy of K and a copy of the left-handed trefoil $-K$, the reflection of K. Thus, if X denotes the exterior of K, then M is homeomorphic to the connected sum $X \# X$, while M' is homeomorphic to $X \# -X$. Tietze notes that $\pi_1(M)$ and $\pi_1(M')$ are both isomorphic to the free product $\pi_1(X) * \pi_1(X)$. On the other hand, it appears that there is no orientation-preserving homeomorphism of S^3 taking K to $-K$, and hence no homeomorphism of S^3 taking L to L', and "hence" no homeomorphism from M to M'. The first assertion was later proved by Dehn [23], and an additional argument (which would have been available to Dehn, for example) can be given to conclude that indeed M and M' are not homeomorphic. It is interesting that these two phenomena pointed out by Tietze, namely, lens spaces, and connected summands with no orientation-reversing homeomorphism, conjecturally account completely for the failure of a closed, orientable 3-manifold to be determined by its fundamental group.

So the lens spaces provide simple examples of complex behavior in 3-manifolds: the properties of having isomorphic fundamental group, having the same homotopy type, and being homeomorphic, are all distinct. On the other hand, they are somewhat misleading; for example, it may have been that their failure to be determined by their fundamental group suggested that this was likely to be common among 3-manifolds, whereas in fact they appear to be the only irreducible examples. The lesson here seems to be: don't worry about simple counterexamples; they may be counterexamples only because they're simple.

8. Kneser's decomposition theorem

The important idea of cutting a 3-manifold along 2-spheres was introduced in the beautiful 1929 paper of Kneser [56]. Apart from his short note [55], this seems to be Kneser's only paper on 3-manifolds, but it turned out to be extremely influential.

In Section 4 of this paper, Kneser considers the operation of cutting a closed 3-manifold M along an embedded 2-sphere S, and capping off each of the resulting boundary components with a 3-ball, giving a (possibly disconnected) 3-manifold M_1. This process he calls a *reduction*. If S bounds a 3-ball in M, then M_1 is just another copy of M together with a copy of S^3, and the reduction is *trivial*. A manifold is *irreducible* if it admits only trivial reductions, i.e. if every 2-sphere in the manifold bounds a 3-ball. Kneser remarks that in order to justify the term "reduction", M_1 should be in some sense simpler than M, but that there is no reasonable topological invariant which can be used to show this. Nevertheless, he is able to prove the following finiteness theorem:

> *Associated to each 3-manifold M is an integer k with the following property: if $k+1$ successive reductions are performed on M, then at least one of them is trivial. By means of k (or fewer) nontrivial reductions M can be transformed to an irreducible manifold.*

Before describing Kneser's proof, we discuss a result of Alexander [4], which is fundamental in this context, and which is needed in the proof. Alexander's theorem asserts that every 2-sphere in S^3 separates it into two regions, the closure of each of which is a 3-ball. In particular, S^3 is irreducible. The corresponding statement one dimension lower, that every circle in S^2 separates it into two components whose closures are disks, is the classical Schönflies theorem, and it is true with no additional hypotheses. Apparently Alexander at one time announced (but did not publish) the same result for 2-spheres in S^3 (see [5, p. 10]), but later constructed counterexamples; the first [6] was based on Antoine's necklace, and the second [5] was Alexander's famous horned sphere. Meanwhile, he gave a proof of the 3-dimensional Schönflies theorem for polyhedral 2-spheres [4]. He does this by considering the intersection of such a 2-sphere S (in $\mathbb{R}^3$) with a generic family of parallel planes. With finitely many exceptions, these will meet S transversely, each of the exceptional planes containing exactly one local minimum, local maximum, or (multiple) saddle point of S. By considering the disk bounded by an innermost simple closed curve in one of the planes containing a saddle point, Alexander replaces S by two 2-spheres, each of which is simpler than S. The theorem now follows easily by induction. (The induction starts with a sphere having only a single local minimum and a single local maximum.)

By a similar argument, Alexander also proves that any (polyhedral) torus in S^3 bounds a solid torus, a fact which was conjectured by Tietze [110]. Later, Fox [33] showed that Alexander's argument generalizes to show that any closed surface in S^3 is compressible. He used this to prove that any compact, connected 3-manifold with boundary embedded in S^3 is *homeomorphic* to the closure of the complement in S^3 of a disjoint union of handlebodies.

We now turn to Kneser's proof of his finiteness theorem.

Fix a triangulation of M, and let Σ be a disjoint union of k 2-spheres in M such that no component of $M - \Sigma$ is a punctured 3-sphere. Kneser shows that Σ may be modified so that each component of the intersection of Σ with any 2-simplex in the triangulation is an arc with its endpoints on distinct edges of the 2-simplex, and each component of the intersection of Σ with any 3-simplex is a disk.

After this, in any 2-simplex, all but at most four of the complementary regions of the intersection of Σ with that 2-simplex have a natural product structure as quadrilaterals, the possible exceptions being a triangle containing a single vertex, and a middle region meeting all three sides of the 2-simplex. Also, for each 3-simplex, if a complementary region X of the intersection of Σ with the 3-simplex meets each face of the 3-simplex in product regions, then this product structure extends over X, so that X is a *prism*.

Now the number of components of $M - \Sigma$ is at least $k - r$, where r is the first mod 2 Betti number of M. It follows from the above discussion that such a component meets every 3-simplex in prisms, unless it contains a vertex or a middle region of a 2-simplex. Therefore, letting α_i be the number of i-simplexes in the triangulation, if $k > r + \alpha_0 + \alpha_2$, then one of the components of M cut along Σ has the structure of an I-bundle over a surface, and hence is either $S^2 \times I$ or a twisted I-bundle over $\mathbb{RP}^2$ (in other words, a punctured $\mathbb{RP}^3$). In the latter case, we can collapse the corresponding 2-sphere onto the $\mathbb{RP}^2$ and repeat the argument. This shows that if $k > r + \alpha_0 + \alpha_2$ then some component of M cut along Σ is homeomorphic to $S^2 \times I$, contrary to assumption. Hence, we can take k to be $r + \alpha_0 + \alpha_2$ in the theorem.

Kneser continues:

> If you study in more detail the different possible ways of transforming by reductions a given 3-manifold into irreducible 3-manifolds, the result is the following theorem, which reduces the topological properties of all 3-manifolds to those of the irreducible ones.

Kneser then gives the following careful statement of his decomposition theorem:

> *Every 3-manifold can be expressed in the following way: take k orientable asymmetric 3-manifolds, ℓ orientable symmetric 3-manifolds, and m non-orientable 3-manifolds* ($k, \ell, m \geqslant 0$)*, all irreducible, and remove a 3-ball from each; from S^3 remove $k + \ell + m + 2r + 2s$ 3-balls (where $r \geqslant 0$; $s = 0$ or 1, and $s = 0$ if $m > 0$); identify the boundary 2-spheres of the punctured manifolds with $k + \ell + m$ boundary 2-spheres of the punctured S^3; identify the remaining boundary 2-spheres in pairs, r pairs being identified in a way that is coherent with the orientation of the punctured S^3, and the last pair, if $s = 1$, so as to give a non-orientable manifold. Two 3-manifolds generated in this way are homeomorphic if and only if the numbers k, ℓ, m, r, s are the same in both cases, the 3-manifolds that are used are homeomorphic in pairs, and in the case of an orientable 3-manifold ($m = s = 0$), the orientations of the asymmetric 3-manifolds are connected in the same way in both cases.*

Kneser omits the details of the proof, but these were later elegantly supplied by Milnor [63]; one guesses that this was very much along the lines that Kneser had in mind. For the non-orientable case, see [111].

Several interesting remarks of Kneser are relegated to footnotes to his decomposition theorem. First, he defines an orientable 3-manifold to be *symmetric* if it has an orientation-reversing self-homeomorphism, and says that the simplest example of an asymmetric 3-manifold is the "torus manifold" $L(3, 1)$, or, more generally, $L(k, \ell)$, provided -1 is a quadratic nonresidue mod k. Second, he gives as examples of irreducible 3-manifolds, the 3-torus T^3 (presumably because its universal cover is $\mathbb{R}^3$), and any 3-manifold covered by S^3. It is here that he mentions in passing that an example of a manifold of this second type is the homology sphere with nontrivial finite fundamental group constructed by Dehn from the trefoil knot. Another footnote makes a reasoned plea for the use of the term "path group" instead of "fundamental group", a plea that seems to have gone unheeded.

Going back to Kneser's proof of his finiteness theorem, this beautiful argument had far-reaching consequences in the work of Haken about thirty years later. Haken observed that Kneser's argument can be applied to a system of disjoint, incompressible (closed) surfaces in a (compact) irreducible 3-manifold M, to show that there is an integer $k(M)$ with the property that the number of such surfaces, no two of which cobound a product in M, is at most $k(M)$. This finiteness theorem allows him to prove that every irreducible manifold which contains an incompressible surface (these are now called *Haken manifolds*), has a *hierarchy*, in other words, it can be reduced to a disjoint union of 3-balls by successively cutting it along incompressible surfaces. This was used to great effect by Waldhausen, to prove, for example, that two Haken manifolds with isomorphic fundamental groups are homeomorphic [115], that the universal cover of a Haken manifold is $\mathbb{R}^3$ [115], and that the fundamental group of a Haken manifold has solvable word problem [116].

Again based on Kneser's idea of controlling his surfaces (spheres) by making them have nice intersections with the simplices of a fixed triangulation of the manifold, Haken developed an algorithmic theory of such *normal surfaces* [42]. This ultimately led, with the work of Waldhausen, Johannson, Jaco-Shalen, and Hemion's solution of the conjugacy problem for automorphisms of surfaces, to the solution of the homeomorphism problem for Haken manifolds; see [117].

So the idea, in Heegaard's words, of "cutting a manifold until it is simply connected", is realized in Haken's concept of a hierarchy, and leads to a solution of the homeomorphism problem for a large class of 3-manifolds, although the notion of a "normal form" survives only as a nebulous logical construct.

Finally, we mention that, very recently, Rubinstein [90] (see also Thompson [106]) has solved the homeomorphism problem for S^3, also using normal surfaces, but in a very different way. So we see how important Kneser's few pages have been for the theory of 3-manifolds.

9. Geometric 3-manifolds

We have already seen how early approaches to the study of 3-manifolds naturally took the form of pursuing analogies with the theory of 2-manifolds. Another feature of 2-manifolds which was well known from the work of Klein and others was that every closed surface can be given a spherical, Euclidean, or hyperbolic structure. That is, it can be represented as the quotient of either the 2-sphere S^2, the Euclidean plane E^2, or the hyperbolic plane H^2, by a group of isometries acting freely and properly discontinuously. This group is of course isomorphic to the fundamental group of the 2-manifold.

The fact that the lens space $L(p, q)$ is the quotient of such an action on S^3 by a cyclic group of order p is essentially in Tietze [110], and was made explicit in Hopf [47], who also gave other examples of spherical 3-manifolds, with noncyclic fundamental groups. In addition, Hopf proved that any n-manifold with a complete Riemannian metric of constant curvature is a quotient of S^n, E^n, or H^n by a free properly discontinuous action of a group of isometries. By finding all the finite subgroups of $SO(4)$ that act freely on S^3, Seifert and Threlfall, in [107, 108], gave a complete description of all spherical 3-manifolds.

Their classification can be roughly described as follows. The quotient of $SO(4)$ (the group of orientation-preserving isometries of S^3) by its center $\{\pm id\}$ is isomorphic to $SO(3) \times SO(3)$, so a finite subgroup G of $SO(4)$ gives rise to two finite subgroups G_L and G_R of $SO(3)$. Now the finite subgroups of $SO(3)$ are the finite cyclic groups C_n, the dihedral groups D_{2n} of order $2n$, and the tetrahedral, octahedral, and icosahedral groups T, O, and I, of orders 12, 24, and 60. If G acts freely on S^3, then G_L (say) must be cyclic, and G can then be described as being of cyclic, dihedral, tetrahedral, octahedral, or icosahedral type, according to the type of G_R. The groups of cyclic type are cyclic, and the corresponding 3-manifolds are the lens spaces. The groups of dihedral, tetrahedral, octahedral, and iccosahedral type include the corresponding binary groups D^*_{2n}, T^*, O^*, and I^*, while the dihedral and tetrahedral types include additional families $D'_{m,n}$ and T'_m. The general group of a given type is the direct product of any one of these with a cyclic group of relatively prime order.

The 3-manifolds M with fundamental groups G of dihedral type are the *prism spaces*: if $G_L \cong C_m$ and $G_R \cong D_{2n}$, then M can be obtained by suitably identifying the faces of a $2mn$-sided prism. A special case of this is quaternionic space, $m = 1, n = 2$, obtained by identifying opposite faces of a cube, as described by Poincaré (see Section 3). The spherical dodecahedral space M_{dodeca} is of icosahedral type, with $G_L = 1$ and fundamental group G

the binary icosahedral group I^*. It is the only homology 3-sphere (apart from S^3) among the spherical 3-manifolds.

The lens spaces and prism spaces also appear, in a different context, in [93]. There, Seifert classifies the 3-manifolds that can be obtained from a solid torus by identifying its boundary with itself via some involution. These manifolds fall into three classes, according to the nature of the involution: the first are the lens spaces and $S^1 \times S^2$, the second are the prism spaces (including the lens spaces $L(4q, 2q-1)$, and $\mathbb{RP}^3 \# \mathbb{RP}^3$, as degenerate cases), and the third consist of $S^1 \times \mathbb{RP}^2$ and the twisted S^1-bundle over S^2.

The question remains whether every 3-manifold with finite fundamental group G is spherical. A more modest goal would be to show that at least G is isomorphic to $\pi_1(M)$ for some spherical 3-manifold M. These questions are still open, but progress was made in 1957 by Milnor, who proved [62] that any such G has at most one element of order 2. Combining this with the fact that the cohomology of G must have period 4, he deduced that any counterexample G to the second assertion must belong to one of two infinite families $Q(8n, k, \ell)$ and $O(48r)$. The second family, and half the first family, were subsequently ruled out by Lee [58].

An important offshoot of the work of Seifert and Threlfall came from their observation that any finite subgroup of $SO(4)$ which acts freely on S^3 commutes with an S^1 subgroup of $SO(4)$, and hence this S^1-action descends to the quotient manifold M, giving M a (singular) fibering by the orbits of the action. This motivated Seifert to investigate 3-manifolds which can be fibered by circles in this fashion, now called *Seifert fibered spaces*. (The special case of circle tangent bundles of surfaces had been studied earlier by Hotelling [50, 51], as the 3-manifolds of states of motion of dynamical systems.) In the introduction to his work [95] on fibered spaces (see also the translation by W. Heil in [97]) Seifert says:

> The question that underlies this paper is the homeomorphism problem for 3-dimensional closed manifolds. The fundamental theorem of surface topology tells us how many topologically distinct 2-manifolds there are. The methods used to prove this have not yet been generalized to three or more dimensions. There are two ways to approach the 3-dimensional problem. The first is to examine the regions of discontinuity[8] of 3-dimensional metric groups of motions. Whereas in two dimensions every closed surface appears as the region of discontinuity of a fixed point free group of motions, there are 3-manifolds for which this does not hold. The regions of discontinuity of 3-dimensional spherical actions are endowed with a certain fibration; the fibers are the orbits of a continuous group of motions of the sphere. ... This leads us to the second approach: instead of investigating a complete system of topological invariants of 3-dimensional manifolds, we search for a system of invariants for fiber-preserving maps of fibered 3-manifolds. This problem is completely solved in this paper. Of course these invariants refer to the fibering of the manifold, not to the manifold itself, so that the question remains open, whether two spaces with different fibrations are topologically distinct. Moreover, there are 3-manifolds that do not admit any fibration. Nevertheless, in many cases the fiber invariants can be used to decide whether 3-manifolds are homeomorphic.

Seifert's paper is a masterpiece of content and clarity. He builds up from scratch the complete and rich theory of his fibered spaces, and in fact his account left little to be added for several decades. With the torus decomposition theorem of Johannson, and Jaco and Shalen, the Seifert fibered spaces emerged as one of the basic building blocks of Haken manifolds,

[8] I.e. quotients.

and this role has been further clarified and emphasized by the work of Thurston on geometrization of 3-manifolds.

Seifert defines a 3-dimensional *fibered space* to be a closed 3-manifold M which is a disjoint union of circles (*fibers*), such that each fiber has a solid torus neighborhood, consisting of fibers, which are the core of the solid torus together with curves that wind around the core α ($\geqslant 1$) times latitudinally and ν times meridionally. If $\alpha > 1$ then the core is a *singular* fiber of M, of *multiplicity* α. The *base* of the Seifert fibration is the quotient surface obtained from M by identifying each fiber to a point.

Seifert shows that oriented fibered spaces, up to orientation- and fiber-preserving homeomorphism, are classified by $(F; b; (\alpha_1, \beta_1), \ldots, (\alpha_r, \beta_r))$, where F is the topological type of the base surface, (α_i, β_i) are the suitably normalized invariants of the singular fibers, $1 \leqslant i \leqslant r$, and b is the Euler number of a certain associated circle bundle over F, i.e. a fibered space with no singular fibers. He also gives a similar, but more involved, classification of the non-orientable fibered spaces.

In [95, 108], it is shown that the Seifert fibered spaces M with $\pi_1(M)$ finite are precisely the spherical 3-manifolds. These are S^3 (whose Seifert fibrations correspond to pairs of nonzero coprime integers m, n, the nonsingular fibers being (m, n)-torus knots); the lens spaces (which have Seifert fibrations with base S^2 and one or two singular fibers); and the Seifert fiber spaces with base S^2 and three singular fibers, whose multiplicities form one of the Platonic triples $(2, 2, n)$, $n \geqslant 2$, $(2, 3, 3)$, $(2, 3, 4)$, or $(2, 3, 5)$, corresponding to the groups of dihedral, tetrahedral, octahedral, and icosahedral type, respectively. Some of the prism spaces also have Seifert fibrations with base $\mathbb{RP}^2$ and one singular fiber.

Turning to Poincaré spaces, i.e. homology 3-spheres not homeomorphic to S^3, Seifert shows that for any sequence $\alpha_1, \ldots, \alpha_r$ of $r \geqslant 3$ pairwise coprime integers $\geqslant 2$, there exists a Seifert fibered Poincaré space, with base S^2 and r singular fibers of multiplicities $\alpha_1, \ldots, \alpha_r$. Conversely, every Seifert fibered Poincaré space is of this form, and two such are homeomorphic if and only if the corresponding sequences of multiplicities $\alpha_1, \ldots, \alpha_r$ are the same, up to order, in which case they are fiber-preservingly homeomorphic. The only Seifert fibered Poincaré space with finite fundamental group is the spherical dodecahedral space, with three singular fibers of multiplicities 2, 3, and 5.

Recalling Dehn's construction of Poincaré spaces by surgery on knots, Seifert shows that the Poincaré spaces obtained in this way from an (m, n)-torus knot are precisely those that have Seifert fibrations with three singular fibers of multiplicities m, n, and $|qmn - 1|$, for some $q \neq 0$. In particular, the manifolds M_{Dehn} and M_{dodeca} are homeomorphic.

Seifert also discusses branched coverings of Seifert fibered spaces, and shows that the Seifert fibered Poincaré spaces can be realized in several different ways as branched coverings of S^3, with branch set a collection of nonsingular fibers in some Seifert fibration of S^3. In particular, if $\alpha_1, \alpha_2, \alpha_3$ are pairwise coprime integers $\geqslant 2$, then the α_i-fold cyclic branched covering of the (α_j, α_k)-torus knot is the Seifert fibered Poincaré space with three singular fibers of multiplicities $\alpha_1, \alpha_2, \alpha_3$, where $\{i, j, k\} = \{1, 2, 3\}$.

This brings us to the proof in [118] that M_{Poin} is homeomorphic to M_{dodeca}. Seifert and Weber start with the genus 2 Heegaard diagram of $M = M_{\mathrm{Poin}}$ given by Poincaré in [81], expressing M as the union of two genus 2 handlebodies V and V', with meridians K_1, K_2 and K_1', K_2', respectively. They note that there is an orientation-reversing involution of the Heegaard surface, with fixed point set a simple closed curve C, which interchanges K_i and K_i', $i = 1, 2$. (Interestingly, this was also observed by Poincaré [81, p. 108], although he made no use of it.) Hence there is an involution on M, interchanging V and V', with fixed point set C. Seifert and Weber show that the quotient of the pair (M, C) by this involution is (S^3, K), where K is a (3,5)-torus knot. Thus M is the 2-fold covering of S^3 branched along K, and hence is the unique Seifert fibered homology sphere with singular fibers of multiplicities 2, 3, and 5.

Turning to Euclidean 3-manifolds, these were classified by Nowacki [70], and Hantzsche and Wendt [43], independently. There are precisely 10 of them, 6 orientable and 4 non-orientable. They are all covered by the 3-torus T^3. Nowacki also classified the open Euclidean 3-manifolds; here there are 4 orientable and 4 non-orientable examples. Nowacki's proof is based on the classification of the 3-dimensional crystallographic groups, while that of Hantzsche and Wendt is more direct.

What about the hyperbolic case, which in dimension 2 is the generic one? Hantzsche and Wendt conclude their paper by saying:

> The Euclidean space form problem is hereby completely settled, and the spherical case has been done in [107, 108]. Much harder is the question of hyperbolic space forms, of which one knows only a few examples.

The first example to be given, of a discrete subgroup of $PSL_2(\mathbb{C})$ (the group of orientation-preserving isometries of H^3) with a fundamental domain of finite volume, was the group $PSL_2(\mathbb{Z}[i])$, described by Picard [76]. More generally, Bianchi studied the groups $PSL_2(R)$, where R is the ring of algebraic integers in an imaginary quadratic number field [12]. However, no-one at that time seems to have found, or looked for, torsion-free subgroups of finite index of these groups, which would give rise to (cusped) hyperbolic 3-manifolds of finite volume. The first hyperbolic 3-manifold of finite volume was described in the 1912 thesis of Gieseking [38], a student of Dehn. (See also [60, Chapter V].) This manifold can be obtained from a regular ideal tetrahedron in H^3 by suitably identifying its faces in pairs. It is non-orientable, and turns out to be the unique noncompact hyperbolic 3-manifold of minimal volume [1].

The first examples of closed hyperbolic 3-manifolds were constructed by Löbell in 1931 [59]; (in the preface to his paper he thanks Koebe for "expressing, in conversation, the desire that the question of the existence of such examples should be decided"). Löbell starts by constructing a 3-dimensional hyperbolic polyhedron whose faces are two right angled hexagons and 12 right angled pentagons. By suitably assembling copies of this polyhedron he builds a compact hyperbolic 3-manifold, whose boundary is totally geodesic and consists of four isometric copies of a surface of genus 2. Taking a finite number of copies of this manifold, and identifying the boundary components in pairs, he then obtains infinitely many closed hyperbolic 3-manifolds, which can be chosen to be either orientable or non-orientable.

A more symmetrical example was described by Seifert and Weber in [118]. This manifold, the *hyperbolic dodecahedral space*, or *Seifert–Weber manifold*, comes from a tiling of H^3 by regular dodecahedra with dihedral angles $2\pi/5$; it is the quotient of H^3 by a fixed point free group of isometries having one of these dodecahedra as a fundamental domain. It can be obtained from a single copy of the dodecahedron by identifying opposite faces by a rotation through $3\pi/5$. They also show that it is a 5-fold cyclic branched covering of S^3, with branch set the Whitehead link.

Permitting ourselves to look ahead, some more examples of closed hyperbolic 3-manifolds were constructed by Best in 1971 [11], using other regular hyperbolic 3-dimensional polyhedra. In 1975 Riley [89] showed that the complement $S^3 - K$ of the figure eight knot K has a complete hyperbolic structure, by finding an explicit discrete, faithful representation of $\pi_1(S^3 - K)$ in the Bianchi group $PSL_2(\mathbb{Z}[e^{2\pi i/3}])$. Nevertheless, at that time it was still the case that only a few examples of hyperbolic 3-manifolds were known. The situation changed dramatically with the work of Thurston, however (see [109]), who showed that hyperbolic 3-manifolds are plentiful, and indeed presented much evidence for his *geometrization conjecture*, which would imply, for example, that any closed orientable 3-manifold which satisfies certain obvious necessary conditions for it to be hyperbolic, namely that it is irreducible, and its fundamental group has no free Abelian subgroup of rank 2, is in fact hyperbolic. So it appears that, just as in dimension 2, hyperbolic geometry is "generic" in dimension 3.

10. The state of play up to 1935

The year 1935 is a convenient place at which to pause and take stock of the state of 3-dimensional topology. By this time the foundations of what we would now call geometric topology had become sufficiently well established that a textbook could be written, Seifert and Threlfall's famous "Lehrbuch der Topologie", published in 1934. (An English translation appeared in 1980 [97].) One chapter of that book is specifically devoted to 3-manifolds.

Let us summarize what has been achieved. The fundamental group has emerged as an important invariant, although it is known that there are nonhomeomorphic 3-manifolds (lens spaces) with isomorphic groups. The homology of a 3-manifold is determined by its fundamental group, and is now seen as a very weak invariant; in particular there are infinitely many homology 3-spheres. There is a complete description of all 3-dimensional spherical manifolds, and of the handful of Euclidean ones. There are also some examples of hyperbolic 3-manifolds. Seifert has given a complete description of all his fibered spaces, and classified them up to fiber-preserving homeomorphism. Dehn has shown how to construct 3-manifolds by "surgery" on knots.

Three methods are known by which all 3-manifolds may be constructed: Heegaard diagrams, identification of faces of polyhedra, and branched coverings of the 3-sphere. However, none of these methods has led to anything approaching a classification. The situation is summarized well by Seifert and Threlfall [97, p. 228]:

> The construction of 3-dimensional manifolds has been reduced to a 2-dimensional problem by means of the Heegaard diagram. This problem is the enumeration of all Heegaard diagrams. Even if the diagrams could all be enumerated, the homeomorphism problem in 3 dimensions would not be solved because a criterion is still lacking for deciding when two different Heegaard diagrams generate the same manifold. The enumeration has been carried out successfully in the simplest case, that of Heegaard diagrams of genus 1, but the problem of coincidence of manifolds, that is, the homeomorphism problem for lens spaces, has not been solved even here.
>
> Another way to attempt the enumeration of all 3-dimensional manifolds would be to construct all polyhedra having pairwise association of faces. This also is a 2-dimensional problem and it has met with as little success at solution as the problem of enumerating the Heegaard diagrams.
>
> It is known from the theory of functions of complex variables that one can obtain any closed orientable surface as a branched covering surface of the 2-sphere, where the branching occurs at finitely many points. Corresponding to this result, it is possible

> to describe each closed orientable 3-dimensional manifold as a branched covering of the 3-sphere. In this case the branching occurs along closed curves (knots) which lie in the 3-sphere. Here also the enumeration and distinguishing of individual covering spaces leads to unanswered questions. On occasion the same manifold can be derived as branched coverings of the 3-sphere with quite distinct knots as branch sets; as an example, three different branch sets are known for the spherical dodecahedron space.

The only general result on the structure of 3-manifolds is Kneser's existence and uniqueness of prime decompositions.

In the introduction to their chapter on n-dimensional manifolds [97, p. 235], Seifert and Threlfall say:

> Because of their clear geometric significance, homogeneous complexes play a distinctive role among the complexes. We have given the name "manifolds" to the homogeneous complexes in 2 and 3 dimensions and we have attempted to gain a complete view of their properties. Our attempt was successful in 2 dimensions. In 3 dimensions we did not get further than a presentation of more or less systematically arranged examples. The complete classification of n-dimensional manifolds is a hopeless task at the present time.

11. Dehn's lemma and the loop theorem

Dehn's 1910 article, "On the topology of 3-dimensional space" [22], contains the following statement, which he refers to simply as *the lemma*, "because of its important place" in the paper.

DEHN'S LEMMA (1). *Let X be a 2-complex in the interior of an n-dimensional manifold M, $n > 2$. On X, let the curve C bound a singular disk D. If D has no singularities on its boundary, then C bounds an embedded disk in M.*

We will discuss the context of this rather curious statement later.

As Dehn says, the lemma is clearly true if $n > 3$; the interesting case is when $n = 3$. In this case, a little thought shows that it may be restated as follows:

DEHN'S LEMMA (2). *Let C be a simple loop on the boundary of a 3-manifold M, which bounds a singular disk in M. Then C bounds an embedded disk in M.*

Note that the hypothesis is equivalent to the statement that C is null-homotopic in M. Perhaps the most natural statement of this kind, which dispenses with the assumption that C is simple, and asserts that if a 3-manifold contains a nontrivial singular disk (homotopical information) then it contains a nontrivial embedded disk (topological information), is the following, which might be called the

DISK THEOREM. *Let M be a 3-manifold and let F be a boundary component of M such that $\pi_1(F) \to \pi_1(M)$ is not injective. Then M contains an embedded disk D, with ∂D contained in F, such that $[\partial D] \neq 1 \in \pi_1(F)$.*

By taking F to be an open annular neighborhood of C, we see that the disk theorem implies Dehn's lemma. On the other, it is implied by Dehn's lemma together with the

LOOP THEOREM. *Let M be a 3-manifold and let F be a boundary component of M such that $\pi_1(F) \to \pi_1(M)$ is not injective. Then there is an essential simple loop in F which is null-homotopic in M.*

We will return to these statements later. But first we note that arguments with singular disks in connection with the fundamental group of a 3-manifold appear in Poincaré's work. For example, in [81], he wishes to show that if $C_1, \ldots, C_p$ are (homologically independent) disjoint simple loops on the boundary of a handlebody V of genus p, which are null-homotopic in V, then they bound disjoint embedded disks in V. The "innermost disk" cutting and pasting argument that he uses to make the disks disjoint is valid if they are already embedded; however, on the latter point he merely says:

> ... an analogous argument will show that since the curves are embedded one may always suppose that the disks [that they bound] are surfaces without double curves.

This is precisely Dehn's lemma.

This fact about embedded disks in a handlebody is not needed for the discussion of Poincaré's homology sphere, but, as we have mentioned in Section 5, in order to compute the fundamental group, the following fact is: if $C_1, \ldots, C_p$ is a system of meridians on the boundary of a handlebody V of genus p, then any element in the kernel of $\pi_1(\partial V) \to \pi_1(V)$ is a product of conjugates of $[C_1]^{\pm 1}, \ldots, [C_p]^{\pm 1}$. Poincaré proves this by taking a disk D in V bounded by a loop C on ∂V, and considering the arcs of intersection of D with disks $A_1, \ldots, A_p$ bounded by $C_1, \ldots, C_p$. The disks A_i are certainly disjoint and embedded, but Poincaré appears to assume that C and D are also nonsingular; however, if one interprets his argument as applying to the inverse image of the union of the A_is under a map of a disk into V, then it is in fact correct. A similar remark applies to his proof that given a Heegaard splitting (V', V'') of a 3-manifold, a loop on the Heegaard surface W is null-homotopic in the manifold if and only if it is a product of elements in the two kernels $\ker(\pi_1(W) \to \pi_1(V'))$ and $\ker(\pi_1(W) \to \pi_1(V''))$.

Dehn attempted to prove his lemma by using the cutting and pasting procedure (he calls the operation a *switch*, or *Umschaltung*), that is hinted at in Poincaré, to remove the singularities of the given singular disk, but although he did give a detailed argument, it was, as is well known, faulty. The error was pointed out by Kneser in a footnote added in proof to his paper [56], and privately in a letter to Dehn dated 22 April, 1929 (see [24, p. 87]). Kneser himself was trying to prove the following:

KNESER'S HILFSATZ. *Let F be a closed surface in a 3-manifold M such that $\pi_1(F) \to \pi_1(M)$ is not injective. Then there exists a disk D in M such that $D \cap F = \partial D$ is an essential loop in F, and the only singularities of D are on ∂D.*

Note that by allowing singularities on ∂D Kneser is taking into account the possibility that F is 1-sided in M.

Kneser starts with a singular disk D in M whose boundary is an essential loop in F, and first shows, by considering the intersection of D with F, that one may assume that $D \cap F = \partial D$. (Modulo this simplification, the Hilfsatz is equivalent to the disk theorem, by cutting M along F.) There follows a cutting and pasting argument to eliminate the double *arcs* of D, and finally an appeal to Dehn's lemma to get a disk of the desired type.

The model for the cutting and pasting approach to Dehn's lemma and Kneser's Hilfsatz is the case where the singular disk has only double points (i.e. no triple points), in which case a proof by these methods may readily be obtained – indeed as we have seen, this is essentially contained in Poincaré's work. More generally, a switch can be used to eliminate any *simple* double curve, i.e. one that comes from the identification of two disjoint simple closed curves in the nonsingular preimage of the disk. But Dehn's attempt to carry this through when the double curves themselves have singularities is unsuccessful; it is not clear in general that the cutting and pasting can be done consistently in the presence of these triple points. The sort of difficulty that arises is illustrated by an example given by Johannson in [52], which he states that Dehn was (by that time) also aware of. Kneser's argument for removing double arcs (which he gives in the case where there are no triple points, the general case being said to follow analogously), is also subject to the same criticism.

Johannson [52] gave necessary conditions for the realizability of *Dehn diagrams* (i.e. patterns of immersed circles) as the double curves of a singular disk in a 3-manifold, and in particular showed that the example mentioned above could not in fact be realized. He states that in fact in all known examples the cutting and pasting argument can be successfully carried out, and hence

> one can still hope that it might be possible to prove the lemma by suitably selected switches.

In [53] Johannson showed that if Dehn's lemma is true for orientable 3-manifolds then it is also true for non-orientable ones.

There the situation remained until the ground-breaking work of Papakyriakopoulos in 1957.

Before discussing this, let us briefly return to the papers of Dehn and Kneser, to see the uses to which they put their "lemmas".

Starting with Kneser, as an application of his Hilfsatz, he proves that every closed surface F in S^3 can be obtained from a 2-sphere by adding handles. In particular, this recovers (or would recover, if the proof of the Hilfsatz were correct) Alexander's theorem that every torus in S^3 bounds a solid torus [4]. The argument is straightforward. If F is not a 2-sphere, then the map $\pi_1(F) \to \pi_1(S^3) = 1$ is not injective, and hence, by the Hilfsatz, there is an embedded disk D in S^3 such that $D \cap F = \partial D$ is essential in F. Compressing F along D gives a simpler surface F', from which F is obtained by adding a handle. The result now follows by induction. Here we first find explicitly the important idea of compressing a surface F in a 3-manifold M, using the noninjectivity of $\pi_1(F) \to \pi_1(M)$, which Kneser used in his "proof" of his "conjecture" (see Section 12), and which played a central role in the later work of Stallings and Waldhausen.

Turning to Dehn's paper, the last chapter consists of two sections. In the first he shows that every closed 3-manifold M can be obtained by sewing a 3-ball along its boundary onto a *seam surface* N_2 in M. (This 2-complex N_2 is exactly Heegaard's "nucleus".) Letting N_1 be the 1-skeleton of N_2, Dehn notes that a neighborhood of N_1 in M is a (possibly non-orientable) solid handlebody, whose complement is also a solid handlebody. He concludes that every closed 3-manifold has a Heegaard splitting, although, oddly, he does not mention Heegaard here at all. He also notes that this implies that every closed 3-manifold is the union of four 3-balls.

The second section is described by Dehn in the introduction to his paper (see [24]):

> Section 2 deals with the important problem of the topological characterization of ordinary space, without, however, resolving the problem. It treats the question of how

> ordinary space may be topologically defined through the properties of its closed curves, and how to make it possible to decide whether or not a given space is homeomorphic to ordinary space. The history of this problem began when first Heegaard (Diss. Copenhagen 1898) and then Poincaré (Pal. Rend. v.13 and Lond. M.S. v.32) pointed out that in order to characterize ordinary space it does not suffice to assume that each curve bounds, possibly when multiply traversed. Indeed the manifolds with *torsion* show this. Then Poincaré proved in Pal. Rend. 1904, by construction of a "Poincaré space" that it is even insufficient for each curve to bound when traversed once.
>
> It now is natural to investigate whether it suffices to suppose that each curve in the space bounds a disk. This is also suggested at the end of Poincaré's work. However, the reduction of the problem given in the present work does not appear to lead directly to a solution. A deeper investigation of the fundamental groups of two-sided closed surfaces seems to be unavoidable.

Dehn's "reduction" of the Poincaré conjecture is the following. He notes that one may assume that N_1 consists of a wedge of circles. Since the manifold M is simply connected by hypothesis, each of these circles bounds a singular disk. *If* it were possible to choose these disks to have no singularities on their boundaries, then Dehn's lemma would give a system of embedded disks, with the same boundaries, meeting only at a single point. A neighborhood of the union of these disks would be a 3-ball, whose boundary S is contained in the solid handlebody A that is the complement of a neighborhood of N_1. Dehn now asserts that the 2-sphere S bounds a 3-ball in A, implying that M is the union of two 3-balls along their boundaries, and therefore homeomorphic to S^3. (Of course, the fact that a 2-sphere in a handlebody bounds a 3-ball requires proof and was not available at the time; if the handlebody is orientable it follows from Alexander's theorem that a (tame) 2-sphere in S^3 separates it into two 3-balls [4], and can be proved for non-orientable handlebodies by passing to the orientable 2-fold cover.)

Although Dehn's lemma finds a very important and natural application in the theorem that a knot with group $\mathbb{Z}$ is trivial, both the title of Dehn's paper and the mention of a 2-complex in the statement of the lemma (presumably Dehn had in mind the seam surface N_2) suggest that Dehn's real motivation was the Poincaré conjecture. Incidentally, Dehn's approach formed the basis for later attacks on this problem, notably by Haken.

As we have noted above, Dehn's lemma is something of a hybrid, and it alone does not enable one to translate purely homotopy theoretic information into topological information. For this reason, Papakyriakopoulos formulated the loop theorem, which he proved in [71]. He explains [74] that his motivation was the following characterization of handlebodies.

CONJECTURE H. *If M is a compact 3-manifold with boundary an orientable surface of genus g, and $\pi_1(M, \partial M) = 1$, then M is a handlebody of genus g.*

This is turn was apparently motivated by the Poincaré conjecture:

> Some years ago I was working on the Poincaré conjecture, and I tried to prove it by proving [Conjecture H]. But I failed, and I may say that I am now convinced that this is not the way to attack the Poincaré conjecture. However, the loop theorem, Dehn's lemma, Poincaré conjecture, and some results from algebraic topology imply [Conjecture H], see [72, Theorem (19.1), p. 297]. This was the reason I worked on the loop theorem, whose proof led me to the proof of Dehn's lemma and the sphere theorem.

The key idea that enabled Papakyriakopoulos to prove the loop theorem, Dehn's lemma, and the sphere theorem, was the use of covering spaces. Another, more elementary, principle that is used in all three proofs is the relation between the first homology of a 3-manifold and that of its boundary. We have already seen a special case of this for submanifolds of $\mathbb{R}^3$ in Poincaré's work; (see Section 3). Of more direct relevance here is the fact, proved by Kneser in [55], that if M is a 3-manifold such that $H_1(M; \mathbb{Z}_2) = 0$, then any two 1-cycles in ∂M have even intersection number. It follows that each component of ∂M is *planar*, i.e. embeds in the 2-sphere. Later, Seifert proved [94] that if M is compact and orientable, with boundary components of genera $p_1, \ldots, p_r$, then $\beta_1(M) \geqslant p_1 + \cdots + p_r$. In particular, if ∂M does not consist of 2-spheres, then $H_1(M)$ is infinite.

Here is a summary of Papakyriakopoulos' proof of the loop theorem. Let C be a loop in a boundary component F of a 3-manifold M, which is essential in F but null-homotopic in M. Let $p: \widetilde{M} \to M$ be the universal cover. Since C is null-homotopic in M, it lifts to a loop $\widetilde{C}$, say, in $p^{-1}(F)$. Now the crucial observation is that the singularities of C are the images under p of, firstly, the singularities of $\widetilde{C}$, and secondly, the intersections of $\widetilde{C}$ with its translates $\tau(\widetilde{C})$ under nontrivial elements τ of the group of covering transformations of $\widetilde{M}$. Hence, one wants to replace $\widetilde{C}$ by an essential loop $\widetilde{C}^*$ in $p^{-1}(F)$ such that (i) $\widetilde{C}^*$ is simple, and (ii) $\widetilde{C}^* \cap \tau(\widetilde{C}^*) = \emptyset$ for all τ. Note that $\widetilde{C}^*$ is automatically null-homotopic in $\widetilde{M}$ since $\pi_1(\widetilde{M}) = 1$. Then $C^* = p(\widetilde{C}^*)$ will be a simple essential loop in F which is null-homotopic in M. Condition (i) is easy to satisfy, and Papakyriakopoulos shows, by a delicate combinatorial argument, that (ii) can also be achieved. The important fact here is that $p^{-1}(F)$ is planar, by Kneser's result.

Whitehead had earlier proved a special case of the loop theorem, by a direct cutting and pasting argument: if C is a simple loop in the boundary of a 3-manifold M such that C^n is null-homotopic in M for some $n > 0$, then C is null-homotopic in M [122].

As Papakyriakopoulos says in [74]:

> Having observed ... that the loop theorem and Dehn's lemma are problems of the same kind, and having proved the loop theorem, the question arises naturally: *can we use the same method, or at least a modification of it, to prove Dehn's lemma*? The answer is affirmative ....

To prove Dehn's lemma, Papakyriakopoulos came up with his famous *tower construction*. In this, a tower of coverings is constructed, by starting with a neighborhood V_0 of the given singular disk $D_0 \subset M = M_0$, taking the universal covering M_1 of V_0, and lifting the map of the disk into V_0 to get a singular disk $D_1 \subset M_1$; now taking a neighborhood V_1 of D_1, taking the universal covering M_2 of V_1, and so on. Since D_i has fewer singularities than D_{i-1}, the tower must terminate, at $D_n \subset V_n \subset M_n$, say. In particular, since $\pi_1(V_n) = 1$, ∂V_n consists of 2-spheres, by Kneser's result.

Recall that we may assume that D_0 has no simple double curve, for otherwise D_0 could be simplified by a Dehn switch. Papakyriakopoulos distinguishes two cases at the top of the tower: (1) D_n is singular, and (2) D_n is nonsingular.

In case (1), since ∂V_n consists of 2-spheres, ∂D_n bounds a disk in ∂V_n. Papakyriakopoulos shows that this disk, when projected down the tower, gives a disk D_0^* in M, with $\partial D_0^* = C$, which (using the fact that D_0 has no simple double curves) has fewer triple points than D_0.

In case (2), first Papakyriakopoulos notes that ∂V_{n-1} does not consist of 2-spheres. (If $n = 1$, this is because otherwise the triple points of D_0 could be decreased, as in case (1) above, and if $n > 1$, it is a consequence of the facts that $V_{n-1} \subset M_{n-1}$, $\pi_1(M_{n-1}) = 1$, and $\pi_1(V_{n-1}) \neq 1$.) Hence $H_1(V_{n-1})$ is infinite, by the result of Seifert mentioned above, and it follows that there is a covering transformation τ of M_n of infinite order such that $D_n \cap \tau(D_n) \neq \emptyset$. This in turn gives rise to a simple double curve in D_0, contrary to hypothesis.

So in the proofs of both the loop theorem and Dehn's lemma, covering spaces are used to select the switches that are to be performed, on C and D_0, respectively. As Papakyriakopoulos says of the proof of Dehn's lemma:

> Actually, looking closer at the proof of Dehn's lemma in [72], we observe that we actually construct the desired disc [72, ll. 34–38, p. 2], and that *the construction is carried out by means of successive cuts.*[9]

Arnold Shapiro suggested the use of 2-fold coverings instead of universal coverings in the tower construction ([74, p. 323]), and Shapiro and Whitehead gave a simplified proof of Dehn's lemma using such coverings [98]. (The advantage of using 2-fold coverings is that, under such a covering $\widetilde{M} \to M$, the image F in M of a nonsingular surface $\widetilde{F}$ in $\widetilde{M}$ will have only double points, and so switches may be performed on F without difficulty.) Finally, again using 2-fold coverings, Stallings gave a proof of the disk theorem, [103] which, as we have noted, combines the loop theorem and Dehn's lemma, and this is the statement that is normally used in practice.

In retrospect, with Stallings' proof of the disk theorem, we can see that the assumption in Dehn's lemma that the boundary of the disk is embedded is in a sense a red herring. On the other hand, it seems to have played an important metamathematical role in this whole development. For it led Papakyriakopoulos to consider the two statements, the loop theorem and Dehn's lemma, separately, and it was in trying to prove the former that the key idea of using covering spaces suggested itself to him, this in turn leading him to take a similar approach to the latter.

The version of the loop theorem for duality spaces proved by Casson and Gordon [17, Theorem 4.5], shows that there are also mathematical grounds for separating the loop theorem from Dehn's lemma. In that version, F is still a surface, but the 3-manifold M is replaced by any complex which satisfies 3-dimensional Poincaré–Lefschetz duality over some field of untwisted coefficients, emphasizing the essentially $2\frac{1}{2}$-dimensional character of the loop theorem.

Recently, an interesting and entirely new proof of the disk theorem has been given by Johannson (Klaus, not Ingebrigt), using hierarchies [54].

Let us now make some remarks on the triviality problem for knots: (how) can you decide whether or not a given knot is trivial? This question clearly lies behind Dehn's result that a knot K is trivial if and only if its group $\pi_1(S^3 - K)$ is isomorphic to $\mathbb{Z}$. Ignoring the fact that the proof uses Dehn's lemma, this statement "reduces" the triviality question for a knot to an algebraic question, namely, is its group Abelian? Although Dehn did, in fact, refer to this as a "solution" to the knot triviality problem, he was very much aware that it is not at all clear that the algebraic question is any easier, or if it can be solved at all. In fact, it was precisely this kind of question, involving fundamental groups of 2- and 3-dimensional manifolds, that led him to articulate and bring to the fore the word problem and isomorphism problem for finitely presented groups. (For a detailed account of the

[9] I.e. Dehn switches.

topological origins of combinatorial group theory, and in particular the influence of the work of Tietze and Dehn, see [18].)

With the proofs, in the 1950's, that there are finitely presented groups with unsolvable word problem, and that the isomorphism problem, or even the triviality problem, for finitely presented groups, is unsolvable, the equivalence of the knot triviality problem to a question about (apparently fairly complicated) finitely presented groups made it seem more likely to be unsolvable. The same held for other, more direct equivalences. For example, in his very nice popular article [112], Turing shows, by considering elementary moves on knots that lie on the unit lattice in $\mathbb{R}^3$, that the knot problem is equivalent to a problem about substitutions on strings of letters which does not seem to have much structure. Later, after listing some decision problems that have been shown to be unsolvable, he says:

> It has recently been announced from Russia that the 'word problem in groups' is not solvable. This is a decision problem not unlike the 'word problem in semi-groups', but very much more important, having applications in topology: attempts were being made to solve this decision problem before any such problems had been proved unsolvable. ... Another problem which mathematicians are very anxious to settle is known as 'the decision problem of the equivalence of manifolds'.... It is probably unsolvable, but has never been proved to be so.[10] *A similar decision problem which might well be unsolvable is the one concerning knots which has already been mentioned.*[11]

In this climate, it was therefore probably something of a shock when, at the International Congress of Mathematicians in Amsterdam in 1954, Haken gave a short address in which he announced that the triviality problem for knots was solvable, using his theory of normal surfaces [41]. The details of the proof appeared in 1961 [42]. Later, in the mid 1970's, the knot problem was also shown to be solvable; see [117].

We have seen that the Poincaré conjecture seems to have been the motivation for both Dehn's formulation of his lemma and Papakyriakopoulos' proof of it. It was also the problem that first led Whitehead into 3-dimensional topology, with his false proof of the conjecture in [119]. In fact this proof also implied that any contractible open 3-manifold is PL-homeomorphic to $\mathbb{R}^3$. But Whitehead soon realized his mistake, and came up with his famous counterexample [121] (an informal description is given in [120]). This *Whitehead manifold* W is defined to be $S^3 - \bigcap_{n=0}^{\infty} T_n$, where $T_0 \supset T_1 \supset \cdots$ is a certain nested sequence of solid tori, derived from the Whitehead link. Thus T_n is unknotted in S^3, and T_{n+1} is null-homotopic in T_n but does not lie in a 3-cell in T_n. It follows that $\pi_1(W) = 1$, $H_2(W) = 0$, and W is irreducible. However, Whitehead shows in [121] that W is not PL-homeomorphic to $\mathbb{R}^3$. In [68], the geometric arguments of [121] (recall that Dehn's lemma was not available) are replaced by algebraic arguments involving the fundamental group, and there it is proved that W is not even homeomorphic to $\mathbb{R}^3$.

But these influences of the Poincaré conjecture are somewhat indirect, and in many ways it has tended to become an isolated problem, with progress in 3-dimensional topology going on independently of it, although recently the work of Thurston [109], and in particular his geometrization conjecture, has put it in a broader context.

---

[10] The unsolvability of the homeomorphism problem for manifolds of dimension $\geqslant 4$ was established by Markov in 1958 [61].

[11] My italics, C. McA. G.

12. π_2 and the sphere theorem

As homotopy theory arose and developed in the 1930's, with the work of Hopf and Hurewicz, investigations were begun on the homotopy properties of 3-manifolds. In 1936 Eilenberg [28] proved that if X is a nonseparating continuum in S^3, for example a knot, such that $\pi_1(S^3 - X) \cong \mathbb{Z}$, then $S^3 - X$ is aspherical (i.e. $\pi_i(S^3 - X) = 0, i \geqslant 2$).

(As an aside, it is interesting to see the terminology adjusting to the unavailability of Dehn's lemma. Eilenberg says:

> ... $\pi_1(S^3 - K) \cong \mathbb{Z}$, which means, in the sense of knot theory (based on the notion of the fundamental group), that K is unknotted.

Later, Whitehead [122] uses the term "ordinary circuit" to mean a knot that does not provide a counterexample to Dehn's lemma, i.e. one that is either unknotted or has the property that its latitude is not null-homotopic in its complement.)

Eilenberg also gave a necessary and sufficient condition for a 2-component link $K_1 \cup K_2$ to have the property that K_1 is a deformation retract of $S^3 - K_2$ (namely, $\pi_1(S^3 - K_2) \cong \mathbb{Z}$ and the linking number of K_1 and K_2 is 1). This led him to ask the following two questions, which turned out to be quite influential.

(1) For which knots K in S^3 is $S^3 - K$ aspherical?

(2) For which 2-component links L in S^3 is $S^3 - L$ aspherical?

Regarding his second question, Eilenberg notes that if L is a split link then $\pi_2(S^3 - L) \neq 0$. He also shows that if $\pi_1(S^3 - L) \cong \mathbb{Z} * \mathbb{Z}$ then $S^3 - L$ is not aspherical, for otherwise (by an earlier theorem of his) $S^3 - L$ would be deformable to a one-dimensional subcomplex, and hence would have 2nd Betti number equal to 0, contradicting Alexander duality. Finally, he shows (by considering the homology of the universal covering space) that a proper, connected, open subset U of S^3 is aspherical if and only if $\pi_2(U) = 0$. (This argument of course applies to any open 3-manifold.)

Inspired by Eilenberg's paper, Whitehead attacked the question of the asphericity of knot and link complements in his 1939 paper [123], (in which he thanks Eilenberg for many valuable suggestions). His approach is essentially algebraic. Starting with a knot K in S^3, he considers the cell decomposition of $S^3 - K$ corresponding to the Wirtinger presentation of $\pi_1(S^3 - K)$, obtaining a 2-complex X homotopy equivalent to $S^3 - K$. If $\widetilde{X}$ denotes the universal covering of X, then $\pi_2(X) \cong \pi_2(\widetilde{X}) \cong H_2(\widetilde{X})$, the last isomorphism being a consequence of the Hurewicz theorem. Moreover, $H_2(\widetilde{X}) \cong \ker \partial$, where $\partial : C_2(\widetilde{X}) \to C_1(\widetilde{X})$ is the boundary homomorphism.

Note also that $C_2(\widetilde{X})$ and $C_1(\widetilde{X})$ are the free $\mathbb{Z}\pi_1(X)$-modules on the 2-cells and 1-cells respectively of X. Thus $\pi_2(X) = 0$ if and only if ∂ is injective, and in this way the asphericity problem becomes equivalent to an assertion about a finite system of linear equations over $\mathbb{Z}\pi_1(X)$. Pointing out that this works for any graph in S^3, Whitehead, by explicit calculation, shows that the complements of the figure eight knot, the Whitehead link [122], and a certain knotted wedge of two circles, are all aspherical.

Next, Whitehead proves that if X_1, X_2 and $X_1 \cap X_2$ are aspherical polyhedra such that $\pi_1(X_1 \cap X_2) \to \pi_1(X_i)$ is injective, $i = 1, 2$, then $X = X_1 \cup X_2$ is aspherical. (He says the proof is mainly due to Eilenberg.) He uses this to show that the asphericity of the complement of a link L is preserved by doubling (in the sense of [122]) a component K of L, provided that $\pi_1(T) \to \pi_1(S^3 - L)$ is injective, where T is the boundary of a tubular neighborhood of K. Recalling that Eilenberg had remarked that the asphericity of

$S^3 - L$ reflects some sort of *linking* of the two components of L, Whitehead points out that his doubling construction shows that $S^3 - L$ may be aspherical even though the two components of L are only linked in a very weak sense, specifically, for any given n, L may be chosen so that the components are not *n-linked* in the sense of Eilenberg [29]. This observation leads him to ask:

> *If X is a closed subset of S^3, is $S^3 - X$ aspherical unless X is a disjoint union $X_1 \coprod X_2$, $X_1 \neq \emptyset \neq X_2$, where X_1 is contained in a 3-cell which does not meet X_2?*

He notes that this is equivalent to:

> *If U is an open subset of S^3, is $\pi_2(U) = 0$ provided every embedded S^2 in U bounds a 3-cell in U?*

Thus Whitehead has arrived at the right "conjecture" about the asphericity of submanifolds of S^3. As Papakyriakopoulos says in [74]:

> It was precisely this conjecture which stimulated the present author to prove during the summer of 1956 the following sphere theorem.

In 1947 Higman took up the asphericity question [46], and used Whitehead's algebraic formulation to show that if L is a link in S^3 such that $\pi_1(S^3 - L)$ is a nontrivial free product, then $\pi_2(S^3 - L) \neq 0$, generalizing Eilenberg's result mentioned above.

The only further progress on the question of the "asphericity of knots and links", until Papakyriakopoulos' complete solution in 1957, was Aumann's proof [10] that complements of alternating knots and links are aspherical. This goes as follows. Let D be a reduced, alternating, connected diagram of a knot or link K in S^3. Shading the complementary regions of the diagram alternately black and white, we see that it determines two spanning surfaces for K. Let F be one of these surfaces, and assume for convenience that F is non-orientable (in fact this can always be arranged if K is a knot and D has a nonzero number of crossings); the orientable case is similar. The surface F has a neighborhood X_1 (a twisted I-bundle over F), such that X_1 and $X_2 = \overline{S^3 - X_1}$ are handlebodies. Then X, the complement of an open neighborhood of K, can be expressed as $X_1 \cup X_2$, where $X_1 \cap X_2 = \widetilde{F}$ is the 2-fold orientable cover of F. Clearly, the map $\pi_1(\widetilde{F}) \to \pi_1(X_1)$ is injective, and Aumann shows, using the fact that D is reduced and alternating, that the map $\pi_1(\widetilde{F}) \to \pi_1(X_2)$ is also injective. The asphericity of X now follows from the result of Whitehead mentioned above.

Appearing as it did just before [72], Aumann's result was overshadowed by that of Papakyriakopoulos. However, it has a feature which was to emerge later as an important notion in knot theory. Namely, taking K to be a knot, one can show that F can be chosen so that its boundary is not a latitude of K, and so the incompressible surface $\widetilde{F}$ represents a nonzero *boundary slope* of K. The potential usefulness of incompressible surfaces with boundary in knot complements was emphasized by Neuwirth [69], and it was later proved by Culler and Shalen [20], using deep results on representations of knot groups in $PSL_2(\mathbb{C})$, that *every* (nontrivial) knot has a nonzero boundary slope. It turns out that the boundary slopes of a knot K play an important role in the study of the manifolds obtained by Dehn surgery on K.

Now we come to the sphere theorem, which asserts that if an orientable 3-manifold contains a singular homotopically essential 2-sphere then it contains a nonsingular one.

SPHERE THEOREM. *Let M be an orientable 3-manifold such that $\pi_2(M) \neq 0$. Then M contains an embedded 2-sphere which is not null-homotopic.*

Papakyriakopoulos proved a "conditional" version of the sphere theorem at the same time that he proved Dehn's lemma [72]. In fact, his proof of the former is modelled exactly on that of the latter, and it is because of this that he needs an extra hypothesis, to deal with the case where n, the height of the tower, is 1, and the sphere S_1 (which plays the role of the disk D_1 in the proof of Dehn's lemma) is nonsingular. To ensure the existence of a covering transformation τ of M_1 of infinite order such that $S_1 \cap \tau(S_1) \neq \emptyset$, he needs to assume that $H_1(V_0)$ is infinite. This follows as before if ∂V_0 does not consist of 2-spheres, but to take account of the possibility that it does, Papakyriakopoulos adds the hypothesis that M embeds in a 3-manifold N such that any nontrivial finitely generated subgroup of $\pi_1(N)$ has infinite commutator quotient group, (in other words, $\pi_1(N)$ is *locally indicable*). Since this condition is vacuously satisfied if $\pi_1(N) = 1$, Papakyriakopoulos' version is enough to prove the asphericity of knots; more generally, it proves Whitehead's conjecture characterizing the aspherical open subsets of S^3.

The additional hypothesis was soon shown to be unnecessary. Quoting Papakyriakopoulos [74, p. 319]:

> In October 1957 J.W. Milnor proved a more general sphere theorem. Finally in December 1957 J.H.C. Whitehead ... proved the sphere theorem in complete generality.

Whitehead [125] achieved this by making the following modifications to the definition of the tower. Firstly, the tower stops when $\pi_1(V_n)$ is finite, as opposed to trivial. Secondly, the coverings $M_i \to V_{i-1}$ are universal, as before, except for the first, $M_1 \to V_0$, which is defined to be that corresponding to the cyclic subgroup of $\pi_1(V_0)$ generated by a nontrivial covering transformation τ of the universal covering $\widetilde{V}_0 \to V_0$ such that $S_1 \cap \tau(S_1) \neq \emptyset$, where S_1 is a lift of the original 2-sphere $S_0 \subset V_0 \subset M$ to $\widetilde{V}_0$. Incidentally, the papers [98, 125] marked Whitehead's return to 3-dimensional topology after an absence of almost twenty years.

As we saw in Section 11, Shapiro and Whitehead, and Stallings, showed that 2-fold coverings could be used to considerably simplify the proofs of Dehn's lemma and the loop theorem, but, interestingly, this does not work for the sphere theorem. Perhaps it is for this reason that Stallings says in [104] that

> The proofs of Dehn's lemma and the Loop Theorem are an order of magnitude easier than is the proof of the Sphere Theorem.

The difference between the theorems is also reflected in Johannson's approach, using hierarchies. While this gives a proof of the disk theorem, in the case of the sphere theorem it merely reduces the problem to proving that if M is a closed, orientable, irreducible, non-Haken 3-manifold then $\pi_2(M) = 0$.

The sphere theorem is false for non-orientable 3-manifolds, as the example $\mathbb{RP}^2 \times S^1$ shows. However, Epstein, in his Cambridge Ph.D. dissertation (see [32]), showed that the following version, which he calls the projective plane theorem, holds without the assumption of orientability: if M is a 3-manifold such that $\pi_2(M) \neq 0$, then there is an essential map $S^2 \to M$ which either is an embedding or has image a 2-sided projective plane. In the case that M is non-orientable, this is proved by going to the 2-fold orientable cover $\widetilde{M}$ of M, taking an essential embedded 2-sphere S in $\widetilde{M}$ (whose existence is guaranteed by the sphere theorem), and doing a cut and paste argument on $S \cup \tau(S)$, where τ is the nontrivial covering transformation of $\widetilde{M}$. So here we see another application of 2-fold coverings in this context, similar to the ones we have already met.

Another line of development here is the relationship between $\pi_2(M)$ and the *ends* of $\pi_1(M)$. The theory of ends was initiated in Freudenthal's 1930 Berlin dissertation (see [37]), and the notion of the number of ends $e(G)$ of a group G was defined by Hopf [49], who proved that $e(G) = 0, 1, 2$, or ∞. In [101], Specker used this theory to show that for a closed 3-manifold M, $\pi_2(M)$ is determined by $\pi_1(M)$, a fact which had been announced without proof by Hopf [48]; more precisely, he showed that $\pi_2(M)$ is a free Abelian group of rank n, where $n = 0, 1, \infty$, according as $e(\pi_1(M))$ is less than 2, 2, or ∞. This is proved by applying Poincaré duality (using cohomology based on finite cochains) in the universal cover of M.

Applying similar considerations to 3-manifolds with boundary, Specker also showed that the asphericity of knots is equivalent to the assertion that the number of ends of a knot group is 1 or 2. In particular, since the group of a torus knot has an infinite cyclic center, it follows from a theorem of Hopf [49] that it has 1 or 2 ends; and, hence, that complements of torus knots are aspherical. But except in this special case, the equivalence established by Specker did not lead to progress on the asphericity question; rather, it was the other way round: after Papakyriakopoulos proved his sphere theorem he could deduce the fact about ends of knot groups. However, the direction of implication that no doubt Specker had in mind did eventually reappear, in the later work of Stallings [104].

Specker's paper also contains an application to the question of which Abelian groups can be fundamental groups of (compact) 3-manifolds. For closed, orientable 3-manifolds this was solved earlier by Reidemeister [87]. Reidemeister observes that for such a manifold M, $\pi_1(M)$ has a finite presentation with the same number of generators as relations, and shows that the only Abelian groups with this property are $\mathbb{Z}$, $\mathbb{Z}_n$ $(n \geqslant 1)$, $\mathbb{Z} \times \mathbb{Z}$, $\mathbb{Z} \times \mathbb{Z}_n$ $(n \geqslant 2)$, and $\mathbb{Z} \times \mathbb{Z} \times \mathbb{Z}$. The main part of the proof is now to rule out the possibilities $\mathbb{Z} \times \mathbb{Z}$ and $\mathbb{Z} \times \mathbb{Z}_n$, which is done by using duality in the cellular chains of the universal covering.

Specker considers manifolds with boundary, and shows that if M is a compact, orientable 3-manifold whose boundary is nonempty and does not consist entirely of 2-spheres, and whose fundamental group $\pi_1(M)$ is Abelian, then $\pi_1(M) \cong \mathbb{Z}$ or $\mathbb{Z} \times \mathbb{Z}$.

Finally, in [32] Epstein proves that the only finitely generated Abelian groups that can be subgroups of the fundamental group of any 3-manifold (not necessarily paracompact or orientable) are $\mathbb{Z}$, $\mathbb{Z}_n$, $\mathbb{Z} \times \mathbb{Z}$, $\mathbb{Z} \times \mathbb{Z}_2$, and $\mathbb{Z} \times \mathbb{Z} \times \mathbb{Z}$. (He also shows that if M is a compact non-orientable 3-manifold with $\pi_1(M)$ finite, then $\pi_1(M) \cong \mathbb{Z}_2$, and in fact M is homotopy equivalent to $\mathbb{RP}^2 \times I$ with a finite number of open 3-balls removed.)

The paper [126] of Whitehead is another which fits into this general end-theoretic context. Here he proves that if M is a 3-manifold, then $\pi_1(M)$ is isomorphic to $\mathbb{Z}$ or a nontrivial free product if and only if M contains an essential embedded 2-sphere. For the nontrivial implication, Whitehead notes that the hypothesis on $\pi_1(M)$ implies that the number of ends of $\pi_1(M)$ is 2 or ∞, and hence, by Specker [101], $\pi_2(M) \neq 0$. If M is orientable, the result is now a consequence of the sphere theorem. Whitehead shows that the statement also holds in the non-orientable case. However, this follows easily from Epstein's projective plane theorem.

We now go back to Kneser's 1929 paper [56]. In the last section of that paper he gives a proof of the following statement.

KNESER'S CONJECTURE. *If M is a 3-manifold such that $\pi(M) \cong G_1 * G_2$, then $M \cong M_1 \# M_2$, where $\pi_1(M_i) \cong G_i$, $i = 1, 2$.*

This is another manifestation of the principle that, in dimension 3, the fundamental group determines the topology.

Kneser's argument is hard to follow, but may be roughly summarized thus. First, a 2-dimensional spine of the 3-manifold M is modified to get two disjoint 2-complexes X_1, X_2 in M, with $\pi_1(X_i) \cong G_i$, $i = 1, 2$, whose inclusions into M induce the natural inclusions of the factors G_i into $G_1 * G_2 \cong \pi_1(M)$. Next one finds a closed surface F in M, separating X_1 and X_2, such that the map $\pi_1(F) \to \pi_1(M)$ is trivial. Using Kneser's Hilfsatz, F may be compressed to a disjoint union of 2-spheres. If two of the 2-spheres can be joined by a path that is null-homotopic in M, then $X = X_1 \coprod X_2$ can be changed so that it misses this path, and then the 2-spheres can be connected by a tube to form a single 2-sphere. Doing this as often as possible, Kneser argues that one must end up with a single 2-sphere, separating M into two components M_1', M_2' with $\pi_1(M_i') \cong G_i$, $i = 1, 2$.

It was soon after writing this paper that Kneser discovered the flaw in Dehn's proof of his lemma, and so he says, in a footnote added in proof, that because of this his proof should be considered incomplete. However, his argument is sufficiently unclear that even when the Hilfsatz was finally established, with the proof of the loop theorem and Dehn's lemma, his theorem was still not regarded as having been proved, and Papakyriakopoulos in [74] therefore termed it "Kneser's conjecture". Papakyriakopoulos envisages that one would approach this conjecture in two steps:

> This suggests that the gap between [the hypothesis] and the conclusion of Kneser's conjecture is so great that it has to be factored, and we first have to prove that [the hypothesis] implies $\pi_2 \neq 0$, and then that $\pi_2 \neq 0$ implies the desired conclusion. It seems that the first step has to be proved by *algebraic topological* techniques, and the second one by using the sphere theorem and *something more*, because the sphere theorem is not enough to provide us with the conclusions of Kneser's conjecture. Thinking now that the algebraic topological techniques were rather undeveloped in 1928, we easily conclude that it was rather hopeless, to expect to have a satisfying proof of this strong statement at that time.

Nevertheless, Stallings, to whom Papakyriakopoulos had suggested the problem, gave a proof of Kneser's conjecture, in his 1959 Princeton Ph.D. thesis [102], which did not follow this scheme, but which was, in outline, very much along the lines indicated by Kneser.

In particular, Stallings' proof did not use the sphere theorem. Note, however, that the conclusion clearly implies (if the groups G_1 and G_2 are nontrivial) that M contains an embedded essential 2-sphere. This leads us to Stallings' work of about ten years later, which brings these ideas about π_2 and ends of π_1 to a full circle. The key is Stallings' result [104] that a finitely generated group G has $e(G) \geqslant 2$ if and only if G is a nontrivial free product with amalgamation $A *_F B$, or an HNN-extension $A*_F$, where the amalgamating subgroup F is finite. Now let M be a (say, closed, orientable) 3-manifold with $\pi_2(M) \neq 0$. Applying the Hurewicz theorem and Poincaré duality to the universal covering of M, it follows easily that $e(\pi_1(M)) \geqslant 2$. Hence, $G = \pi_1(M)$ splits as described above over a finite group F. Constructing a $K(G, 1)$ space K_G containing a bicollared copy of a $K(F, 1)$ space K_F, we have a map $f : M \to K_G$ inducing an isomorphism on fundamental groups, and we may assume by transversality that $f^{-1}(K_F)$ is a 2-sided surface S in M. Using the disk theorem, the map f may be homotoped so that for each component S_0 of

S, $\pi_1(S_0) \to \pi_1(M)$ is injective. Since $\pi_1(K_F) \cong F$ is finite, this implies that S_0 is a 2-sphere. Since null-homotopic components S_0 may be eliminated by a further homotopy of f, Stallings concludes that M contains an essential embedded 2-sphere.

This work demonstrates a deep relationship between the fundamental group of a 3-manifold and its topology, and indeed Stallings sees the connection between group theory and 3-dimensional topology in even broader terms. In [104, pp. 1, 2] he makes the following remarks, which have been vindicated by recent work in geometric group theory:

> Philosophically speaking, the depth and beauty of 3-manifold theory is, it seems to me, mainly due to the fact that its theorems have off-shoots that eventually blossom in a different subject, namely group theory. Thus I tend to believe that new results in the theory, such as Waldhausen's [115], may eventually have relatives in group theory; the solution of the Poincaré Conjecture [81], if it ever occurs, will have group-theoretic consequences of a wider nature.

Bibliography

[1] C.C. Adams, *The noncompact hyperbolic 3-manifold of minimal volume*, Proc. Amer. Math. Soc. **100** (1987), 601–606.

[2] J.W. Alexander, *Note on two 3-dimensional manifolds with the same group*, Trans. Amer. Math. Soc. **20** (1919), 339–342.

[3] J.W. Alexander, *Note on Riemann spaces*, Bull. Amer. Math. Soc. **26** (1919), 370–372.

[4] J.W. Alexander, *On the subdivision of 3-space by a polyhedron*, Proc. Nat. Acad. Sci. **10** (1924), 6–8.

[5] J.W. Alexander, *An example of a simply connected surface bounding a region which is not simply connected*, Proc. Nat. Acad. Sci. **10** (1924), 8–10.

[6] J.W. Alexander, *Remarks on a point set constructed by Antoine*, Proc. Nat. Acad. Sci. **10** (1924), 10–12.

[7] J.W. Alexander, *New topological invariants expressible as tensors*, Proc. Nat. Acad. Sci. **10** (1924), 99–101.

[8] J.W. Alexander, *On certain new topological invariants of a manifold*, Proc. Nat. Acad. Sci. **10** (1924), 101–103.

[9] J.W. Alexander, *Some problems in topology*, Verhandlungen des Internationalen Mathematiker-Kongresses Zürich (1932), Kraus Reprint (1967), 249–257.

[10] R.J. Aumann, *Asphericity of alternating knots*, Ann. of Math. **64** (1956), 374–392.

[11] L.A. Best, *On torsion-free discrete subgroups of PSL(2, C) with compact orbit space*, Can. J. Math. **23** (1971), 451–460.

[12] L. Bianchi, *Geometrische Darstellung der Gruppen linearer Substitutionen mit ganzen complexen Coefficienten nebst Anwendungen auf die Zahlentheorie*, Math. Ann. **38** (1891), 313–333.

[13] R.H. Bing, *An alternative proof that 3-manifolds can be triangulated*, Ann. of Math. **69** (1959), 37–65.

[14] J.S. Birman, F. González-Acuña and J.M. Montesinos, *Heegaard splittings of prime 3-manifolds are not unique*, Michigan Math. J. **23** (1976), 97–103.

[15] E.J. Brody, *The topological classification of lens spaces*, Ann. of Math. **71** (1960), 163–184.

[16] F.E. Browder (ed.), *The Mathematical Heritage of Henri Poincaré*, Proc. Sympos. Pure Math. vol. 39, Amer. Math. Soc., Providence, RI (1983).

[17] A.J. Casson and C.McA. Gordon, *A loop theorem for duality spaces and fibred ribbon knots*, Invent. Math. **74** (1983), 119–137.

[18] B. Chandler and W. Magnus, *The History of Combinatorial Group Theory: A Case Study in the History of Ideas*, Stud. Hist. Math. Phys. Sci. vol. 9, Springer, New York (1982).

[19] R. Craggs, *A new proof of the Reidemeister–Singer theorem on stable equivalence of Heegaard splittings*, Proc. Amer. Math. Soc. **57** (1976), 143–147.

[20] M. Culler and P.B. Shalen, *Bounded, separating, incompressible surfaces in knot manifolds*, Invent. Math. **75** (1984), 537–545.

[21] M. Dehn, *Berichtigender Zusatz zu III AB 3 Analysis Situs*, Jahresber. Deutsch. Math.-Verein. **16** (1907), 573.

[22] M. Dehn, *Über die Topologie des dreidimensionalen Raumes*, Math. Ann. **69** (1910), 137–168.
[23] M. Dehn, *Die beiden Kleeblattschlingen*, Math. Ann. **75** (1914), 1–12.
[24] M. Dehn, *Papers on Group Theory and Topology*, translated and introduced by J. Stillwell, Springer, New York (1987).
[25] M. Dehn and P. Heegaard, *Analysis situs*, Enzyklopädie Math. Wiss. III, AB 3, Teubner, Leipzig (1907), 153–220.
[26] J. Dieudonné, *A History of Differential and Algebraic Topology 1900–1960*, Birkhäuser, Boston (1989).
[27] W. Dyck, *On the "Analysis Situs" of 3-dimensional spaces*, Report of the Brit. Assoc. Adv. Sci. (1884), 648.
[28] S. Eilenberg, *Sur les courbes sans noeuds*, Fund. Math. **28** (1936), 233–242.
[29] S. Eilenberg, *Sur les espaces multicoherents II*, Fund. Math. **29** (1937), 101–122.
[30] S. Eilenberg, *On the problems of topology*, Ann. of Math. **50** (1949), 247–260.
[31] R. Engmann, *Nicht-homöomorphe Heegaard-Zerlegungen vom Geschlecht 2 der zusammenhängenden Summe zweier Linsenräume*, Abh. Math. Sem. Univ. Hamburg **35** (1970), 33–38.
[32] D.B.A. Epstein, *Projective planes in 3-manifolds*, Proc. London Math. Soc. **11** (1961), 469–484.
[33] R.H. Fox, *On the imbedding of polyhedra in 3-space*, Ann. of Math. **49** (1948), 462–470.
[34] F. Frankl, *Zur Topologie des dreidimensionalen Raumes*, Monatsh. Math. Phys. **38** (1931), 357–364.
[35] W. Franz, *Über die Torsion einer Überdeckung*, J. Reine Angew. Math. **173** (1935), 245–254.
[36] W. Franz, *Abbildungsklassen und Fixpunktklassen dreidimensionaler Linsenräume*, J. Reine Angew. Math. **185** (1943), 65–77.
[37] H. Freudenthal, *Über die Enden topologischer Raüme und Gruppen*, Math. Z. **33** (1931), 692–713.
[38] H. Gieseking, *Analytische Untersüchen über topologische Gruppen*, Ph.D. thesis, Münster (1912).
[39] L. Goeritz, *Die Heegaard-Diagramme des Torus*, Abh. Math. Sem. Univ. Hamburg **9** (1932), 187–188.
[40] J. Hadamard, *L'oeuvre mathématique de Poincaré*, Acta Math. **38** (1921), 203–287.
[41] W. Haken, *Über Flächen in 3-dimensionalen Mannigfaltigkeiten Lösung des Isotopieproblems für den Kreisknoten*, Proc. of the Internat. Congress of Mathematicians, Amsterdam (1954), Vol. 1, North-Holland, Amsterdam (1957), 481–482.
[42] W. Haken, *Theorie der Normalfächen*, Acta Math. **105** (1961), 245–375.
[43] W. Hantzsche and H. Wendt, *Dreidimensionale euklidische Raumformen*, Math. Ann. **110** (1935), 593–611.
[44] P. Heegaard, *Forstudier til en topologisk Teori for de algebraiske Fladers Sammenhaeng*, Dissertation, Copenhagen (1898).
[45] P. Heegaard, *Sur l'"Analysis Situs"*, Bull. Soc. Math. France **44** (1916), 161–242.
[46] G. Higman, *A theorem on linkages*, Quart. J. Math. Oxford **19** (1948), 117–122.
[47] H. Hopf, *Zum Clifford–Kleinscher Raumformen*, Math. Ann. **95** (1925), 313–339.
[48] H. Hopf, *Räume, die Transformationsgruppen mit kompacten Fundamentalbereichen gestatten*, Verhand. Schweizer. Natur. Gesellschaft (1942), 79.
[49] H. Hopf, *Enden offener Räume und unendliche diskontinuierliche Gruppen*, Comment. Math. Helv. **16** (1944), 81–100.
[50] H. Hotelling, *Three-dimensional manifolds and states of motion*, Trans. Amer. Math. Soc. **27** (1925), 329–344.
[51] H. Hotelling, *Multiple-sheeted spaces and manifolds of states of motion*, Trans. Amer. Math. Soc. **28** (1926), 479–490.
[52] I. Johannson, *Über singuläre Elementarflächen und das Dehnsche Lemma*, Math. Ann. **110** (1935), 312–320.
[53] I. Johannson, *Über singuläre Elementarflächen und das Dehnsche Lemma II*, Math. Ann. **115** (1938), 658–669.
[54] K. Johannson, *On the loop- and sphere-theorem*, Low-Dimensional Topology, K. Johannson, ed., Conf. Proc. and Lecture Notes in Geometry and Topology vol. 3, International Press (1994), 47–54.
[55] H. Kneser, *Eine Bemerkung über dreidimensionale Mannigfaltigkeiten*, Nachr. Ges. Wiss. Göttingen (1925), 128–130.
[56] H. Kneser, *Geschlossene Flächen in driedimensionale Mannigfaltigkeiten*, Jahresber. Deutsch. Math.-Verein. **38** (1929), 248–260.
[57] M. Kreines, *Zur Konstruktion der Poincaré-Räume*, Rend. Circ. Mat. Palermo **56** (1932), 277–280.
[58] R. Lee, *Semicharacteristic classes*, Topology **12** (1973), 183–200.

[59] F. Löbell, *Beispiele geschlossener dreidimensionaler Clifford–Kleinscher Räume negativ Krümmung*, Ber. Sächs. Akad. Wiss. **83** (1931), 168–174.
[60] W. Magnus, *Non-Euclidean Tesselations and Their Groups*, Academic Press, New York (1974).
[61] A.A. Markov, *The insolubility of the problem of homeomorphy*, Dokl. Akad. Nauk USSR **121** (1958), 218–220.
[62] J. Milnor, *Groups which act on S^n without fixed points*, Amer. J. Math. **79** (1957), 623–630.
[63] J. Milnor, *A unique factorization theorem for 3-manifolds*, Amer. J. Math. **84** (1962), 1–7.
[64] J. Milnor, *Hyperbolic geometry: The first 150 years*, Bull. Amer. Math. Soc. **6** (1982), 9–24.
[65] E. Moise, *Affine structures in 3-manifolds V. The triangulation theorem and Hauptvermutung*, Ann. of Math. **56** (1952), 96–114.
[66] J.W. Morgan and H. Bass (eds), *The Smith Conjecture*, Academic Press, New York (1984).
[67] E.S. Munkholm and H.J. Munkholm, *Poul Heegaard*, History of Topology, I.M. James, ed., Elsevier, Amsterdam (1999), 925–946.
[68] M.H.A. Newman and J.H.C. Whitehead, *On the group of a certain linkage*, Quart. J. Math. Oxford **8** (1937), 14–21.
[69] L.P. Neuwirth, *Interpolating manifolds for knots in S^3*, Topology **2** (1963), 359–365.
[70] W. Nowacki, *Die euklidischen, dreidimensionalen, geschlossen und offenen Raumformen*, Comment. Math. Helv. **7** (1934), 81–93.
[71] C.D. Papakyriakopoulos, *On solid tori*, Proc. London Math. Soc. **7** (1957), 281–299.
[72] C.D. Papakyriakopoulos, *On Dehn's lemma and the asphericity of knots*, Ann. of Math. **66** (1957), 1–26.
[73] C.D. Papakyriakopoulos, *On the ends of the fundamental groups of 3-manifolds with boundary*, Comment. Math. Helv. **32** (1957), 85–92.
[74] C.D. Papakyriakopoulos, *Some problems on 3-dimensional manifolds*, Bull. Amer. Math. Soc. **64** (1958), 317–335.
[75] C.D. Papakyriakopoulos, *The theory of three-dimensional manifolds since 1950*, Proc. of the Internat. Congress of Mathematicians, Cambridge (1958), Cambridge Univ. Press, Cambridge (1960), 433–440.
[76] E. Picard, *Sur un groupe de transformations des points de l'espace situés du même côté d'un plan*, Bull. Soc. Math. France **12** (1884), 43–47.
[77] H. Poincaré, *Sur l'analysis situs*, Comptes Rendus **115** (1882), 633–636.
[78] H. Poincaré, *Mémoire sur les Groupes Kleiniens*, Acta Math. **3** (1883), 49–92.
[79] H. Poincaré, *Analysis Situs*, J. École Polytech. Paris (2) **1** (1895), 1–121.
[80] H. Poincaré, *Second complément à l'Analysis Situs*, Proc. London Math. Soc. **32** (1900), 277–308.
[81] H. Poincaré, *Cinquième complément à l'Analysis Situs*, Rend. Circ. Mat. Palermo **18** (1904), 45–110.
[82] H. Poincaré, *Analyse de ses travaux scientifiques*, Acta Math. **38** (1921), 3–135.
[83] J. Przytycki, *Knot theory from Vandermonde to Jones* (with the translation of the topological part of Poul Heegaard's dissertation, by A.H. Przybyszewska), Preprint 43, Odense Universitet, Denmark (1993).
[84] K. Reidemeister, *Zur dreidimensionalen Topologie*, Abh. Math. Sem. Univ. Hamburg **9** (1933), 189–194.
[85] K. Reidemeister, *Heegaarddiagramme und Invarianten von Mannigfaltigkeiten*, Abh. Math. Sem. Univ. Hamburg **10** (1934).
[86] K. Reidemeister, *Homotopieringe und Linsenräume*, Abh. Math. Sem. Univ. Hamburg **11** (1935), 102–109.
[87] K. Reidemeister, *Kommutative Fundamentalgruppen*, Monatsh. Math. Phys. **43** (1936), 20–28.
[88] G. de Rham, *Sur l'Analysis situs des variétés à n dimensions*, J. Math. Pures Appl. **10** (1931), 115–120.
[89] R. Riley, *A quadratic parabolic group*, Math. Proc. Cambridge Philos. Soc. **77** (1975), 281–288.
[90] J.H. Rubinstein, *An algorithm to recognize the 3-sphere*, Proc. of the Internat. Congress of Mathematicians, Zürich (1994), Vol. 1, Birkhäuser (1995), 601–611.
[91] M. Rueff, *Beiträge zur Untersuchung der Abbildungen von Mannigfaltigkeiten*, Compositio Math. **6** (1938), 161–202.
[92] P. Scott, *The geometries of 3-manifolds*, Bull. London Math. Soc. **15** (1983), 401–487.
[93] H. Seifert, *Konstruktion dreidimensionaler geschlossener Räume*, Ber. Sächs. Akad. Wiss. **83** (1931), 26–66.
[94] H. Seifert, *Homologiegruppen berandeter driedimensionaler Mannigfaltigkeiten*, Math. Z. **35** (1932), 609–611.
[95] H. Seifert, *Topologie dreidimensionaler gefaserter Räume*, Acta Math. **60** (1932), 147–238.
[96] H. Seifert, *Verschlingungsinvarianten*, Sitzungsber. Preuss. Akad. Wiss. **16** (1933), 811–828.

[97] H. Seifert and W. Threlfall, *A Textbook of Topology*, Academic Press, New York (1980), Translation of *Lehrbuch der Topologie*, Teubner, Leipzig (1934).
[98] A.S. Shapiro and J.H.C. Whitehead, *A proof and extension of Dehn's lemma*, Bull. Amer. Math. Soc. **64** (1958), 174–178.
[99] L.C. Siebenmann, *Les bisections expliquent le théorème de Reidemeister–Singer, un retour aux sources*, Preprint, Orsay (1979).
[100] J. Singer, *Three-dimensional manifolds and their Heegaard diagrams*, Trans. Amer. Math. Soc. **35** (1933), 88–111.
[101] E. Specker, *Die erste Cohomologiegruppe von Überlagerungen und Homotopieeigenschaften dreidimensionaler Mannigfaltigkeiten*, Comment. Math. Helv. **23** (1949), 303–332.
[102] J.R. Stallings, *Some topological proofs and extensions of Gruško's theorem*, Dissertation, Princeton University (1959).
[103] J.R. Stallings, *On the loop theorem*, Ann. of Math. **72** (1960), 12–19.
[104] J. Stallings, *Group Theory and 3-Dimensional Manifolds*, Yale Math. Mongraphs vol. 4, Yale Univ. Press, New Haven, CT (1971).
[105] J. Stillwell, *Sources of Hyperbolic Geometry*, Hist. Math. vol. 10, Amer. Math. Soc., Providence, RI (1996).
[106] A. Thompson, *Thin position and the recognition problem for S^3*, Math. Research Letters **1** (1994), 613–630.
[107] W. Threlfall and H. Seifert, *Topologische Untersuchung der Discontinuitätsbereiche endlicher Bewegungsgruppen der dreidimensionalen sphärischen Raumes I*, Math. Ann. **104** (1930), 1–70.
[108] W. Threlfall and H. Seifert, *Topologische Untersuchung der Discontinuitätsbereiche endlicher Bewegungsgruppen der dreidimensionalen sphärischen Raumes II*, Math. Ann. **107** (1932), 543–586.
[109] W.P Thurston, *Three-dimensional manifolds, Kleinian groups and hyperbolic geometry*, Bull. Amer. Math. Soc. **6** (1982), 357–381.
[110] H. Tietze, *Über die topologischen invarianten mehrdimensionaler Mannigfaltigkeiten*, Monatschr. Math. Phys. **19** (1908), 1–118.
[111] B. Trace, *Two comments concerning the uniqueness of prime factorizations for 3-manifolds*, Bull. London Math. Soc. **19** (1987), 75–77.
[112] A.M. Turing, *Solvable and unsolvable problems*, Sci. News **31** (1954), 7–23.
[113] K. Volkert, *The early history of Poincaré's conjecture*, Preprint, Heidelberg (1994).
[114] F. Waldhausen, *Heegaard-Zerlegungen der 3-Sphäre*, Topology **7** (1968), 195–203.
[115] F. Waldhausen, *On irreducible 3-manifolds which are sufficiently large*, Ann. of Math. **87** (1968), 56–88.
[116] F. Waldhausen, *The word problem in fundamental groups of sufficiently large 3-manifolds*, Ann. of Math. **88** (1968), 272–280.
[117] F. Waldhausen, *Recent results on sufficiently large 3-manifolds*, Proc. Sympos. Pure Math. vol. 32, Amer. Math. Soc., Providence, RI (1978), 21–38.
[118] C. Weber and H. Seifert, *Die beiden Dodekaederräume*, Math. Z. **37** (1933), 237–253.
[119] J.H.C. Whitehead, *Certain theorems about 3-dimensional manifolds* (1), Quart. J. Math. Oxford **15** (1934), 308–320; *3-dimensional manifolds (corrigendum)*, ibid. **6** (1935), 80.
[120] J.H.C. Whitehead, *A certain region in Euclidean 3-space*, Proc. Nat. Acad. Sci. **21** (1935), 364–366.
[121] J.H.C. Whitehead, *A certain open manifold whose group is unity*, Quart. J. Math. Oxford **6** (1935), 268–279.
[122] J.H.C. Whitehead, *On doubled knots*, J. London Math. Soc. **12** (1937), 63–71.
[123] J.H.C. Whitehead, *On the asphericity of regions in a 3-sphere*, Fund. Math. **32** (1939), 149–166.
[124] J.H.C. Whitehead, *On incidence matrices, nuclei and homotopy types*, Ann. of Math. **42** (1941), 1197–1239.
[125] J.H.C. Whitehead, *On 2-spheres in 3-manifolds*, Bull. Amer. Math. Soc. **64** (1958), 161–166.
[126] J.H.C. Whitehead, *On finite cocycles and the sphere theorem*, Colloq. Math. **6** (1958), 271–281.

This Page Intentionally Left Blank

CHAPTER 16

A Short History of Triangulation and Related Matters

N.H. Kuiper* (1920–1994)

1. Triangulation in the work of L.E.J. Brouwer

Real understanding in mathematics means an intuitive simple grasp of a fact. Therefore the urge to understand will seek satisfaction in simplicity of stated theorems, simplicity of methods and proofs, and simplicity of tools. It is this simplicity which can give rise to a sensation of beauty that goes with real understanding. This does not exclude admiration for a proof that is difficult by necessity.

Thus the specific interest of a geometrically-minded mathematician, who deals with figures like curves, surfaces, with structures like metric, group, and with relations like embedding, map, is influenced by this simplicity as well as by the success of methods and tools. Emphasis on existing tools sometimes leads to unnecessary overgrowth. As a consequence the historical development of mathematics is irregular like that of other forms of life and creation. We can see this in the stream of developing mathematics, at the origin of which Brouwer's work on manifolds, related to triangulation, has a prominent place.

Poincaré [1895] developed the *analysis situs* (the origin of algebraic topology) of algebraic manifolds V. He showed by examples that the Betti [1871] numbers do not suffice for a complete topological classification. He defined Betti groups with the help of a division of V into embedded images of convex polyhedra. Aiming at a complete classification of objects like algebraic varieties, by fitting together simple building stones one was led to take as standard parts the embedded images of straight k-simplices of various dimensions $k = 0, 1, 2, \ldots$, in number space $\mathbb{R}^N$. Any two ought to fit together in a simple way, namely by meeting, if at all in one common subsimplex. The two parametrizations by barycentric coordinates with respect to the common vertices ought to be the same also. The division of V into such simplices is called a triangulation τ. It is well-defined in so far as it consists of objects, namely simplices, whose only property is thus the dimension k, and with as relations only the incidence at certain vertices between simplices. The division can therefore be described by a "scheme" T consisting of the finite or countable set of vertices,

*Editor's note: this survey, which dates from 1977, has not been updated for the present volume, although some very minor corrections have been made.

HISTORY OF TOPOLOGY
Edited by I.M. James

together with the set of those finite subsets that carry a simplex. Nowadays we define a *topological* ($= C^0$) *triangulation* as a homeomorphism $\tau : |T| \to V$ of a simplicial complex $|T|$, the "geometric realisation" of a finite or countable scheme T (realized say in $\mathbb{R}^N$, and consisting of affine simplices), onto a topological space V.

If T and T' are "schemes" and $h : |T'| \to |T|$ is a homeomorphism which sends every simplex of $|T'|$ linearly into a simplex of $|T|$, so that every vertex of T is an image of one vertex of T', and if $\tau : |T| \to V$ is a triangulation, then the triangulation $\tau \circ h : |T'| \to V$ is called a *subdivision* of τ.

The study of the topology of a real algebraic variety or manifold V aims first of all at the definition of invariants of the underlying topological space top(V), and their calculation. At the beginning top(V) was considered too slippery to deal with. Therefore it was replaced by the triangulation $\tau : |T| \to V$, or rather the "scheme" T. The *dimension* of T is n, if $n+1$ is the maximal number of vertices of simplices of T. Also the *Euler–Poincaré number* is defined in terms of T, and so are the *Betti numbers* from the incidence matrices. But are all properties that are invariant under subdivision of a triangulation topological properties of V? They would be if the following crucial problems had a positive solution.

The triangulation problem. Is there a triangulation for every algebraic variety? For every algebraic manifold? For every topological metrizable manifold?

The Hauptvermutung. This is the affirmation of the following question. Call two triangulations $\tau_1 : |T_1| \to V$ and $\tau_2 : |T_2| \to V$ TRI-*equivalent* in case there are subdivisions $h_1 : |T_1'| \to |T_1|$ and $h_2 : |T_2'| \to |T_2|$ for two realisations $|T_1'|$ and $|T_2'|$ of one and the same "scheme" $T = T_1' = T_2'$.[1] Are any two triangulations of a given V TRI-equivalent? (Observe that the composition of homeomorphisms

$$|T_1'| \xrightarrow{h_1} |T_1| \xrightarrow{\tau_1} V \xrightarrow{\tau_2^{-1}} |T_2| \xrightarrow{h_2^{-1}} |T_2'|$$

is only required to be a homeomorphism.)

For many years people wrote inconclusive papers on these two problems.

In 1911 two papers of Brouwer [1911, 1912] on topology appeared, both outstanding in this century. In the first, only five pages long, he proves

THEOREM (*The invariance of dimension*). *If h is a homeomorphism* (1–1 *continuous map*) *of an open set $U \subset \mathbb{R}^n$ onto an open set $h(U) \subset \mathbb{R}^m$, then $m = n$.*

Brouwer's revolutionary idea and method was to approximate a continuous map f of an n-cube $D \subset \mathbb{R}^n$ (in the case at hand $f = h$), into $\mathbb{R}^m$ by a simplicial (piecewise linear, (PL)) map g: i.e. a map linear on each simplex of a triangulation of D by linear simplices.

In his key lemma, the cube has sides of length one, m equals n, $\mathbb{R}^m = \mathbb{R}^n$, and f moves every point of D over a distance at most $d < 1/2$. If g is ε-near to f for small $\varepsilon > 0$, then the image of g covers completely a concentric cube D' with sides of length $1-2d-2\varepsilon > 0$,

[1] In Dehn and Heegard [1907], the word homeomorph was still used as a synonym of TRI-equivalent between finite simplicial or convex-polyhedral complexes. Compare also the definition of pseudo-manifold in Seifert and Threlfall [1934].

because, as he shows, the "*Brouwer degree*", that is the algebraic number of oriented n-simplices covering an image point in D', is almost everywhere *one*. Therefore also the image of f covers such concentric discs D'. A simple argument completes the proof of the topological invariance of dimension.

In the second paper Brouwer defines a closed n-manifold as a topological space V *with* (in our terminology) a finite triangulation $\tau : |T| \to V$, of dimension n, whose simplices at a common vertex meet "like the linear simplices of a star in $\mathbb{R}^n$". This is now called a *Brouwer-triangulated manifold*. He proceeds with the *method of* PL-approximation and define the *degree* of a continuous map $f : M \to M'$ between closed orientable Brouwer triangulated n-manifolds. Then he proves the *invariance* of the degree *under homotopy* of f, *as well as* the invariance under any *modification of the Brouwer-triangulations* of the topological spaces underlying M and M'. This means that the degree is an invariant of a homotopy class of maps between Brouwer triangulable closed oriented n-manifolds. He applies degree theory to obtain the *Brouwer fixed point theorem*.

The notions and tools in this work were new. The papers are clear now, but they were found hard to understand at the time. Their influence became clear and effective only several years later.[2] They were fundamental for later algebraic theories of topology. Brouwer assumed triangulations in his definitions of manifolds and he used them in an exemplary way to obtain purely topological results. He must have liked his definition of manifold to be rather constructive.[3] He also must have been aware of the difficulty of the triangulation problem.

It was only many years later that S.S. Cairns [1934] gave in two papers the first proof that a smooth n-manifold (embedded in $\mathbb{R}^N$, respectively, abstractly given) has a topological Brouwer-triangulation. Brouwer [1939] presented independently a proof in a lecture for the Wiskundig Genootschap in 1937. This paper had not much impact, also because it had an unusual intuitionistic terminology. It is interesting to observe that neither Cairns nor Brouwer showed interest in C^1-triangulations nor in the Hauptvermutung. Freudenthal [1939], quoting Brouwer, extended the result and gave a proof of the existence of a C^q-triangulation $\tau : |T| \to M$ (q-times continuously differentiable on each simplex of $|T|$) for a C^q-manifold M, $q \geqslant 1$. J.H.C. Whitehead [1940], went further and completed the work by proving uniqueness as well, obtaining the TRI-equivalence of any two C^q-triangulations of M, $q \geqslant 1$. So he got a kind of smooth Hauptvermutung theorem for smooth manifolds. All TRI-triangulations obtained here were Brouwer triangulations. We denote the class of C^1-equivalence classes of C^1-manifolds by C^1, and the class of TRI-equivalence classes of Brouwer triangulated manifolds (respectively, simplicial complexes) by PL (respectively, TRI). Then the essence of the above theorems is expressed by the existence of a natural map concerning *manifolds*:

$$C^1 \to \mathrm{PL} \subset \mathrm{TRI}. \tag{1}$$

Cairns [1940a] discovered non-Brouwer triangulations of $\mathbb{R}^n$ for $n \geqslant 3$, that admit Brouwer subdivisions. He [1940b] also proposed the smoothing problem for Brouwer-triangulated n-manifolds and solved it for $n \leqslant 3$.

[2] Early, in the book of H. Weyl [1913] and in the work of J.W. Alexander [1915] who proved the topological invariance of the homology groups.

[3] Not quite constructive because it still cannot be decided whether the double cone $\Sigma(\Sigma^3)$ of a Brouwer-triangulated homotopy 3-sphere (manifold) Σ^3 is Brouwer-triangulated, by lack of a solution of the Poincaré conjecture in dimension three.

This concludes our short commentary on Brouwer's papers of 1911 and 1939 concerning triangulation.

2. Manifolds, algebraic varieties, and their triangulations

We will recall various interesting theorems and facts more or less in the chronological order of their discovery. The main diagram below organizes the problems while giving their relations. Every arrow represents a map between one class of equivalence classes of spaces into another one. The main problems and discoveries concern the injectivity and the surjectivity of these maps. The conclusions often depend on dimension.

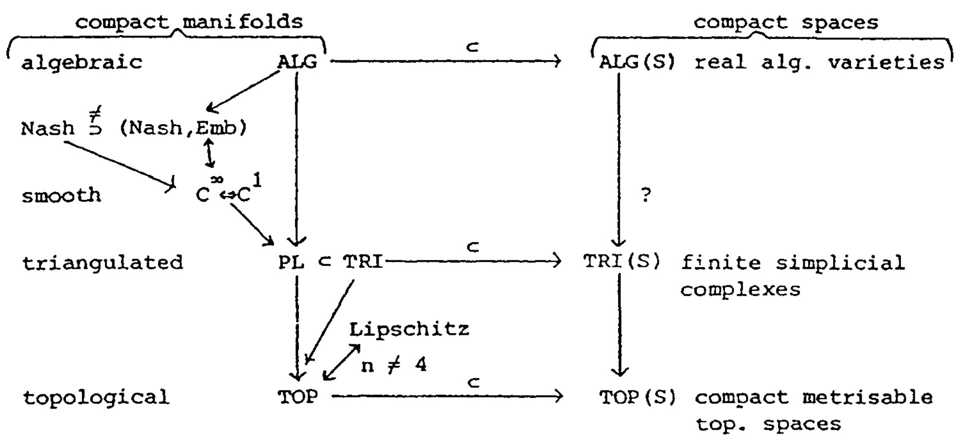
We start from the topological analysis of real algebraic varieties, because this seems, also historically, the most natural problem. It is the study of the *forget* map from equivalence classes of real algebraic varieties to their underlying topological spaces, allowing singularities (as suggested in the notation by the letter S),

$$\mathrm{ALG}(S) \to \mathrm{TOP}(S) \tag{2}$$

that arises naturally by "forgetting" part of the structure. For manifolds, for which we delete the above letter S in our notation, this map (2) factorizes with (1) and some natural forget maps to give a *diagram on manifolds*

$$\mathrm{ALG} \to \mathrm{C}^{\infty} \to \mathrm{C}^{1} \overset{(1)}{\to} \mathrm{PL} \to \mathrm{TOP}. \tag{3}$$

This is part of the *main diagram*:



For manifolds of small dimension, the expected existence and uniqueness of triangulations for topological manifolds was obtained for $n = 2$ by T. Rado [1925] and for $n = 3$ a quarter of a century later by E. Moise [1952]:

$$n \leqslant 3\colon\ \mathrm{PL} \longleftrightarrow \mathrm{TRI} \longleftrightarrow \mathrm{TOP}\ \text{bijections}.$$

Papakyriokopoulos [1943] proved the uniqueness of the TRI-structure of simplicial complexes of dimension 2:

$$n \leqslant 2:\ \mathrm{TRI}(S) \to \mathrm{TOP}(S) \text{ is injective.}$$

Geometric topology of combinatorial (PL-) manifolds developed slowly. M.H.A. Newman [1926] complained that it could not even be decided whether two subdivisions of a given Brouwer-triangulated manifold were TRI-equivalent. He started the foundations of "geometric topology", a topic much developed by E.C. Zeeman [1963]. Compare Hudson [1969] with important later work of M. Cohen in this field.

In the course of time the need for triangulations and a solution of the Hauptvermutung decreased because new homology theories of Vietoris, Čech, Alexander and the singular theory permitted purely topological definitions of invariants, although subdivisions in simplices or cells remained useful for calculating them. A milestone in algebraic topology was the axiomatic theory of Eilenberg and Steenrod [1952], which covered all older (co-) homology theories. Category and functor, notions due to S. Eilenberg and S. Mac Lane appeared as new powerful tools. Naturally algebraic topology, including the fast developing homotopy theory, dominated the field, giving a wealth of new invariants distinguishing spaces, while most people hardly dreamed of the complete classification of manifolds. The results (1) concerning the smooth triangulation of smooth (say C^1-) manifolds were isolated.

Of course manifolds existed since Grassman and Riemann [1854], and for dimension 2 the notion developed and became "more abstract" in H. Weyl's [1913] *Idee der Riemannschen Fläche*. Veblen and Whitehead [1932] formalized the definition of n-manifold M with structure S as follows. M is a connected metrizable topological space covered by images of embedded open $\mathbb{R}^n$-sets given by charts $h_i : U_i \to M$, that are related in their intersections $h_i(U_i) \cap h_j(U_j)$ by homeomorphisms of open sets in $\mathbb{R}^n$, $h_{ij} = h_j^{-1} \circ h_i$, belonging to some pseudo group S. In our present day applications, S can be the pseudogroup of homeomorphisms (TOP), C^1- or C^∞- or analytic diffeomorphisms, piecewise-linear homeomorphisms (PL), locally algebraic homeomorphisms (Nash), Lipschitz homeomorphisms, giving rise to most of the entries in our main diagram. As differentiable manifolds, embedded in $\mathbb{R}^N$ as well as abstract, became better understood, in particular under the influence of H. Whitney, it was not difficult to obtain a C^∞-structure, unique but for equivalence on any C^1-manifold:

$$C^\infty \longleftrightarrow C^1 \text{ is bijective.}$$

Manifolds being "slippery" bothered mathematicians less and less. It became also clear that PL-manifolds have a Brouwer triangulation, unique up to TRI-equivalence.

J. Nash [1952] proved that every embedded (in $\mathbb{R}^N$) compact C^1- (or C^∞-) manifold can be approximated by a diffeomorphic manifold that is also a component of a real algebraic variety. He also proved that any two mutually diffeomorphic *embedded Nash-manifolds*, are related by a diffeomorphism which is algebraic, and which is locally defined by polynomial equations:

$$(\text{Nash, embedded}) \longleftrightarrow C^\infty \longleftrightarrow C^1 \text{ bijections.}$$

There passed again a quarter of a century before A. Tognoli [1973] proved that every compact C^1-manifold is diffeomorphic to a manifold that is a whole real algebraic variety:

$$\text{ALG} \to \text{C}^1 \text{ is surjective.}$$

The Veblen–Whitehead definition of manifolds gives a larger class of Nash-manifolds:

$$\text{Nash} \overset{\neq}{\supset} (\text{Nash, embedded}).$$

An example of a nonembeddable Nash structure on the circle is obtained by identifying points in $\mathbb{R}$ by the algebraic relation $x' = x + 1$. Any function on the quotient space M yields a periodic function on $\mathbb{R}$ and cannot be algebraic unless it is constant. Hence M cannot be Nash-embedded in $\mathbb{R}^N$. It would be interesting to study all Nash structures on the circle. Perhaps all homogeneous ones admit compatible locally projective structures, as described by Kuiper [1953]. The work of J. Hubbard and Chillingworth [1971] suggests that there may be so many nonequivalent Nash-structures that a complete classification is uninteresting. Is there more than one on the two sphere?

J. Milnor [1956] made the sensational discovery of a manifold M, which is homeomorphic and PL-equivalent to the usual 7-sphere S^7, without being diffeomorphic to it:

$$\text{C}^1 \to \text{PL is not injective.}$$

This manifold M, a certain S^3-bundle over S^4, is homeomorphic to S^7 because it has a nondegenerate function with exactly two critical points (maximum and minimum). In order to prove M not diffeomorphic to S^7, Milnor used Hirzebruch's [1956] sophisticated theory and calculation of the index of a manifold in terms of Pontrjagin numbers with Thom's [1954] cobordism theory, both powerful and fundamental tools in the further development of manifold theory.

R. Thom [1958] proposed an obstruction theory concerning the introduction of a differential structure (or smoothing) on a PL-manifold. The obstruction was to be in cohomology groups with coefficients in the group Γ_n of smoothings of the n-sphere with its usual PL-structure. As $\Gamma_n = 0$ for $n < 7$ the first obstruction turned out to be in Γ_7, a cyclic group with 28 elements. A very hard case was $\Gamma_4 = 0$, proved by J. Cerf [1962]. For the groups Γ_n see M. Kervaire and J. Milnor [1963]. The ideas of Thom were made into a solid smoothing theory by J. Munkres [1960, 1964] and much improved by M. Hirsch. (See M. Hirsch and B. Mazur [1974].) M. Kervaire [1960] was the first to produce effectively a PL-manifold (of dimension 10) which could not have the structure of a smooth manifold:

$$\text{C}^1 \to \text{PL is not surjective.}$$

J. Eells and N. Kuiper [1961] and Tamura [1961] gave simple examples in the lowest possible dimension 8. These are manifolds that can be obtained by compactifying $\mathbb{R}^8$ by an S^4, as is the case with the smooth quaternion projective plane. Although the PL-structures of the various exotic n-spheres are all the same, this does not mean that each has the same set of smooth triangulations. N. Kuiper [1965] proved that a smooth triangulation with $n+1$ vertices of an n-sphere exists only for the customary differential structure. A triangulation of an exotic n-sphere requires many more vertices.

We mention as a side remark that the number of vertices e_0 of a triangulation of a closed surface of Euler characteristic χ obeys

$$e_0 \geqslant \text{minimum}\left\{k \in \mathbb{Z}\colon\ 2k \geqslant 7 + \sqrt{49 - 24\chi}\right\}$$

and equality can arise for many surfaces, but not for the Klein-bottle ($\chi = 0,\ e_0 > 7$). Compare Ringel [1974]. For the real projective 3-space a triangulation with 11 vertices exists and this seems to be the minimal number possible. E. Brieskorn [1966] found that a complex algebraic variety with a singularity can have the topology of a manifold in some neighborhood of that singularity. For example, the set

$$\left\{(z_1, \ldots, z_6)\colon\ z_1^2 + z_2^2 + z_3^2 + z_4^3 + z_5^{6k-1} = 0,\ \sum\{z_j \bar{z}_j \leqslant 1\} \subset \mathbb{C}^5\right\}$$

is homeomorphic to an 8-ball, and its boundary is the seven sphere with exotic differential structure $k \cdot \gamma$, exotic if $k \neq 0$, where γ is the generator of Γ^7. So exotic spheres may have rather simple equations. N. Kuiper [1968] used Brieskorn's examples and "generalized" Nash [1952] to obtain algebraic equations for all nonsmoothable PL-8-manifolds. Akbulut [1976] following Tognoli [1973] proved that every PL-8-manifold (as well as some other PL-manifolds of higher dimensions), can be made into a whole algebraic variety and not only a component. Akbulut and Henry King at present are making progress in obtaining algebraic equations for many more PL-manifolds.

Milnor [1961] *disproved* the *Hauptvermutung* for simplicial complexes: the one point compactifications of $L(7,1) \times \mathbb{R}^4$ and $L(7,2) \times \mathbb{R}^4$ (concerning lens spaces $L(7,k)$, see H. Seifert and W. Threlfall [1934]) are homeomorphic without being TRI-equivalent

$$\mathrm{TRI}(S) \to \mathrm{TOP}(S) \text{ is not injective.}$$

The next most important phase in the study of manifolds started with the work of S. Smale [1961] proving the Poincaré conjecture for dimensions $n \geqslant 5$. (For $n = 5$ with the help of J. Stallings and E.C. Zeeman.) If f is a nondegenerate C^∞-function on a compact manifold M, then for increasing values of t, the manifold $\{x\colon\ f(x) \leqslant t\}$ changes at critical values, and these changes can be realized by attaching handles and thickening them. The Morse relations (see Milnor [1963]) among the Betti numbers restrict the possible numbers of nondegenerate critical points of various indices on a given manifold M. Smale succeeded, for a function on a manifold M of the homotopy type of S^n, in cancelling critical points (and handles) until two remained (maximum and minimum). Therefore M is seen to be homeomorphic as well as PL-equivalent to S^n ($n \geqslant 5$): the Poincaré conjecture, as well as the *Hauptvermutung* were proved for S^n, $n \geqslant 5$.

A tremendous activity in manifold theory took place between 1960 and 1970, in which the merging theories for smooth, PL- and topological manifolds developed with new tools like surgery and handlebody theory, h- and s-cobordism theory (see Milnor [1965]), transversality, microbundles and via homotopy theory to algebraic problems, which were particularly deep and hard for non simply-connected manifolds (see C.T.C. Wall [1970]). It will be impossible to go into much detail. I might mention S. Novikov and W. Browder as leaders. Compare the contributions on topology in the proceedings of the International Mathematical Congress in Nice, in particular the paper of L. Siebenmann [1970].

See also the proceedings of *Manifolds Amsterdam* [1970] and R. Kirby and L. Siebenmann [1977]. D. Sullivan [1967] proved the *Hauptvermutung* for simply connected PL-manifolds of dimension $\geqslant 6$, for which $H_3\ (M; \mathbb{Z})$ has no 2-torsion. R. Kirby [1969] made the final break-through by proving that every orientable homeomorphism of S^n onto itself is a product of homeomorphisms, each of which is identical on some open set. This was the crucial and longstanding *stable manifold conjecture*. It carried with it the positive answer to the *annulus conjecture*. R. Kirby and L. Siebenmann [1969] (see Siebenmann [1970]) then solved the triangulation problem and the Hauptvermutung for manifolds of dimension $n \geqslant 5$. They deduced, using in an essential way results on homotopy tori of C.T.C. Wall and others, that there is exactly *one* well defined (by Siebenmann [1970] in a counter-example) obstruction in $H^4(M; \pi_3(\mathrm{TOP/O})) = H^4(M; \mathbb{Z}_2)$ to imposing a PL-structure on a topologically closed n-manifold M^n, $n \geqslant 5$, and, given one PL-structure, the equivalence (isotopy-) classes PL-structures biject onto $H^3(M; \mathbb{Z}_2)$. So *for certain topological manifolds no Brouwer-triangulation exists, and for certain* PL-*manifolds the* PL-*Haupvermutung is false.*

$$\mathrm{PL} \to \mathrm{TOP} \text{ is neither injective nor surjective.}$$

It may be true still, and there is hope for the conjecture, that every topological manifold has some triangulation, which of course cannot always be a Brouwer triangulation (PL). If true then one can hope for algebraic equations as well. R. Edwards [1976] constructed triangulations of S^n, $n \geqslant 5$, with the property that no subdivision is a Brouwer triangulation. So for manifolds:

$$n \geqslant 5,\ \mathrm{PL} \subset \mathrm{TRI} \text{ is not bijective.}$$

He uses B. Mazur [1961] and V. Poenaru [1960], who constructed long ago a contractible 4-manifold M with boundary ∂M that is not simply-connected, although it necessarily has the homology of S^3. Edwards proved, and this is hard, that the $(n-3)$-fold suspension

$$W^n = \Sigma^{n-3}(\partial M) = S^{n-4} * \partial M$$

(obtained by joining every point of S^{n-4} by a line segment to every point of ∂M), for $n \geqslant 5$, is homeomorphic to S^n. If we triangulate W then naturally $S^{n-4} \subset W$ is triangulated by a subcomplex of dimension $n-4$. Every $(n-4)$-simplex in it has a copy of ∂M (and not of S^3) as link, which shows that the triangulation of W is not a Brouwer-triangulation.

In the spirit of our interest in the topology of real algebraic varieties, we mention a real algebraic variety (which is due to C. Gordon), which is a Mazur–Poenaru 3-manifold, that is it bounds a contractible 4-manifold:

$$\partial M = \left\{ (z_1, z_2, z_3) \in \mathbb{C}^3 \colon z_1^2 + z_2^5 + z_3^7 = 0,\ \sum_{j=1}^{3} z_j \bar{z}_j = 1 \right\}.$$

Siebenmann observed, and the reader can check that the double suspension of ∂M is

$$V^5 = \Sigma^2(\partial M) = \left\{ (z_1, z_2, z_3, z_4) \in \mathbb{C}^4 \colon z_1^2 + z_2^5 + z_3^7 = 0,\ \sum_{j=1}^{4} z_j \bar{z}_j = 1 \right\},$$

an algebraic variety, which is homeomorphic to S^5 by Edwards, but whose *natural triangulation is not* PL for the same reasons as above. Observe that the singular curve S^1 with equations $z_1 = z_2 = z_3 = 0$ in the topological 5-sphere V has no normal microbundle (compare P.S. at end of chapter).

Let us recall here that S. Lojaciewicz [1964] was the first to give an accepted proof that *every real algebraic variety can be triangulated.* In order to define uniqueness one first of all has to distinguish certain triangulations of an algebraic variety to be natural, like smooth triangulations for smooth manifolds, and then to show that any two such natural triangulations are TRI-equivalent. Such a kind of uniqueness proof does not exist in the literature for $n \geqslant 3$, although there is some hope that it could be deduced from Lojaciewicz's work. (For $n \leqslant 2$: Papakyriokopoulos [1943].)

It should be noted that the natural singular version of the triangulation conjecture is false: Siebenmann [1970], §3 gave explicit examples of compact locally triangulable spaces that are not triangulable. His example are even locally real algebraic.

Very recently (as I learned from L. Siebenmann and R. Stern) the triangulation problem for topological manifolds has again much advanced. J. Cannon showed, generalising Edwards [1976], that the double suspension $\Sigma^2 W^3$ of *every* homology 3-sphere W^3 is homeomorphic to S^5. With Siebenmann's work this implies that all orientable topological 5-manifolds are triangulable, and there are many of them without any PL-structure.

D. Galewshi and R. Stern, *and independently* T. Matumoto, even define an obstruction element $\tau \in H^5(M; \rho)$ such that the topological manifold M^n, $n \geqslant 5$, is triangulable if and only if $\tau = 0$; and if M is triangulable there are $|H^5(M; \rho)|$ such triangulations up to "concordance".

Unfortunately, although the group ρ is well defined it is also completely unknown. Even so we can conclude that every simply connected topological 6-manifold can be triangulated. It is also known now that necessary and sufficient for triangulability of all manifolds of dimension $\geqslant 5$ is the existence of a smooth closed homology 3-sphere (manifold) with Rohlin invariant 1 (that is, bounding a parallelizable 4-manifold of index 8) such that the connected sum $H \# H$ bounds a homology 4-disc.

Not every compact simplicial complex is homeomorphic to a real algebraic variety. Hardly anything is known about this question. D. Sullivan [1971] discovered that in every triangulation of a real algebraic variety the link of a vertex or simplex has even Euler characteristic. For example, a double cone on the real projective plane cannot be a real algebraic variety. Compact simplicial complexes of dimension one are algebraic if and only if an even number of edges meet at every vertex. Sullivan's condition is perhaps also sufficient to decide which simplicial complexes of dimension two are algebraic. For higher dimensions the problem is completely open.

Smooth as well as PL-manifolds are Lipschitz manifolds: they can be covered with charts for which the transition functions h_{ij} (see above) obey the condition that locally

$$\frac{\|h_{ij}(x) - h_{ij}(y)\|}{\|x - y\|}$$

is bounded away from 0 and from ∞. D. Sullivan [1977] proved that the structure of every topological manifold can be strengthened as much: *every* closed topological *manifold* of dimension $\neq 4$ has a *Lipschitz structure*, and it is unique up to equivalence.

Every PL-structure on S^4 has a unique smoothing and visa versa, but it still remains undecided whether there are more nonequivalent PL-structures on S^4 or closed 4-manifolds in general: The Hauptvermutung and the triangulation conjecture remain open for 4-manifolds. With the Poincaré conjecture for dimensions 3 and 4, the subject of the classification of 3- and 4-manifolds is active, but the main interest in geometry and topology has shifted since 1970 to structures on manifolds like foliations, vector fields, differential equations, Riemannian metrics, functions and maps, their topology and their singularities. The topology of complex algebraic varieties remains very active too.

I mentioned that it was hard for me to do justice to all mathematicians involved in the subject. As it seems appropriate, I will go into some more detail concerning the tremendous development between 1960 and 1970. Several people helped me again to clarify points.

In this decade, 1960 and 1970, the emerging theories of smooth, PL- and topological manifolds were developed using new tools such as surgery and handle body theory (see Milnor [1965]), transversality, microbundles, and block bundles to transfer geometric questions to homotopy theory and to algebraic questions, which were particularly deep for non simply-connected manifolds. Following the work of Kervaire and Milnor mentioned above, the powerful general theory of simply connected manifolds was developed by Browder and Novikov, and the overall non simply-connected theory was put into place by Wall.

In a short space one cannot describe all the outstanding contributions made by the many talented mathematicians who worked in this area. Perhaps the most significant achievement was the resolution of the Hauptvermutung and triangulation problem for manifolds. S.P. Novikov contributed the first striking step when he proved the topological invariance of rational Pontrjagin classes. Together with the surgery exact sequence (the "Sullivan sequence"), this already implied the Hauptvermutung for some special cases. By developing a canonical version of Novikov's argument (with the aid of Siebenmann's thesis) Lashof–Rothenberg and Sullivan were then able to prove the Haupvermutung for 4-connected manifolds of dimension $\geqslant 6$. But, Casson and Sullivan (independently) had developed such penetrating (and *complete* in the case of Sullivan) analyses of the classifying space G/PL that appears in the surgery sequence that they were able to extend the proof to cover all simply connected manifolds for which $H_3(M; \mathbb{Z})$ has no 2-torsion.

The final breakthrough began when R. Kirby showed how to reduce the stable homeomorphisms conjecture to some questions about homotopy tori. This conjecture says that every homeomorphism of $\mathbb{R}^n$ to itself is the product of homeomorphisms, each of which is the identity on some open set, and it also implies the well-known annulus conjecture. But Hsiang–Shaneson and Wall had just classified homotopy tori, and so they could easily resolve the questions of Kirby in the affirmative.

With the same ideas plus topological immersion theory (Lashof and Rothenberg [1968], Lashof [1971]), Kirby and Siebenmann [1969] and Lashof and Rothenberg [1969] solved the triangulation and Hauptvermutung problems for $n \geqslant 5$. Kirby and Siebenmann then deduced, still using the results on homotopy tori in an essential way, that there is exactly one well-defined (by Siebenmann [1970] in a counter-example) obstruction in $H^4(M; \pi_3(\text{TOP/O})) = H^4(M; \mathbb{Z}_2)$ to imposing a PL-structure on a closed topological n-manifold, $n \geqslant 5$ and, given one PL-structure, the equivalence (isotopy) classes of PL structures biject onto $H^3(M; \mathbb{Z}_2)$. *So for certain topological manifolds no Brouwer triangulation exists, and for certain PL-manifolds* (e.g., the torus itself) *the* PL-*Hauptvermutung is false.*

P.S. Here are equations of Siebenmann of a topological 5-manifold that can be triangulated, but has no Brouwer-triangulation or PL-structure:

$$\left\{(z_1, z_2, z_3, z_4):\ z_1^2 + z_2^5 + z_3^7 + e\left(\sum_{j=1}^{3} z_j \bar{z}_j\right)^5 = 0,\ z_4\bar{z}_4 = 1\right\} \subset \mathbb{C}^4.$$

Bibliography

Akbulut, S. (1977), *Algebraic equations for a class of PL-manifolds*, Math. Ann. **231**, 19–31.

Alexander, J.W. (1915), *A proof of the invariance of certain constants in Analysis situs*, Trans. Amer. Math. Soc. **16**, 148–154.

Betti, E. (1871), *Sopra gli spazi di un numero qualunque di dimensioni*, Annali di Matematica **IV** (2) 140–158.

Brieskorn, E. (1966), *Beispiele zur Differentialtopologie von Singularitäten*, Inv. Math. **2**, 1–14.

Brouwer, L.E.J. (1911), *Beweis der Invarianz der Dimensionenzahl*, Math. Ann. **70**, 161–165. *Complete Works II* (1911 C).

Brouwer, L.E.J. (1912), *Über Abbildung von Mannigfaltigkeiten*, Math. Ann. **71**, 97–115, *Complete Works II* (1911 D).

Brouwer, L.E.J. (1939), *Zum Triangulationsproblem*, Proc. Konink. Acad. Wetensch. Amsterdam **41**, 701–706.

Cairns, S.S. (1934), *On the triangulation of regular loci*, Ann of Math. **35**, 579–587.

Cairns, S.S. (1935), *Triangulation of the manifold of class one*, Bull. Amer. Math. Soc. **41**, 549–552.

Cairns, S.S. (1940a), *Triangulated manifolds which are not Brouwer manifolds*, Ann. of Math. **41**, 792–795.

Cairns, S.S. (1940b), *Homeomorphisms between topological manifolds and analytic manifolds*, Ann. of Math. **41**, 796–808.

Cerf, J. (1968a), *La nullité de π_0 (Diff. S^3)*, Séminaire Henri Cartan 1961/63.

Cerf, J. (1968b), *Sur les diffeomorphismes de la sphère de dimension trois* ($\Gamma_4 = 0$), Lecture Notes in Math. vol. 53, Springer, Berlin.

Cohen, M.M. (1968), *A proof of Newman's theorem*, Proc. Cambridge. Phil. Soc. **64**, 961–963.

Dehn, M. and Heegard, P. (1907), *Analysis Situs*, Enz. der Math. Wiss. III **AB 3**, 154–222.

Edwards, R.D. (1976), *The double suspension of a certain homology 3-sphere is S^5*, Unpublished. See a paper of Giffin in Annals of Mathematics, 1978.

Eells, J. and Kuiper, N.H. (1961), *Manifolds which are like projective planes*, Publ. Math. IHES **14**, 5–46.

Eilenberg, S. and Steenrod, N. (1952), *Foundations of Algebraic Topology*, Princeton Univ. Press.

Freudenthal, H. (1939), *Die Triangulation der differenzierbaren Mannigfaltigkeiten*, Proc. Konink. Acad. Wetensch. Amsterdam **42**, 880–901.

Hsiang, W.C. and Shaneson, J. (1970), *Fake tori*, Topology of Manifolds, Proceedings of the 1969 Georgia Conference, Markham Press, Chicago.

Hirsch, M. and Mazur, B. (1974), *Smoothings of piecewise linear manifolds*, Ann. of Math. Studies **80**.

Hirzebruch, F. (1956), *Neue Topologische Methoden in der Algebraischen Geometrie*, Springer, Berlin. (English enlarged edition (1966)).

Hubbard, J. and Chillingworth, D.R.J. (1971), *A note on non-rigid Nash structures*, Bull. Amer. Math. Soc. **77**, 429–431; and with Douady, A., *Cohomology of Nash sheaves*, unpublished.

Hudson, J.F.P. (1969), *Piecewise-Linear Topology*, Benjamin, New York.

Kervaire, M.A. (1960), *A manifold which does not admit any differentiable structure*, Comment. Math. Helvetici **34**, 257–270.

Kervaire, M.A. and Milnor, J. (1963), *Groups of homotopy spheres*, Ann. of Math. **77**, 504–537.

Kirby, R. (1969), *Stable homeomorphisms and the annulus conjecture*, Ann. of Math. **89**, 575–582.

Kirby, R. and Siebenmann, L. (1977), *Foundational essays on topological manifolds, smoothings and triangulations*, Ann. of Math. Studies **88**.

Kirby, R. and Siebenmann, L. (1969), *On the triangulation of manifolds and the Hauptvermutung*, Bull. Amer. Math. Soc. **75**, 742–749.

Kuiper, N.H. (1953), *Locally projective spaces of dimension one*, Mich. Math. J. **2**, 95–97.

Kuiper, N.H. (1965), *On the Smoothings of Triangulated and Combinatorial Manifolds*, *Differential and Combinatorial Topology*, Princeton Univ. Press, 3–22.

Kuiper, N.H. (1968), *Algebraic equations for nonsmoothable* 8*-manifolds*, Publ. Math. IHES **33**, 139–155.

Lashof, R.K. (1969), *Lees' immersion theorem and the triangulation of manifolds*, Bull. Amer. Math. Soc. **75**, 535–538.

Lashof, R. (1971), *The immersion approach to triangulation and smoothing*, Proceedings of Symposia in Pure Math. XXII, Algebraic Topology, Amer. Math. Soc., Providence, RI.

Lashof, R. and Rothenberg, M. (1968), *Hauptvermutung for manifolds*, Proceedings of the Conference on the Topology of Manifolds, Michigan State University (1967), Complementary Series in Math. vol. 13, Prindel, Weber and Schmidt, 81–106.

Lashof, R. and Rothenberg, M. (1969), *Triangulation of Manifolds I, II*, Bull. Amer. Math. Soc. **75**, 750–757.

Lojasiewicz, S. (1964), *Triangulation of semi-analytic sets*, Ann. Scuola Normale Sup. Pisa **III** (18), 449–474.

Mazur, B. (1961), *A note on some contractible* 4*-manifolds*, Ann. of Math. **73**, 221–228.

Milnor, J. (1956), *On manifolds homeomorphic to the* 7*-sphere*, Ann. of Math. **64**, 399–405.

Milnor, J. (1961), *Two complexes with are homeomorphic but combinatorially distinct*, Ann. of Math. **74**, 575–590.

Milnor, J. (1963), *Morse Theory*, Ann. of Math. Study vol. 51, Princeton Univ. Press.

Milnor, J. (1965), *Lectures on the h-cobordism theorem*, Notes Princeton Univ. Press.

Moise, E. (1952), *Affine structures on* 3*-manifolds*, Ann. of Math. **56**, 96–114.

Munkres, J. (1960), *Obstructions to the smoothing of piecewise differentiable homeomorphisms*, Ann. of Math. **72**, 521–554.

Munkres, J. (1964), *Obstructions to imposing differentiable structures*, III, J. of Math. **8**, 361–376.

Nash, J. (1952), *Real algebraic manifolds*, Ann. of Math. **56**, 405–421.

Newman, M.H.A. (1926), *On the foundations of combinatorial analysis situs*, Proc. Konink. Acad. Wetensch. Amsterdam **21**, 611–641.

Papakyriokopoulos, C.D. (1943), *A new proof of the invariance of the homology groups of a complex*, Bull. Soc. Math. Grèce **22**, 1–154.

Poenaru, V. (1960), *Les décompositions de l'hypercube en produit topologique*, Bull. Soc. Math. France **88**, 113–129.

Poincaré, H. (1895), *Analysis Situs*, Journal de L'Ecole Polytechnique 1–121; *Complete Works VI.*

Rado, T. (1925), *Ueber den Begriff der Riemannschen Fläche*, Acta Litt. Scient. Univ. Szegd **2**, 101–121.

Riemann, B. (1854), *Ueber die Hypothesen welche der Geometrie zu Grunde liegen*, Habilitationschrift, Gesammelte Werke (1892), 272–287.

Ringel, G. (1974), *Map Colour Theorem*, Grundlehre der Math. Wiss. vol. 209, Springer, Berlin.

Rourke, C. and Sanderson, B. (1972), *Introduction to Piecewise Linear Topology*, Ergebnisse der Math. vol. 69, Springer, Berlin.

Seifert, H. and Threlfall, W. (1934), *Lehrbuch der Topologie*, Teubner, Leipzig.

Siebenmann, L. (1970), *Topological manifolds*, Proc. ICM Nice **2**, 143–163.

Smale, S. (1961), *Generalized Poincaré's conjecture in dimensions greater than four*, Ann. of Math. **74**, 391–406.

Sullivan, D. (1967), *On the Hauptvermutung for manifolds*, Bull. Amer. Math. Soc. **73**, 598–600.

Sullivan, D. (1971), *Combinatorial invariants of analytic spaces*, Proc. Liverpool Singularities Symposium I, Lecture Notes in Math. vol. 192, Springer, Berlin, 165–168.

Sullivan, D. (1977), *Every manifold has a Lipschitz structure*, to appear.

Tamura, I. (1961), 8*-manifolds admitting no differentiable structure*, J. Math. Soc. Japan **13**, 377–382.

Thom, R. (1954), *Quelques propriétés globales des variétés différentiables*, Comment. Math. Helvetici **28**, 17–86.

Thom, R. (1958), *Des variétés triangulées aux variétés différentiables*, Proc. Int. Congr. Math. Edinburgh, 248–255.

Thom, R. (1971), *Le degré brouwerien en topologie différentielle moderne*, Brouwer Memorial Lecture (1970), Nieuw Archief voor Wiskunde **III** (21), 10–16.

Tognoli, A. (1973), *Su una congettura di Nash*, Ann. Scuola Normale Sup. Pisa **27**, 176–185.

Veblen, O. and Whitehead, J.H.C. (1932), *Foundations of Differential Geometry*, Cambridge Univ. Press.

Wall, C.T.C. (1970), *Surgery on Compact Manifolds*, Academic Press.

Weyl, H. (1913), *Die Idee der Riemannschen Fläche*, Teubner, Leipzig.

Whitehead, J.H.C. (1940), *On* C^1*-complexes*, Ann. of Math. **41**, 809–824.

Zeeman, C. (1963), *Seminar on combinatorial topology*, Notes IHES, Bures-sur-Yvette and University of Warwick, Coventry.

Manifolds Amsterdam (1970), Lecture Notes in Math. vol. 197, Springer, Berlin (1971).

Steenrod, N., *Mathematical Reviews (1940–1967) of Papers in Algebraic and Differential Topology*, Part I.

CHAPTER 17

Graph Theory

Robin J. Wilson

Faculty of Mathematics and Computing, The Open University, Walton Hall, Milton Keynes MK7 6AA, UK
E-mail: r.j.wilson@open.ac.uk

> The origins of graph theory are humble, even frivolous. Whereas many branches of mathematics were motivated by fundamental problems of calculation, motion, and measurement, the problems which led to the development of graph theory were often little more than puzzles, designed to test the ingenuity rather than to stimulate the imagination. But despite the apparent triviality of such puzzles, they captured the interest of mathematicians, with the result that graph theory has become a subject rich in theoretical results of a surprising variety and depth.

So begins the book *Graph Theory 1736–1936* by Biggs, Lloyd and Wilson [1998], which outlines the history of graph theory from Euler's treatment of the Königsberg bridges problem in the 1730s to the explosion of activity in the area in the 20th century. This book contains extracts (translated into English where necessary) from many original writings in the subject, including several discussed below.

This chapter largely follows the account of the above-mentioned book. Also of substantial use has been an extensive unpublished manuscript *Origins of Graph Theory* by P.J. Federico, who spent many years working on the history of graph theory but who died before his book was completed.

1. Traversability

In this subsection we describe Euler's solution of the Königsberg problem and mention some subsequent related work on diagram-tracing puzzles, leading to the concept of an *Eulerian graph*. We also describe some other traversability problems that led to the idea of a *Hamiltonian graph*.

1.1. *Euler and the Königsberg bridges*

On 26 August 1735 Leonhard Euler presented a paper on 'the solution of a problem relating to the geometry of position' to the Academy of Sciences in St. Petersburg, where he had worked since 1727. In his paper, Euler discussed the solution of a problem, which he

HISTORY OF TOPOLOGY
Edited by I.M. James

believed to be widely known, of the *Königsberg bridges problem*, which asks whether it is possible to find a route crossing each of the seven bridges of Königsberg once and once only (Fig. 1); more generally, given any division of a river into branches and any arrangement of bridges, is there a general method for determining whether such a route is possible?

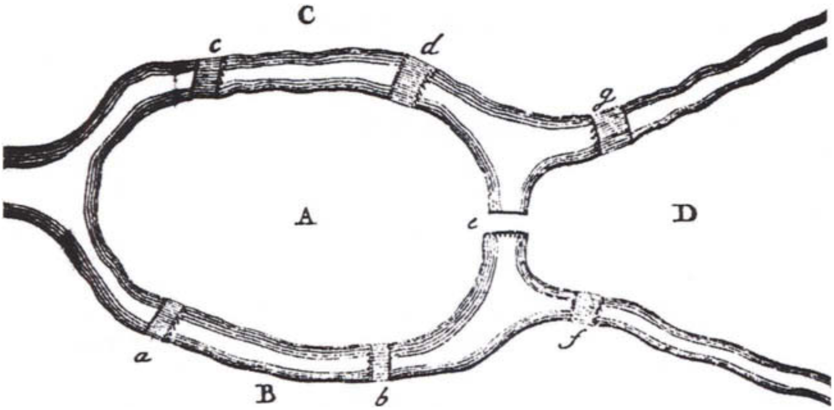


Fig. 1.

In 1736 Euler communicated his solution of the problem to his friend Carl Ehler, Mayor of Danzig, and the Italian mathematician Giovanni Marinoni; an extract of his letter to Marinoni, showing Euler's own drawing of the arrangement of bridges, is shown in Fig. 2 (see Sachs, Stiebitz and Wilson [1988]). Euler also wrote up his solution for publication in the *Commentarii Academii Scientiarum Imperialis Petropolitanae* under the title 'Solutio problematis ad geometriam situs pertinentis'. Although dated 1736, it did not appear until 1741 and was later republished in the new edition of the *Commentarii* (*Nova Acta Commentarii* ...) which appeared in 1752.

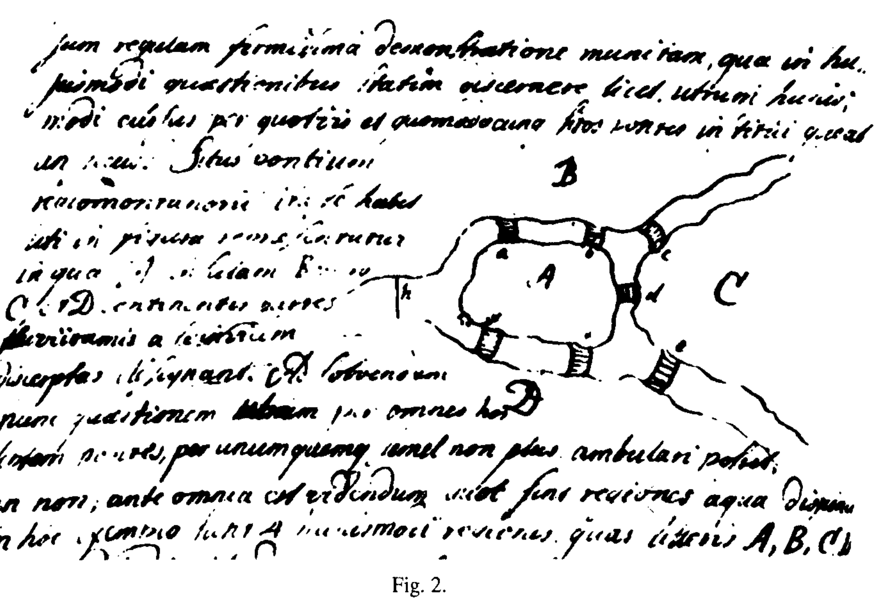


Fig. 2.

Euler's paper is divided into twenty-one numbered paragraphs, of which the first nine show the impossibility of solving the Königsberg bridges problem and the rest are concerned with the general situation. Euler first described the problem as relating to the *geometry of position* (geometria situs), a branch of mathematics first mentioned by Leibnitz and concerned solely with aspects of position rather than the calculation of magnitudes; the various interpretations that have been put on this phrase are discussed in [Pont, 1974]. He then reformulated the problem as one of trying to *find a sequence of eight letters A, B, C or D (the land areas) such that the pairs AB and AC are adjacent twice (corresponding to the two bridges between A and B and between A and C), and the pairs AD, BD and CD are adjacent just once* and showed that this is impossible.

In discussing the general problem, Euler observed that

> 'the number of bridges written next to the letters A, B, C, etc. together add up to the twice the total number of bridges. The reason for this is that, in the calculation where every bridge leading to a given area is counted, each bridge is counted twice, once for each of the two areas that it joins.'

This is the earliest statement of what graph theorists now call the *handshaking lemma*. The paper continues with Euler's main conclusions:

> 'If there are more than two areas to which an odd number of bridges lead, then such a journey is impossible. If, however, the number of bridges is odd for exactly two areas, then the journey is possible if it starts in either of these two areas. If, finally, there are no areas to which an odd number of bridges lead, then the required journey can be accomplished starting from any area.'

Finally, Euler noted the converse result, that if the above conditions hold, then a route is possible, and gave a heuristic reason why this should be so. However, his discussion does not amount to a proof, presumably because he considered the result self-evident, and a valid demonstration did not appear until a related result was proved by C. Hierholzer in 1873. Hierholzer's discussion was given in the language of diagram tracing, to which we now turn.

1.2. *Diagram-tracing puzzles*

In 1809 the French mathematician Louis Poinsot wrote a memoir on polygons and polyhedra (Poinsot [1809–10]), in which he described the four non-convex regular polyhedra and posed several geometrical problems, including the following:

> 'Given some points situated at random in space, it is required to arrange a single flexible thread uniting them two by two in all possible ways, so that finally the two ends of the thread join up, and so that the total length is equal to the sum of all the mutual distances.'

For example, we can arrange a thread joining seven points in the order

0–1–2–3–4–5–6–0–2–4–6–1–3–5–0–3–6–2–5–1–4–0

(see Fig. 3). In fact, there are many millions of such arrangements, as was subsequently observed by M. Reiss [1871–73] in the context of determining the number of ways that one can lay out a ring of dominoes; the above ordering corresponds to the ring of dominoes

0–1, 1–2, 2–3, . . . , 1–4, 4–0. Poinsot noted that a solution is possible only for an odd number of points, and gave an ingenious method for joining the points in each such case.

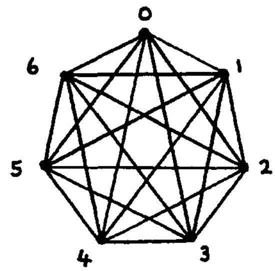


Fig. 3.

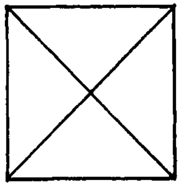


Fig. 4.

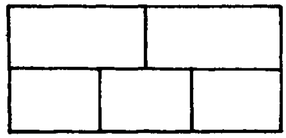


Fig. 5.

Puzzles that require one to draw a given diagram in the smallest possible number of connected strokes have been of interest for many hundreds of years; see, for example, the early African examples in [Ascher, 1991]. In particular, as Poinsot observed, the diagram in Fig. 4 has four points at which three adjacent lines meet and so cannot be drawn with fewer than two strokes; similarly, four separate strokes are needed to trace all the edges of a cube. A few years later, T. Clausen (1844) observed that four strokes are needed to draw the diagram in Fig. 5.

In 1847 Johann Benedict Listing wrote a short treatise entitled *Vorstudien zur Topologie*, in which he investigated a number of non-metrical geometrical problems and discussed the solution of diagram-tracing puzzles; these included Clausen's example and the complicated diagram in Fig. 6 which can be drawn in a single stroke. His treatise is noteworthy for being the first place that the word 'topology' had appeared in print; Listing had coined the word in 1836 in a letter to a friend.

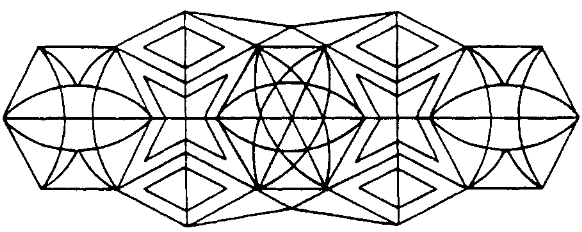


Fig. 6.

As mentioned above, C. Hierholzer [1873] was the first to give a complete account of the theory of diagram-tracing puzzles, proving in particular that

> 'If a line-system can be traversed in one path without any section of line being traversed more than once, then the number of odd nodes is either zero or two'

and, conversely, that

> 'if a connected line-system has either no odd node or two odd nodes, then the system can be traversed in one path.'

The precise connection between Euler's bridge-crossing problems and the tracing of diagrams was not noticed until the end of the 19th century. Euler's discussion of such problems had been popularized through a French translation of E. Coupy [1851] that included an application to the bridges over the River Seine, and by a lengthy account in Volume 1 of E. Lucas' *Récréations Mathématiques* [Lucas, 1882], but it was W.W. Rouse Ball [1892] who first represented the four land areas by points and the bridges by lines joining the appropriate pairs of points, thereby producing the well-known diagram in Fig. 7.

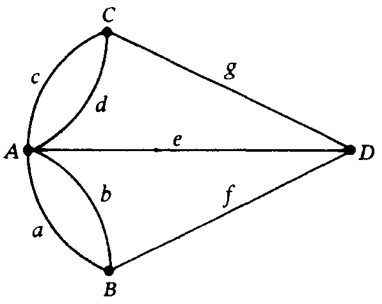


Fig. 7.

Such a diagram is now called a **connected graph**, the points are **vertices**, the lines are **edges**, and the number of edges appearing at a vertex is the **degree** of that vertex; thus, the above graph has three vertices of degree 3 and one vertex of degree 5. It follows from the above results that *a connected graph has a path that includes each edge just once if and only if there are exactly* 0 *or* 2 *vertices of odd degree*. When there are no vertices of odd degree, the graph is called an **Eulerian graph**, even though the concept of such a graph did not make its first appearance until over 150 years after the paper by Euler that originally inspired it.

1.3. *Hamiltonian graphs*

A type of graph problem that is superficially similar to the Eulerian problems described above is that of *finding a cycle that passes just once through each vertex*, rather than just once along each edge; for example, if we are given the graph in Fig. 8, then it is impossible to cover each *edge* just once, because there are eight vertices of degree 3, but we can find a cycle (shown with heavy lines) passing through each *vertex* just once. Such graphs are now called **Hamiltonian graphs** although, as we shall see, this is perhaps not the most appropriate name for them.

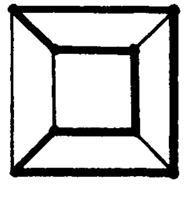


Fig. 8.

An early example of a Hamiltonian-type problem is the celebrated *knight's-tour problem*. The problem is to find a succession of knight's moves on a chessboard visiting each of the 64 squares just once and returning to the starting point. The connection with Hamiltonian graphs may be seen by regarding the squares as vertices of a graph, and joining two squares whenever they are connected by a single knight's move.

Solutions of the knight's-tour problem have been known for many hundreds of years, including solutions by De Montmort and De Moivre in the 17th century, but it was not until the mid-18th century that it was subjected to systematic mathematical analysis, by Leonhard Euler [1759]; Euler showed in particular that no solution is possible for the analogous problem on a chessboard with an odd number of squares. Shortly afterwards, A.-T. Vandermonde [1771] discussed the original problem, obtaining the knight's tour in Fig. 9 and remarking that

> 'whereas that great geometer presupposes that one has a chessboard to hand, I have reduced the problem to simple arithmetic.'

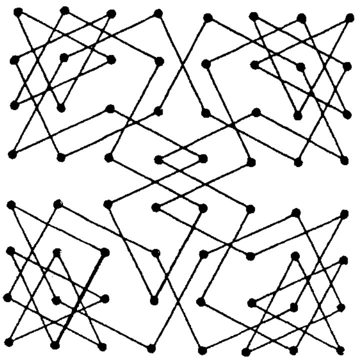


Fig. 9.

50	11	24	63	14	37	26	35
23	62	51	12	25	34	15	38
10	49	64	21	40	13	36	27
61	22	9	52	33	28	39	16
48	7	60	1	20	41	54	29
59	4	45	8	53	32	17	42
6	47	2	57	44	19	30	55
3	58	5	46	31	56	43	18

Fig. 10.

Many mathematicians have since attempted to generalize the problem to other types of board, or to find solutions that satisfy extra conditions; for example, Jaenisch [1862–63] wrote a 3-volume account of the knight's tour problem, and included the ingenious solution in Fig. 10 where the successive knight's moves yield a semi-magic square in which the entries in each row or column add up to 260.

In 1855 the Royal Society of London received a paper by the Revd. Thomas Penyngton Kirkman that asked *for which polyhedra can one find a cycle passing through all the vertices just once?*; for example, a cube has the cycle given (in flattened form) in Fig. 8. Kirkman claimed to have a sufficient condition for the existence of such a cycle, but his reasoning was faulty; however, he did prove that any polyhedron with even-sided faces and an odd number of vertices has no such cycle, and gave as an example the polyhedron obtained by 'cutting in two the cell of a bee' (see Fig. 11). In 1884 P.G. Tait asserted that every 3-valent polyhedron has a cycle passing through every vertex; if true, this assertion would have yielded a simple proof of the four colour theorem (see Section 4), but it was eventually disproved by W.T. Tutte [1946], who produced the 3-valent polyhedron in Fig. 12.

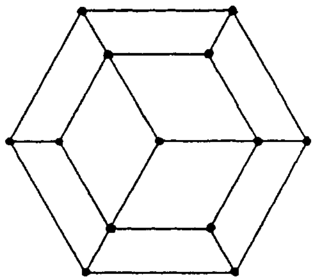


Fig. 11.

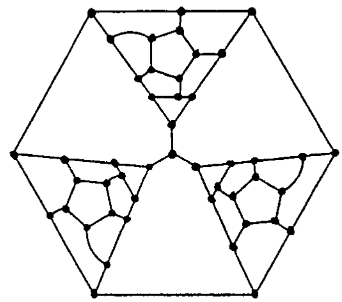


Fig. 12.

Another mathematician who was intrigued with cycles on polyhedra was Sir William Rowan Hamilton. Arising from his work on quaternions and non-commutative algebra, Hamilton was led to the *icosian calculus*, in which he considered cycles of faces on an icosahedron or, equivalently, cycles of vertices on a dodecahedron; such a cycle is given in Fig. 13. Hamiltonian subsequently invented the *icosian game*, a solid or flat dodecahedron with holes at the vertices and pegs to be inserted along paths and cycles according to certain instructions that he had written (see Fig. 14). He sold the game to a games manufacturer for £25 who marketed it under the name *A voyage round the world*, with the vertices $B, C, D, \ldots, Z$, standing for Brussels, Canton, ... , Zanzibar; it was not a commercial success.

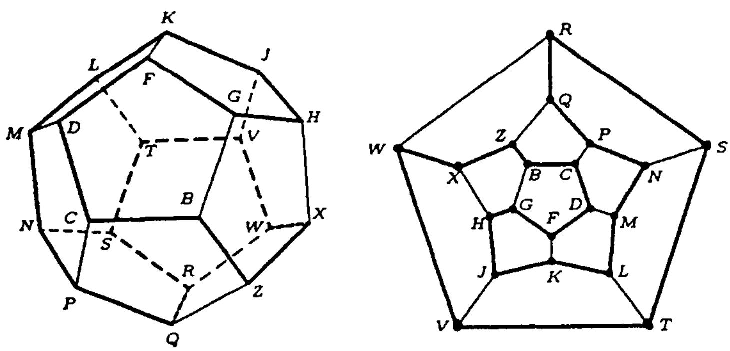


Fig. 13.

Because of Hamilton's influence, his name has become associated with such cycle problems and with the corresponding Hamiltonian graphs, even though Kirkman, who considered these problems in greater generality, had preceded him by a few months. Unlike the Eulerian problem, no necessary and sufficient condition has been discovered for the existence of a Hamiltonian cycle in a general graph, although a number of sufficient conditions have been found – most notably by G.A. Dirac [1952] and O. Ore [1960]. A survey of results on Hamiltonian graphs appears in a survey by J.-C. Bermond [1978].

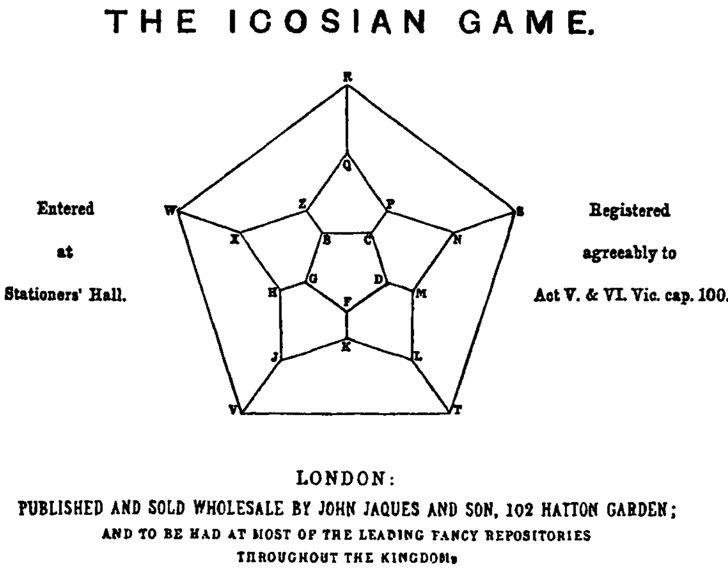


Fig. 14.

2. Trees

A **tree** is a connected graph with no cycles; for example, Fig. 15 illustrates the six possible trees with six vertices.

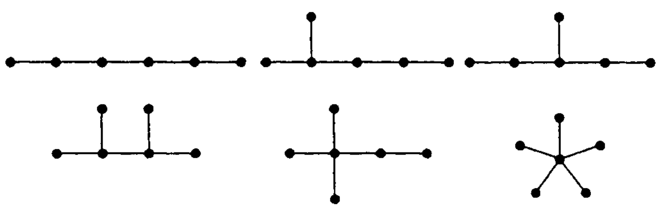


Fig. 15.

The concept of a tree appeared implicitly in the work of Gustav Kirchhoff [1847], who used graph-theoretical ideas in the calculation of currents in an electrical network. In this section, however, our concern is mainly with the enumeration of certain types of chemical molecule. Such problems can be reduced to the counting of trees and were investigated by Arthur Cayley and James Joseph Sylvester. We outline their contributions and indicate how their ideas were developed in the first half of the 20th century.

2.1. *Chemical trees*

By 1850 it was already known that chemical elements combine in fixed proportions, and chemical formulae such as H_2O (water) and C_2H_5OH (ethanol) were well established. But it was not understood exactly how the various elements combine to form these substances. The breakthrough occurred in the 1850s when August Kekulé (Germany), Edward Frankland (England), A.M. Butlerov (Russia) and A.S. Couper (Scotland) proposed what is now the theory of *valency* (see the book by Russell [1971]); in this theory, each atom has several bonds by which it is linked to other atoms: *carbon atoms have four bonds, oxygen atoms have two, and hydrogen atoms have one.*

As the idea of valency became established, it became increasingly necessary for chemists to find a method for representing molecules diagrammatically. Various people tried and failed, including some of those mentioned above, but it was not until the 1860s that Alexander Crum Brown [1864] proposed what is essentially the form we use today. In his system, each atom is represented by a circled letter, and the bonds are indicated by lines joining the circles. Fig. 16 shows:

(a) Crum Brown's representation of ethanol,
(b) the present-day representation with the circles omitted,
(c) the associated chemical tree with vertices corresponding to atoms and edges representing bonds.

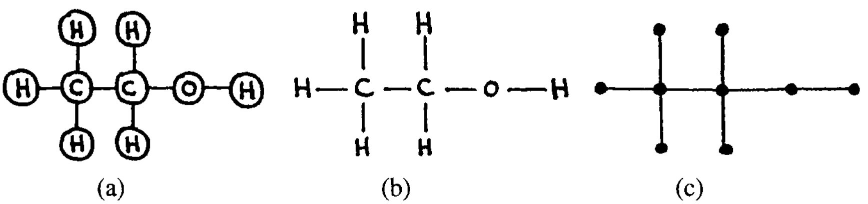


Fig. 16.

Crum Brown's 'graphic notation', as it came to be called, was quickly adopted by Frankland [1866], who used it in his *Lecture Notes for Chemical Students*. Its great advantage was that its use explained, for the first time, the phenomenon of *isomerism*, whereby there can exist pairs of molecules (*isomers*) with the same chemical formula but different chemical properties. Fig. 17 shows a pair of isomers, each with chemical formula C_4H_{10}; the difference between them is that the atoms are arranged in different ways inside the molecule.

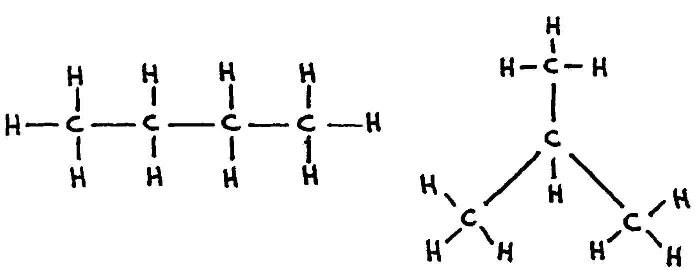


Fig. 17.

This idea leads naturally to problems of *isomer enumeration*, in which we determine the number of different molecules with a given chemical formula. The most celebrated of these problems is that of enumerating the alkanes (paraffins), with formula C_nH_{2n+2}; the following table gives the numbers of such molecules for $n = 1, 2, \ldots, 8$.

Formula	CH_4	C_2H_6	C_3H_8	C_4H_{10}	C_5H_{12}	C_6H_{14}	C_7H_{16}	C_8H_{18}
Number	1	1	1	2	3	5	9	18

2.2. *Counting trees*

In 1874, Arthur Cayley observed that the diagrams corresponding to the alkanes all have a tree-like structure, and that removing the hydrogen atoms yields a tree in which each vertex has degree 1, 2, 3 or 4 (see Fig. 18); thus, the problem of enumerating such isomers is the same as counting trees with this property.

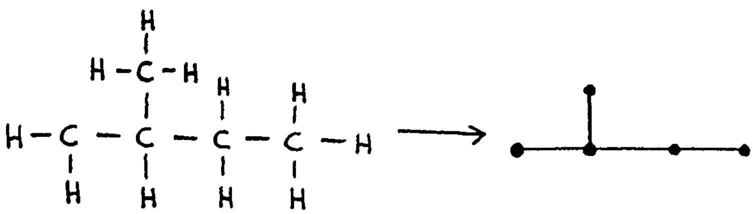


Fig. 18.

Cayley had been interested in tree-counting problems for some time. In 1857, while trying to solve a problem inspired by Sylvester relating to the differential calculus, he managed to enumerate all *rooted trees* – that is, trees in which one particular vertex has been singled out as the 'root' of the tree (see Fig. 19).

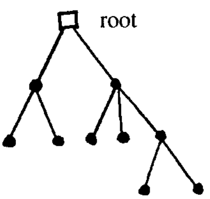


Fig. 19.

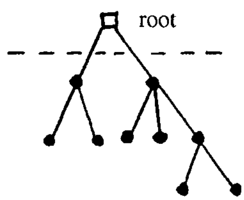


Fig. 20.

Cayley's method was to take a rooted tree and remove its root, thereby obtaining a number of smaller rooted trees, as in Fig. 20. If A_n is the number of rooted trees with n branches, then this reduction enables one to express A_n in terms of certain of the numbers A_k, where k is less than n. Specifically, Cayley considered the generating function

$$1 + A_1x + A_2x^2 + A_3x^3 + \cdots$$

and proved that it is equal to the product

$$(1-x)^{-1} \cdot (1-x^2)^{-A_1} \cdot (1-x^3)^{-A_2} \cdot \cdots .$$

Using this equality he was then able to calculate the coefficients $A_1, A_2, A_3, \ldots$ one at a time.

It was not until several years later, in 1874, that Cayley found a systematic method for counting *unrooted* trees – a much more difficult problem. He applied this method to the enumeration of various isomers, building up the molecules step by step from their 'centres', and succeeded in finding the numbers of alkanes with up to 11 vertices. However, his methods were cumbersome and impractical, in spite of improvements suggested independently by Sylvester and Camille Jordan involving a redefinition of the 'centre' of a tree, and it was not for many years that any substantial progress was made on the counting of chemical molecules.

At around the same time that Cayley was enumerating isomers, his friends James Joseph Sylvester and William Kingdon Clifford were trying to establish a link between the study of chemical molecules and the algebraic topic of *invariant theory*. Each chemical atom was to be compared with a 'binary quantic', a homogeneous expression in two variables such as

$$ax^3 + 3bx^2y + 3cxy^2 + dy^3,$$

and a chemical substance composed of atoms of various valencies was to be compared with an 'invariant' of a system of binary quantics of the corresponding degrees. Indeed, profoundly influenced by Frankland's *Lecture Notes*, Sylvester was later to write [1878]:

> 'The more I study Dr. Frankland's wonderfully beautiful little treatise the more deeply I become impressed with the harmony or homology . . . which exists between the chemical and algebraical theories. In travelling my eye up and down the illustrated pages of "the Notes", I feel as Aladdin must have done in walking in the garden where every tree was laden with precious stones . . .'.

Both Cayley and Sylvester had made important contributions to the theory of invariants, and Sylvester and Clifford tried to introduce the 'graphic notation' of chemistry into the subject (see Fig. 21); indeed, the use of the word *graph* for such a diagram arose from one of Sylvester's papers [1877–78] in this area.

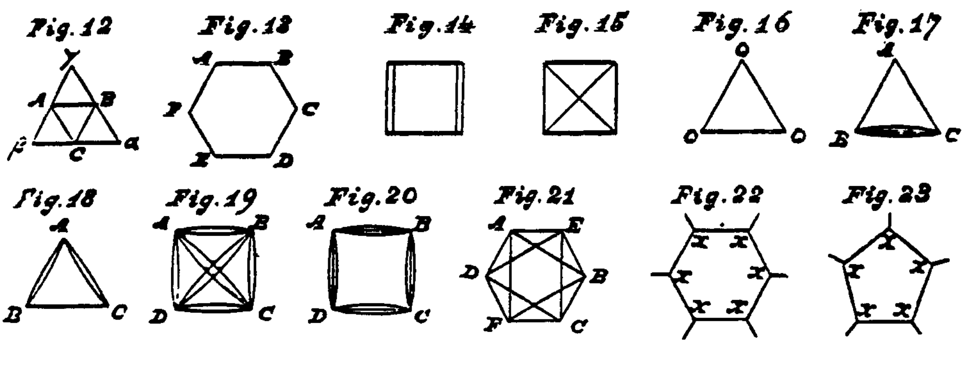


Fig. 21.

Unfortunately, the 'chemico-algebraic' ideas of Sylvester and Clifford proved to be less useful than their originators had hoped – two largely unrelated ideas linked by a notation that was only superficially similar. Invariant theory quickly became submerged in the work of David Hilbert and others, while the theory of graphs increasingly took on a life of its own. Perhaps Sylvester feared this all along; in a letter to Simon Newcomb, he nervously admitted that:

> 'I feel anxious as to how it will be received as it will be thought by many strained and over-fanciful. It is more a 'reverie' than a regular mathematical paper ... [Nevertheless,] it may at the worst serve to suggest to chemists and Algebraists that they may have something to learn from each other.'

In 1889, Cayley tackled another tree-counting problem – that of determining the number $t(n)$ of labelled trees with n vertices; for example, if $n = 4$, the number of such trees is 16 (see Fig. 22). Unlike the earlier problems we considered, this one has a very simple answer – namely, $t(n) = n^{n-2}$. Cayley stated this result and demonstrated it for $n = 6$, but Prüfer [1918] was the first to publish a complete proof.

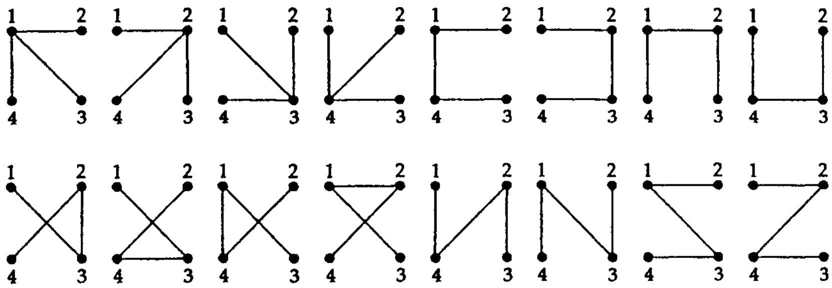


Fig. 22.

It was not until the 1920s and 1930s that any substantial theoretical progress was made in the counting of chemical molecules. In 1927, J.H. Redfield produced a paper that foreshadowed the later work of George Pólya, but it was written in obscure language and overlooked for many years. Shortly afterwards, A.C. Lunn and J.K. Senior [1929] recognized that the theory of permutation groups was appropriate to the enumeration of isomers, and their ideas were considerably developed in a fundamental paper of Pólya [1937] in which the classical method of generating functions was combined with the idea of a permutation group. Pólya's main result was a powerful theorem that enables one to enumerate certain types of configuration under the action of a group of symmetries; his results have been used to enumerate both graphs and molecules and many other configurations arising in mathematics. An English translation of, and commentary on, Pólya's paper appears in the book by Pólya and Read [1987].

3. Topological graph theory

In this section we investigate the origins of *Euler's polyhedron formula* for both polyhedra and planar graphs, and show how one of its generalizations led to the work of Listing and, ultimately, Poincaré. We also study the structure of graphs embedded on the plane or sphere, and describe some work on the embedding of non-planar graphs on surfaces other than the sphere.

3.1. *Euler's polyhedron formula*

Although the Greeks were familiar with the five regular solids (the tetrahedron, cube, octahedron, dodecahedron and icosahedron) and several other polyhedra, there is no evidence that they knew the simple formula relating the numbers of vertices, edges and faces of such a polyhedron – namely,

$$\text{(number of vertices)} - \text{(number of edges)} + \text{(number of faces)} = 2.$$

In the 17th century, René Descartes also missed the formula. He obtained a formula for the sum of the angles in all the faces of a polyhedron, from which the above formula can be deduces, but he never made the deduction.

It was Leonhard Euler [1750] who first stated the above result, in a letter to Christian Goldbach. Euler considered a solid polyhedron and obtained various equalities and inequalities relating the numbers of faces, solid angles (vertices) and joints where two faces come together (edges) to other quantities; in particular, denoting them by H, S and A, respectively, he asserted that

> '6. In every solid enclosed by plane faces the aggregate of the number of faces and the number of solid angles exceeds by two the number of edges, or $H + S = A + 2$ ...
> 11. The sum of all plane angles is equal to four times as many right angles as there are solid angles, less eight, that is $= 4S - 8$ right angles ...
> I find it surprising that these general results in solid geometry have not previously been noticed by anyone, so far as I am aware; and furthermore, that the important ones, Theorems 6 and 11, are so difficult that I have not yet been able to prove them in a satisfactory way.'

Euler verified these results for several families of polyhedra and two years later produced a dissection proof, but his proof was deficient. The first valid proof was a metrical one given by A.M. Legendre [1794].

Euler's formula also holds for any **planar graph** – the map obtained by stereographically projecting the polyhedron onto the plane (see Fig. 23) – provided that we remember

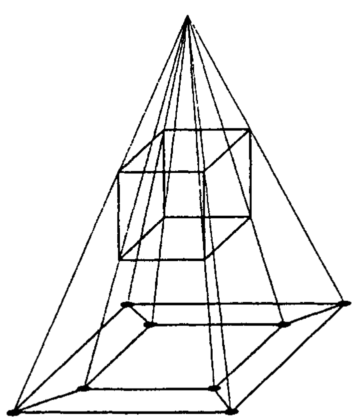


Fig. 23.

to include the 'infinite' (unbounded) face. In 1813, Augustin-Louis Cauchy used a triangulation process to give topological proofs of both versions of Euler's formula, and deduced that there are only four regular non-convex polyhedra, as Poinsot had predicted.

Around the same time, Simon-Antoine-Jean Lhuilier [1811] gave a topological proof that there are only five regular convex polyhedra and anticipated the idea of duality by remarking that four of them occur in reciprocal pairs; he also found three types of polyhedra for which Euler's formula fails – those with an interior cavity, those with indentations in their faces, and ring-shaped polyhedra drawn on a torus (that is, polyhedra containing a 'tunnel'). For ring-shaped polyhedra, he obtained the formula

$$(\text{number of vertices}) - (\text{number of edges}) + (\text{number of faces}) = 0.$$

Lhuilier then extended this discussion to prove that if g is the number of 'tunnels' in a surface on which a polyhedral map is drawn, then

$$(\text{number of vertices}) - (\text{number of edges}) + (\text{number of faces}) = 2 - 2g.$$

The number g is now called the **genus** of the surface, and the quantity $2 - 2g$ is its **Euler characteristic**; these numbers depend only on the surface on which the polyhedron is embedded, and not on the map itself.

Lhuilier's result was the starting point for an extensive investigation by Listing [1861–62], entitled *Der Census räumliche Complexe*, which proved to be influential in the subsequent development of topology; these 'complexes' are built up from simpler pieces, and Listing studied the question of how their topological properties affect the above generalization of Euler's formula.

Listing's ideas were soon taken up by other mathematicians. In particular, Henri Poincaré developed them in his papers of 1895 to 1904 that laid the foundations of algebraic topology. Like Listing, Poincaré developed a method for constructing complexes from basic 'cells', such as 0-*cells* (vertices) and 1-*cells* (edges). In order to fit the cells together, he adapted a technique of Kirchhoff from the theory of electrical networks, replacing sets of linear equations by matrices. These matrices could then be studied from an algebraic point of view.

Poincaré's work was an instant success, and appeared in M. Dehn and P. Heegaard's article [1907] on *analysis situs* (topology) in the *Encyklopädie der Mathematischen Wissenschaften*. His ideas were subsequently developed further by Oswald Veblen in a series of colloquium lectures for the American Mathematical Society on analysis situs; these lectures were delivered in 1916 and published in book form six years later.

3.2. *Planar graphs*

Like many other aspects of graph theory, the origins of the study of planar graphs can be found in recreational puzzles. One such puzzle was given by August Möbius in his lectures around the year 1840:

> 'There was once a king with five sons. In his will he stated that after his death the sons should divide the kingdom into five regions so that the boundary of each region should

> have a frontier line in common with each of the other four regions. Can the terms of the will be satisfied?'

This question asks whether it is possible to find five mutually neighbouring regions in the plane. We can turn this into a graph theory problem by 'dualizing' it, replacing regions by capital cities and frontier lines by connecting roads, as follows:

> 'There was once a king with five sons. In his will he stated that after his death the sons should join the five capital cities of his kingdom by roads so that no two roads intersect. Can the terms of the will be satisfied?'

Note that if there had been only four sons, then both problems would have been easily solved; Fig. 24 gives the solutions and the dual connection between them. Note that each region on the left corresponds to a vertex on the right, each vertex on the left corresponds to a region on the right, and there is a one-one correspondence between the edges on the left and those on the right.

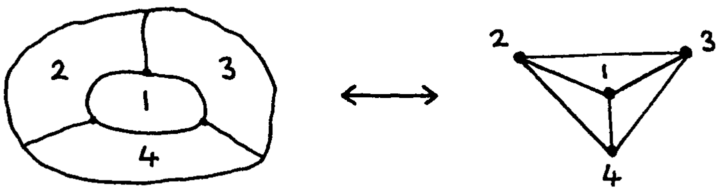
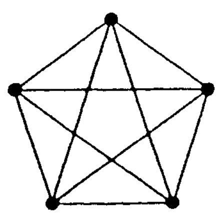


Fig. 24.

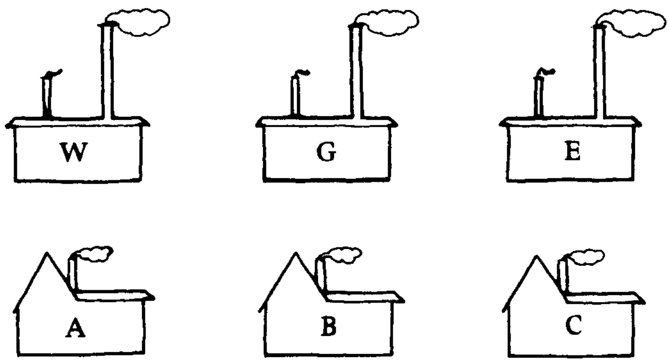
A little experimentation, or use of Euler's formula, shows that Möbius' original problem is insoluble. If we define the **complete graph** K_n to be the graph obtained by drawing edges connecting n vertices in all possible ways, then our result is that *the graph* K_5 *is non-planar* (see Fig. 25).



K_5

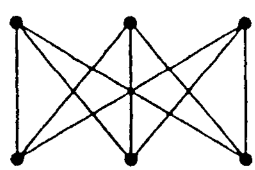
Fig. 25.

A related problem is the *'gas–water–electricity' problem.* The origins of this problem are obscure, but in 1913 H.E. Dudeney presented the problem as follows, describing it as 'as old as the hills':



'The puzzle is to lay on water, gas, and electricity, from W, G, and E, to each of the three houses, A, B, and C, without any pipe crossing another. Take your pencil and draw lines showing how this should be done. You will soon find yourself in difficulties . . .'.

This problem is also impossible, although Dudeney claimed to have solved it by running a pipe through one of the houses. If we define the **complete bipartite graph** $K_{r,s}$ to be the graph obtained by connecting each of r independent vertices to each of s independent vertices in all possible ways, then our result is that *the graph* $K_{3,3}$ *is non-planar* (see Fig. 26).



$K_{3,3}$

Fig. 26.

In 1929, Kazimierz Kuratowski proved the surprising result that *the graphs in Figs. 25 and 26 are the 'basic' non-planar graphs*, in the sense that every non-planar graph must contain a subdivision of at least one of them (see Kuratowski [1930]); this result was obtained independently by O. Frink and P.A. Smith.

For some time mathematicians had tried to find characterizations of planar graphs that depended on combinatorial considerations rather than topological ones. The clue to solving this turned out to be through duality; note that only planar graphs have (geometrical) duals. In 1931, Hassler Whitney formulated an abstract definition of duality that is purely combinatorial, involving the cycles and cutsets of two graphs, and that agrees with the geometrical definition of dual graph when the graph is planar. He then proved that, with this abstract form of dual, *a graph is planar if and only if it has an abstract dual.* Extending these ideas led him eventually to the concept of a matroid, which generalizes ideas of independence in both graphs and vector spaces (Whitney [1935]); in particular, the dual of a matroid is a natural concept that extends and clarifies the duality of planar graphs. Interest in matroids took time to develop, but in the 1950s Tutte obtained a Kuratowski-type condition for a matroid to arise from a graph (Tutte [1959]).

In recent years, much attention has been paid to extending these results to graphs embedded on surfaces other than the plane, or equivalently the sphere. We have seen that Euler's polyhedron formula extends to such graphs, but the problem of determining whether there is a set of 'forbidden subgraphs' for non-planar graphs, analogous to the graphs K_5 and $K_{3,3}$ for the sphere, remained elusive for a long time. We say that a graph has **genus** g if it can be embedded on a surface of a sphere with g handles but not on the surface of a sphere with fewer handles; for example, the complete graph K_5 is of genus 1 since it can be embedded on a torus but not on a sphere (see Fig. 27). In a remarkable series of papers in the 1980s, Neil Robertson and Paul Seymour proved that for each genus g, there is a *finite* set of forbidden subgraphs; however, for $g > 0$, the number of forbidden subgraphs may be large – even for $g = 1$ there are over a hundred of them. For non-orientable surfaces there is a similar result, and H.H. Glover, J.P. Huneke and C.S. Wang [1979] obtained a set of 103 forbidden subgraphs for the projective plane. A survey of this area can be found in Robertson and Seymour [1985].

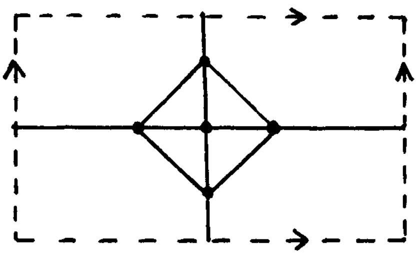


Fig. 27.

A natural question is to ask for the genus of certain important families of graphs. For the complete graphs K_n, the answers came after a long and difficult struggle involving many people. In 1968 Gerhard Ringel and Ted Youngs proved that the genus of K_n is

$$\lceil (n-3)(n-4)/12 \rceil$$

(see the book by Ringel [1974]). This result is closely related to the Heawood conjecture discussed in the next section. Their proof involved the ingenious use of a related 'electrical current graph', and split up into no fewer than twelve separate arguments, depending on the residue class of n modulo 12. Some of these cases were particularly intransigent, and three values of n that did not fit into the general pattern proved to be too difficult for mathematicians and were eventually sorted out by Jean Mayer [1969], a professor of French literature!

4. Graph colouring

In this section we describe the origins of the celebrated *four colour problem* and outline its solution. We also indicate other types of graph colouring problem.

4.1. *Map colouring*

The earliest known reference to the four colour problem occurs in a letter dated 23 October 1852, from Augustus De Morgan to Sir William Rowan Hamilton. In this letter, De Morgan described how one of his students had asked him whether every map can be coloured with only four colours. The student was later identified as Frederick Guthrie, who claimed that the problem was due to his brother Francis; Francis Guthrie had formulated it while colouring the counties of a map of England.

In his letter, De Morgan observed that four colours are needed for some maps; for example, if there are four neighbouring countries, then each country must be differently coloured from its neighbours. But four colours may be needed even if four neighbouring countries do not appear. An example is given in Fig. 28.

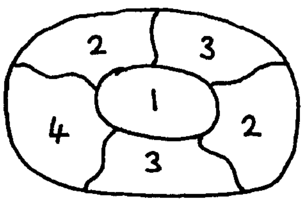


Fig. 28.

De Morgan quickly became intrigued by the problem and communicated it to several other mathematicians, so that it soon became part of mathematical folklore. In 1860 he stated it, in rather obscure terms, in an unsigned book review in the *Athenaeum*, a scientific and literary journal; it is likely that this is the first printed reference to the problem (De Morgan [1860]). This review was read in the USA by the logician and philosopher C.S. Peirce, who subsequently presented an attempted proof to a mathematical society at Harvard University.

It was not until after De Morgan's death in 1871 that any progress was made in solving the four colour problem. On 13 June 1878, at a meeting of the London Mathematical Society, Arthur Cayley enquired whether the problem had been solved, and soon afterwards he wrote a short paper (Cayley [1879]) for the Royal Geographical Society in which he attempted to explain in simple terms where the difficulties lie. He also proved that one can make the simplifying assumption that exactly three countries meet at each point.

In 1879 there appeared one of the most famous fallacious proofs in mathematics. Its author was Alfred Bray Kempe, a London barrister who had studied with Cayley at Cambridge and had become well known for his work on linkages. On learning of this proof, Cayley suggested that Kempe submit it to the *American Journal of Mathematics*, newly founded and edited by J.J. Sylvester (Kempe [1879]).

Although Kempe's argument contained a fatal flaw, it also included some important ideas that were to feature in many subsequent attempts on the problem. His proof was in two parts. He first showed, using Euler's polyhedron formula, that every map necessarily contains a digon, triangle, quadrilateral or pentagon (see Fig. 29); since at least one of these configurations must appear, we call such a set of configurations an **unavoidable set**.

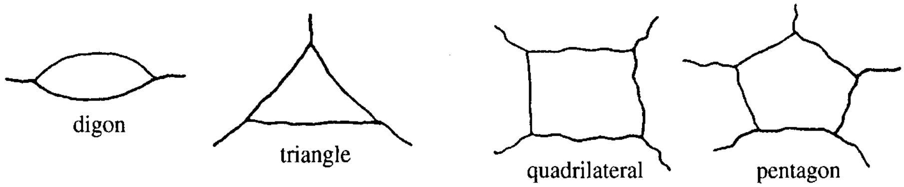


Fig. 29.

Kempe then took each of these configurations in turn and showed that any colouring containing it can be extended to the whole map; a configuration for which this is true is called a **reducible configuration**. Now, it is simple to prove that the digon and triangle are reducible configurations. To prove that a quadrilateral is reducible, Kempe looked at a two-coloured piece of the map – for example, the part of the map containing countries coloured red and green – and he was able to interchange the colours so as to enable the colouring to be extended to the whole map as required. To prove that a pentagon is reducible, Kempe simply repeated the process, making *two* colour interchanges simultaneously. Since all possible cases have been considered, the proof is complete.

Kempe's proof was greeted with enthusiasm, and he published two further papers indicating various simplifications. In 1880 the natural philosopher P.G. Tait reformulated the result in terms of the colouring of boundary *edges* (rather than countries), believing that such considerations would simplify the proof still further (Tait [1878–80]). The headmaster of a famous school set the problem as a challenge problem to his pupils, Frederick Temple (Bishop of London, later Archbishop of Canterbury) produced a 'proof' during a lengthy meeting, and Lewis Carroll reformulated the problem as a game between two players.

In 1890, Percy Heawood, who had learned of the problem while a student at Oxford University, published a paper that pointed out Kempe's error (Heawood [1890]). In this paper Heawood gave a specific example (see Fig. 30) to show that, whereas one colour inter-

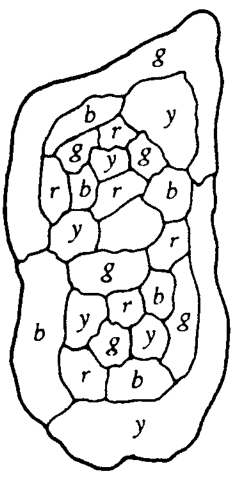


Fig. 30.

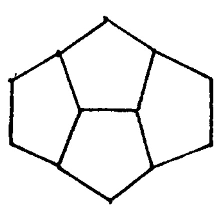


Fig. 31.

change is always permissible, one cannot carry out two interchanges at the same time; thus, Kempe's treatment of the pentagon was deficient. Heawood managed to salvage enough from Kempe's proof to prove that *every map can be coloured with five colours* – itself a remarkable result – but he was unable to fill the gap.

In fact, the gap took another eighty-six years to fill. In 1904 P. Wernicke proved that the pentagon can be replaced by a pair of adjacent pentagons and a pentagon adjacent to a hexagon, thereby obtaining a more complicated unavoidable set that could be tested for reducibility. A few years later, G.D. Birkhoff [1913], who learned of the four colour problem from Veblen while studying at Princeton University, showed that various other configurations, such as the 'Birkhoff diamond' of four adjacent pentagons (see Fig. 31), are reducible. By this means he was able to prove that if there were a map that required five colours, then it could not contain a ring of four or five regions; from this, he deduced that any such map must have at least 13 countries.

This two-pronged attack of contructing unavoidable sets and proving configurations to be reducible would eventually prove successful. On the one hand, one would replace the pentagon by more and more complicated unavoidable sets. For example, in 1922 Philip Franklin proved that every map must have at least twelve pentagons, and must contain either a pentagon adjacent to two other pentagons, a pentagon adjacent to a pentagon and a hexagon, or a pentagon adjacent to two hexagons; using this unavoidable set he proved the four colour theorem for maps with up to 25 countries. Unavoidable sets were also given by C.N. Reynolds, Henri Lebesgue and others, and over the years the number of countries in a map that required five colours continued to grow.

On the other hand, one could try to obtain larger and larger lists of reducible configurations. The ultimate aim was to find *an unavoidable set of reducible configurations*, since every map would have to contain at least one such configuration, and whichever it was, the colouring of the configuration could then be extended to the whole map.

Around 1970, H. Heesch believed, for probabilistic reasons, that a *finite* unavoidable set of reducible configurations must exist, and that the number of such configurations need not exceed 9000. In addition, Heesch developed a technique for constructing unavoidable sets, called the *discharging method*, and noticed that there are certain features that seem to prevent a configuration from being reducible. These ideas were developed by Kenneth Appel and Wolfgang Haken, who spent several years developing computer programs that would help in the search for unavoidable configurations and assist in the testing of reducibility. By this means they were eventually able to produce after some 1200 hours of computer time, an unavoidable set of almost 2000 reducible configurations, thereby completing the proof of the four colour theorem (see Appel and Haken [1977a, 1977b] and Appel, Haken and Koch [1977]). Indeed, since they had constructed many thousands of such unavoidable sets, they had thousands of proofs of the theorem, and if any individual configuration were subsequently to be proved irreducible, this would not invalidate their work.

Since then, the technical details of the proof have been simplified somewhat, and the configurations have been checked on other computers, but no easily verifiable proof has yet been found. Because of this, and because Appel and Haken's work raised a number of interesting philosophical questions about the nature of mathematical proof, the mathematical world was slow to acclaim their magnificent achievement.

4.2. *Maps on other surfaces*

The above discussion on the colouring of maps drawn on the plane applies equally well to maps drawn on a sphere, and this leads us to ask how many colours are needed for maps drawn on a torus. Indeed, in the 1890 paper in which he demolished Kempe's 'proof', Heawood raised this very question, proving that seven colours are sufficient for maps drawn on a torus and that for some such maps seven colours are actually needed. A example of such a map is given in Fig. 32.

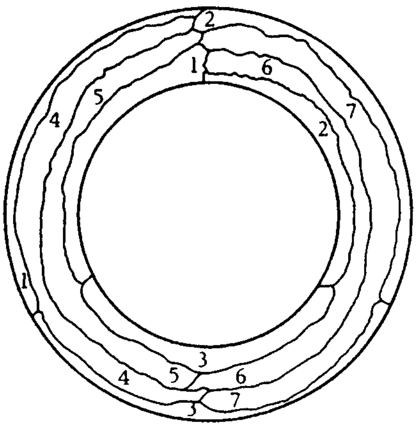


Fig. 32.

Heawood also extended the idea to the colouring of maps on a sphere with g handles where, by using Lhuilier's result on the Euler characteristic of such a surface, he was able to show that the appropriate number of colours is

$$\lfloor (7 + \sqrt{1 + 48g})/2 \rfloor.$$

Unfortunately, he omitted to prove that there are maps that actually require this number of colours, and the result became known as the *Heawood conjecture*. It took almost eighty years for the gap to be filled. The proof splits up into no fewer than twelve separate arguments, depending on the residue class of g modulo 12, and some of these cases turned out to be particularly intransigent.

A related question is to ask how many colours are needed for maps drawn on *non-orientable* surfaces. For a surface of non-orientable genus g (a sphere with g cross-caps), the Euler characteristic is $2 - g$, and the appropriate number of colours becomes

$$\lfloor (7 + \sqrt{1 + 24g})/2 \rfloor,$$

as proved by Heinrich Tietze [1910]. In 1935 I. Kagno proved that there are maps that require this number of colours when $g = 3$, 4 or 6, but in the previous year Philip Franklin [1934] showed that for the Klein bottle, where $g = 2$, the correct number of colours turns out to be 6, rather than the value 7 given by the above formula. Eventually, Ringel obtained the complete solution, proving that the above formula does indeed give the correct number of colours, with the single exception of the Klein bottle (see Ringel [1974]).

4.3. *Other colouring problems*

As stated, the four colour map problem is not a problem in graph theory. However, as Kempe pointed out in his 1879 paper, the problem can be dualized to give a problem on the colouring of vertices (see Fig. 33); in this formulation, we are required to colour the vertices of a planar graph in such a way that adjacent vertices are differently coloured; this reformulation was the version in which Appel and Haken's solution was presented.

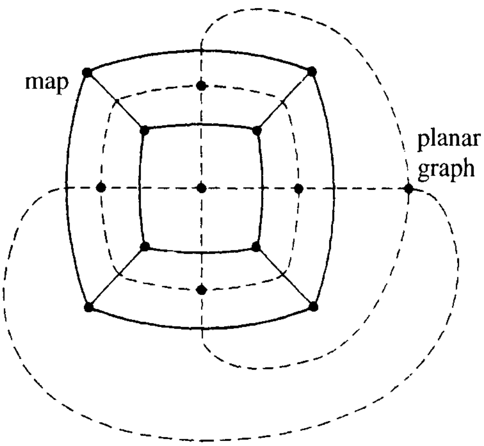


Fig. 33.

This idea of colouring the vertices of a graph so that adjacent vertices are differently coloured developed a life of its own in the 1930s, mainly through the work of Whitney who wrote his Ph.D. thesis on the colouring of graphs, R.L. Brooks [1941], who obtained a good upper bound on the number of colours required, and G.A. Dirac [1952], who introduced the idea of a *critical graph*. Whitney developed for graphs an idea of G.D. Birkhoff [1912]; this is the *chromatic polynomial* of a map, where the number of possible colourings is a polynomial function of the number of colours available. Such a polynomial can be usefully studied in its own right, as has been done to great effect by G.D. Birkhoff and D.C. Lewis [1946], W.T. Tutte [1970] and others.

The above-mentioned contribution of Tait [1878–80], in which one colours the edges of a graph in such a way that any two edges that meet are differently coloured, also developed a life of its own. It is clear that if we have a graph with maximum vertex degree k, then we need at least k colours to colour its edges. Dénes König proved in 1916 that for *bipartite* graphs k colours are sufficient; for example, the edges of the bipartite graph in Fig. 34 can be coloured with 3 colours. In 1949, Claude Shannon formulated a problem involving electrical relays as an edge-colouring graph problem. Then, in a pair of fundamental papers, V.G. Vizing [1964, 1965] proved that $k + 1$ colours are always sufficient, leading to the *classification problem* of trying to decide which graphs need only k colours and which graphs need $k+1$ colours.

The study of such colouring problems blossomed throughout the 1970s and 1980s; further information about these developments can be found in Fiorini and Wilson [1977] and Jensen and Toft [1995].

4.4. *Factorization*

A graph is **k-regular** if each of its vertices has degree k. Such graphs can sometimes be split into regular subgraphs, each with the same vertex set as the original graph. For example, the complete graph K_5, which is 4-regular, can be split into two 2-regular subgraphs (see Fig. 35), and the 3-regular bipartite graph in Fig. 34 can be split into three 1-regular subgraphs. An **k-factor** in a graph is a k-regular subgraph that contains all the vertices of the original graph; for example, the graphs in Figs. 34 and 35 split into three 1-factors and two 2-factors, respectively. Note that, if we assign a different colour to each 1-factor in Fig. 34, then we obtain the above 3-colouring of the edges of the graph.

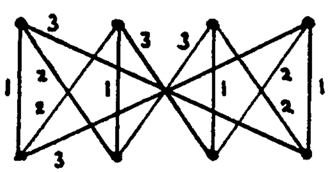


Fig. 34.

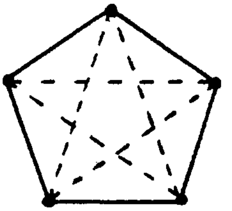


Fig. 35.

In 1891 Julius Petersen wrote a fundamental paper on the factorization of regular graphs, arising from a problem in the theory of invariants. In this paper he proved that *if k is even, then any k-regular graph can be split into 2-factors*. He also proved that *any 3-regular graph possesses a 1-factor, provided that it has not more than two 'leaves'*; a *leaf* is a subgraph joined to the rest of the graph by a single edge. A few years later, he produced a trivalent graph (Petersen [1898]) with no leaves, now called the **Petersen graph** (Fig. 36), which cannot be split into three 1-factors; it can, however, be split into a 1-factor (the spokes) and a 2-factor (the pentagon and pentagram).

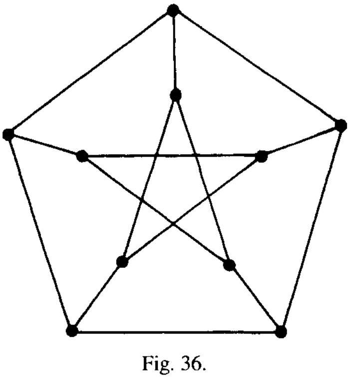


Fig. 36.

5. Algorithmic graph theory

Graph theory algorithms can be traced back over one hundred years to when Fleury gave a systematic method for tracing an Eulerian graph and G. Tarry [1895] showed how to escape from a maze. During the 20th century such algorithms increasingly came into their own, with the solutions of such problems as the *shortest and longest path problems*, the *minimum connector problem*, and the *Chinese postman problem*. In each of these problems we are given a network, or *weighted graph*, to each edge (and/or vertex) of which has been assigned a number, such as its length or the time taken to traverse it.

There are several efficient algorithms for finding the shortest path in a given network, of which the best known is due to E.W. Dijkstra [1959]. Finding a longest path, or *critical path*, in an activity network also dates from the 1950s, with PERT (Program Evaluation and Review Technique) used by the US Navy for problems involving the building of submarines and CPM (Critical Path Method) developed by the Du Pont de Nemours Company in order to minimize the total cost of a project. The *Chinese postman problem*, for finding the shortest route that covers each edge of a given weighted graph, was solved by Meigu Guan (Mei-Ku Kwan) [1960]. The greedy algorithm for the *minimum connector problem*, in which we seek a minimum-length spanning tree in a weighted graph, can be traced back to O. Boruvka [1926] and was later rediscovered by J.B. Kruskal [1956].

A related problem is the *travelling salesman problem*, in which a salesman wishes to to make a cyclic tour of a number of cities in minimum time or distance. This problem appeared in rudimentary form in a practical book written for the *Handlungsreisende* (see Voigt [1831]), but its first appearance in mathematical circles was not until the early 1930s, at Princeton. It was later popularized at the RAND Corporation, leading eventually to the fundamental paper of G.B. Dantzig, D.R. Fulkerson and S.M. Johnson [1954] which included the solution of a travelling salesman problem with 49 cities. Over the years the number of cities was gradually increased, and in the 1980s a problem with 2392 cities was settled by Padberg and Rinaldi [1987]. An extensive survey of the travelling salesman problem can be found in E.L. Lawler et al. [1985].

The travelling salesman problem was not the only significant combinatorial problem studied at the RAND Corporation in the mid-20th century. In particular, algorithms were developed by Ford and Fulkerson [1956] for finding the maximum flow of a commodity between two nodes in a capacitated network, and by Gomory and Hu [1961] for determining maximum flows in multi-terminal networks. Algorithms for solving matching and assignment problems were developed, where one wishes to assign people as appropriately as possible to jobs for which they are qualified; this work developed from the above-mentioned work of König and from a celebrated result on matching due to Philip Hall [1935], later known as the 'marriage theorem' (see Halmos and Vaughan [1950]).

These investigations led eventually to the subject of *polyhedral combinatorics*, and were combined with the newly emerging study of linear programming. Further information can be found in a historical article by Dantzig [1982] and in a lengthy survey of matching theory by L. Lovász and M.D. Plummer [1986].

6. Conclusion

Our aim in this article has been to survey the development of the main themes in graph theory, tracing them from the earliest times and showing how current research has evolved from earlier problems. Inevitably some important topics have had to be partly or completely overlooked, and some mathematicians have been slighted by the omissions of their contributions; nevertheless, we hope that we have been able to convey some idea of the nature and content of graph theory, both past and present.

Bibliography

Appel, K. and Haken, W. (1977a), *Every planar map is four colorable: Part 1, Discharging*, Illinois J. Math. **21**, 429–490.

Appel, K. and Haken, W. (1977b), *The solution of the four-color-map problem*, Scientific American **237** (4) (October), 108–121.

Appel, K., Haken, W. and Koch, J. (1977), *Every map is four colourable: Part 2, Reducibility*, Illinois J. Math. **21**, 491–567.

Ascher, M. (1991), *Ethnomathematics: A Multicultural View of Mathematical Ideas*, Brooks & Cole, Pacific Grove, CA.

Bermond, J.-C. (1978), *Hamiltonian graphs*, Selected Topics in Graph Theory, L.W. Beineke and R.J. Wilson, eds, Academic Press, London, 127–167.

Biggs, N.L., Lloyd, E.K. and Wilson, R.J. (1998), *Graph Theory 1736–1936*, Clarendon Press, Oxford, revised paperback edition.

Birkhoff, G.D. (1912), *A determinantal formula for the number of ways of coloring a map*, Ann. of Math. **14**, 42–46.

Birkhoff, G.D. (1913), *The reducibility of maps*, Amer. J. Math. **35**, 115–128.

Birkhoff, G.D. and Lewis, D.C. (1946), *Chromatic polynomials*, Trans. Amer. Math. Soc. **60**, 355–451.

Boruvka, O. (1926), *O jistém problému minimálním*, Acta Soc. Sci. Natur. Moravicae **3**, 37–58.

Brooks, R.L. (1941), *On colouring the nodes of a network*, Proc. Cambridge Philos. Soc. **37**, 194–197.

Cauchy, A.-L. (1813), *Recherches sur les polyèdres – premier mémoire*, J. Ecole Polytech. **9** (Cah. 16), 68–86.

Cayley, A. (1857), *On the theory of the analytical forms called trees*, Phil. Mag. **13** (4), 172–176.

Cayley, A. (1874), *On the mathematical theory of isomers*, Phil. Mag. **47** (4), 444–446.

Cayley, A. (1879), *On the colouring of maps*, Proc. Roy. Geog. Soc. (new Ser.) **1**, 259–261.

Cayley, A. (1889), *A theorem on trees*, Quart. J. Pure Appl. Math. **23**, 376–378.

Clausen, T. (1844), [*Second postscript to*] *De linearum tertii ordinis proprietatibus*, Astron. Nachr. **21**, col. 209–216.

Coupy, E. (1851), *Solution d'un problème appartenant à la géométrie de situation par Euler*, Nouv. Ann. Math. **10**, 106–119.

Crum Brown, A. (1864), *On the theory of isomeric compounds*, Trans. Roy. Soc. Edinburgh **23**, 707–719.

Dantzig, G.B. (1982), *Reminiscences about the origins of linear programming*, Oper. Res. Lett. **1**, 43–48.

Dantzig, G.B., Fulkerson, D.R. and Johnson, S.M. (1954), *Solution of a large-scale traveling-salesman problem*, Oper. Res. **2**, 393–410.

Dehn, M. and Heegaard, P. (1907), *Analysis situs*, Encyklopädie der mathematischen Wissenschaften, Vol. III AB3, 153–220.

De Morgan, A. (1860), *A review of* The philosophy of discovery, chapters historical and critical, *by W. Whewell, D.D.*, Athenaeum No. 1694, 501–503.

Dijkstra, E.W. (1959), *A note on two problems in connexion with graphs*, Numer. Math. **1**, 269–271.

Dirac, G.A. (1952), *Some theorems on abstract graphs*, Proc. London Math. Soc. **2** (3), 69–81.

Dudeney, H.E. (1913), *Perplexities*, Strand Mag. **46** (July 1913), 110 and (August 1913), 221.

Euler, L. (1736), *Solutio problematis ad geometriam situs pertinentis*, Commentarii Academiae Scientiarum Imperialis Petropolitanae **8**, 128–140.

Euler, L. (1750), [*Letter to Christian Goldbach*], Leonhard Euler und Christian Goldbach: Briefwechsel 1729–1764, A.P. Juškevič and E. Winter, eds, Akademie-Verlag, Berlin, 1965.

Euler, L. (1759), *Solution d'une question curieuse qui ne paroit soumise à aucune analyse*, Mem. Acad. Sci. Berlin **15**, 310–337.

Fiorini, S. and Wilson, R.J. (1977), *Edge-Colourings of Graphs*, Research Notes in Mathematics, Vol. 16, Pitman, London.

Ford, L.R. and Fulkerson, D.R. (1956), *Maximal flow through a network*, Canad. J. Math. **8**, 399–404.
Frankland, E. (1866), *Lecture Notes for Chemical Students*, London.
Franklin, P. (1922), *The four color problem*, Amer. J. Math. **44**, 225–236.
Franklin, P. (1934), *A six color problem*, J. Math. Phys. **13**, 363–369.
Glover, H.H., Huneke, J.P. and Wang, C.S. (1979), 103 *graphs that are irreducible for the projective plane*, J. Combin. Theory (B) **27**, 332–370.
Gomory, R.E. and Hu, T.C. (1961), *Multi-terminal network flows*, SIAM J. Appl. Math. **9**, 551–556.
Guan Meigu (1960), *Graphic programming using odd or even points*, Acta Math. Sinica **10**, 263–266; (1962), Chinese Math. **1**, 273–277.
Hall, P. (1935), *On representatives of subsets*, J. London Math. Soc. **10**, 26–30.
Halmos, P.R. and Vaughan, H.E. (1950), *The marriage problem*, Amer. J. Math. **72**, 214–215.
Hamilton, W.R. (1856), *Memorandum respecting a new system of roots of unity*, Phil. Mag. **12** (4), 446.
Heawood, P.J. (1890), *Map-colour theorem*, Quart. J. Pure Appl. Math. **24**, 332–338.
Heesch, H. (1969), *Untersuchungen zum Vierfarbenproblem*, B.I. Hochschulscripten, 810/810a/810b, Bibliographisches Institut, Mannheim–Vienna–Zürich.
Hierholzer, C. (1873), *Uber die Möglichkeit, einen Lineanzug ohne Wiederholung und ohne Unterbrechnung zu umfahren*, Math. Ann. **6**, 30–32.
Jaenisch, M. (1862–63), *Applications de l'Analyse Mathématiques au Jeu des Echecs*, 3 Vols, St. Petersburg.
Jensen, T.R. and Toft, B. (1995), *Graph Coloring Problems*, Wiley-Interscience, New York.
Kagno, I.N. (1935), *A note on the Heawood color formula*, J. Math. Phys. **14**, 228–231.
Kempe, A.B. (1879), *On the geographical problem of the four colours*, Amer. J. Math. **2**, 193–200.
Kirchhoff, G.R. (1847), *Über die Auflösung der Gleichungen, auf welche man bei der Untersuchung der linearen Vertheilung galvanischer Ströme geführt wird*, Ann. Phys. Chem. **72**, 497–508.
Kirkman, T.P. (1856), *On the representation of polyedra*, Phil. Trans. Roy. Soc. London **146**, 413–418.
König, D. (1916), *Uber Graphen und ihre Anwendung auf Determinantentheorie und Mengenlehre*, Math. Ann. **77**, 453–465.
Kruskal, J.B. (1956), *On the shortest spanning subtree of a graph and the traveling salesman problem*, Proc. Amer. Math. Soc. **7**, 48–50.
Kuratowski, K. (1930), *Sur le problème des courbes gauches en topologie*, Fund. Math. **15**, 271–283.
Lawler, E.L., Lenstra, J.K., Rinnooy Kan, A.H.G. and Schmoys, D.B. (eds) (1985), *The Traveling Salesman Problem: A Guided Tour through Combinatorial Optimization*, Wiley, Chichester.
Legendre, A.M. (1794), *Eléments de Géométrie* (1st edn), Firmin Didot, Paris.
Lhuilier, S. (1811), *Démonstration immédiate d'un théorème fondamental d'Euler sur les polyhèdres, et exceptions dont ce théorème est susceptible*, Mém. Acad. Imp. Sci. St. Pétersb. **4**, 271–301.
Listing, J.B. (1847), *Vorstudien zur Topologie*, Göttinger Studien (Abt. 1) Math. Naturwiss. Abh. **1**, 811–875.
Listing, J.B. (1861–62), *Der Census räumliche Complexe*, Abh. K. Ges. Wiss. Göttingen Math. Cl. **10**, 97–182.
Lovász, L. and Plummer, M.D. (1986), *Matching Theory*, Annals of Discrete Mathematics vol. 29, North-Holland.
Lucas, E. (1882), *Récréations Mathématiques*, Vol. 1, Gauthier-Villars, Paris.
Lunn, A.C. and Senior, J.K. (1929), *Isomerism and configuration*, J. Phys. Chem. **33**, 1027–1079.
Mayer, J. (1969), *Le problème des régions voisines sur les surfaces closes orientables*, J. Combin. Theory **6**, 177–195.
Ore, O. (1960), *Note on Hamiltonian circuits*, American Math. Monthly **67**, 55.
Padberg, M.W. and Rinaldi, G. (1987), *Optimization of a 532-city symmetric traveling salesman problem by branch and cut*, Oper. Res. Lett. **6**, 1–7.
Petersen, J. (1891), *Die Theorie der regulären Graphs*, Acta Math. **15**, 193–220.
Petersen, J. (1898), [*Sur le théorème de Tait*], Interméd. Math. **5**, 225–227.
Poinsot, L. (1809–10), *Sur les polygones et les polyèdres*, J. Ecole Polytech. **4** (Cah. 10), 16–48.
Pólya, G. (1937), *Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen*, Acta Math. **68**, 145–254.
Pólya, G. and Read, R.C. (1987), *Combinatorial Enumeration of Groups, Graphs, and Chemical Compounds*, Springer, Berlin–Heidelberg–New York.
Pont, J.C. (1974), *La Topologie Algébrique des Origines à Poincaré*, Bibl. de Philos. Contemp., Presses Universitaires de France, Paris.
Prüfer, H. (1918), *Neuer Beweis eines Satzes über Permutationen*, Arch. Math. Phys. **27** (3), 142–144.
Redfield, J.H. (1927), *The theory of group-reduced distributions*, Amer. J. Math. **49**, 433–455.

Reiss, M. (1871–73), *Evaluation du nombre de combinaisons desquelles les 28 dés d'un jeu du domino sont susceptibles d'après la règle de ce jeu*, Ann. Mat. Pura Appl. **5** (2), 63–120.

Ringel, G. (1974), *Map Color Theorem*, Springer, Berlin.

Robertson, N. and Seymour, P.D. (1985), *Graph minors – a survey*, Surveys in Combinatorics 1985, I. Anderson, ed., London Math. Soc. Lecture Notes Series, Vol. 103, Cambridge Univ. Press, 153–171.

Rouse Ball, W.W. (1892), *Mathematical Recreations and Problems of Past and Present Times* (later entitled *Mathematical Recreations and Essays*), Macmillan, London.

Russell, C.A. (1971), *The History of Valency*, Leicester Univ. Press.

Sachs, H., Stiebitz, M. and Wilson, R.J. (1988), *An historical note: Euler's Königsberg letters*, J. Graph Theory **12**, 133–139.

Shannon, C.E. (1949), *A theorem on coloring the lines of a network*, J. Math. Phys. **28**, 148–151.

Sylvester, J.J. (1877–78), *Chemistry and algebra*, Nature **17**, 284.

Sylvester, J.J. (1878), *On an application of the new atomic theory to the graphical representation of the invariants and covariants of binary quantics*, Amer. J. Math. **1**, 64–125.

Tait, P.G. (1878–80), [*Remarks on the colouring of maps*], Proc. Roy. Soc. Edinburgh **10**, 729.

Tarry, G. (1895), *Le problème des labyrinthes*, Nouv. Ann. Math. (3) **14**, 187–190.

Tietze, H. (1910), *Einige Bemerkungen über das Problem des Kartenfärbens auf einseitigen Flächen*, Jahresber. Deut. Math.-Ver. **19**, 155–179.

Tutte, W.T. (1946), *On hamiltonian circuits*, J. London Math. Soc. **21**, 98–101.

Tutte, W.T. (1959), *Matroids and graphs*, Trans. Amer. Math. Soc. **90**, 527–552.

Tutte, W.T. (1970), *On chromatic polynomials and the golden ratio*, J. Combin. Theory **9**, 289–296.

Vandermonde, A.-T. (1771), *Remarques sur les problèmes de situation*, Mém. Acad. Sci. (Paris), 556–574.

Veblen, O. (1922), *Analysis Situs*, Amer. Math. Soc. Colloq. Lect. 1916, New York.

Vizing, V.G. (1964), *On an estimate of the chromatic class of a p-graph*, Diskret. Analiz **3**, 25–30.

Vizing, V.G. (1965), *The chromatic class of a multigraph*, Diskret. Analiz **5**, 9–17.

Voigt, B.F. (1831), *Der Handlungsreisende, wie er sein soll und was er zu thun hat, um Aufträge zu erhalten und einer glücklichen Erfolgs in seinen Geschäften gewiss zu sein* (*von einem alten Commis-Voyageur, Ilmenau*); reprinted 1981, Schramm, Kiel.

Wernicke, P. (1904), *Über den kartographischen Vierfarbensatz*, Math. Ann. **58**, 413–426.

Whitney, H. (1931), *Non-separable and planar graphs*, Proc. Nat. Acad. Sci. USA **17**, 125–127.

Whitney, H. (1935), *On the abstract properties of linear dependence*, Amer. J. Math. **57**, 509–533.

This Page Intentionally Left Blank

CHAPTER 18

The Early Development of Algebraic Topology*

Solomon Lefschetz (1884–1972)

1. Early history

The beginnings of algebraic topology share this with the beginnings of any important chapter of mathematics that its roots are more or less obscure. Those of algebraic topology are found mostly in geometry and did not contain the promise of a major field. Since my proposed excursion has nothing archeological and hardly any historical aspect, I will concentrate on the following major points. First Euler's characteristic, then the Möbius strip and its significance for orientability. I will conclude (for special reasons) with a section on knots.

1.1. *Euler's characteristic*

This is certainly one of the earliest manifestations of algebraic topology. Let a convex polyhedron Π in a 3-space have F faces, E edges and V vertices. Euler's formula asserts that always

$$F - E + V = 2. \tag{1.1}$$

The expression at the left is the *characteristic* $\chi(\Pi)$ of Π.

Let 0 be an interior point of the polyhedron and S a sphere of center 0. Project Π onto S from 0. This results in a partition of the sphere into F polygonal regions, with E sides and V vertices and (1.1) still holds. It is interesting, however, to observe that it is known to hold for any partition of a sphere into a finite number of polygonal regions. In other words it represents actually a property of the sphere S itself: topological property. In fact it holds as well, for example, for an ellipsoid, or for any "like" figure.

In order to calculate this fixed value of $\chi(\Pi)$ one may, therefore, take any simple decomposition, for example: a great circle made into a polygon with one vertex and one arc plus the two hemispheres. Thus there are one vertex, one edge and two faces, so that $\chi(\Pi) = 2$. Euler's proof was à la old geometry, but his proof is easily topologized, as done much later by Poincaré (1895).

*Extracted from Bol. Soc. Bras. Matem. **1** (1970), 1–48.

HISTORY OF TOPOLOGY
Edited by I.M. James

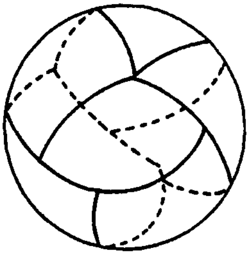


Fig. 1.

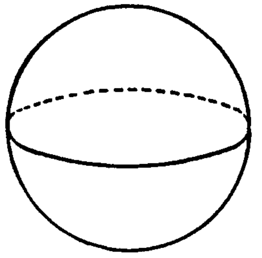


Fig. 2.

1.2. *The Möbius strip (1850)*

Let ABCD be a plane rectangle. Match AD with BC so that A coincides with C and B with D. One sees then readily that one cannot match the orientation of AD with that of BC so that any side common to two triangles is oppositely oriented to both. Intuitively one finds that a small oriented circuit on the strip may be so displaced as to return to its original position with reversed orientation. Poincaré described this as the return upside down of a fly crawling on the strip.

A smooth surface is *orientable* when it contains no part like a Möbius strip, and it is nonorientable when it does contain a part like this strip. Thus a 2-sphere is orientable, but the projective plane is not. The second statement is not quite obvious, but is easily proved along the following lines. An open line L in an ordinary plane is orientable (evident) and remains so when its two end points are made to coincide turning the line into a circle C. Take now an origin 0 in a plane and a circular region of center 0 bounded by a circle D (Fig. 4). Let any diameter have end points AA′ on the circle D. The open interval (A, A′) is the perfect image of the line L. One closes it by bringing the two points A and A′ into coincidence. The operation on the circle D has for effect to bring all diameter pairs of points into coincidences and then one has the perfect image of a projective plane. Let (A, A′) and (B, B′) be two terminal pairs of points. Upon joining A to B′ and A′ to B one finds that the projective plane contains the perfect image of a Möbius strip and so it is nonorientable.

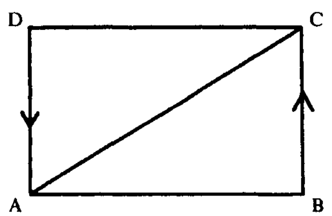


Fig. 3.

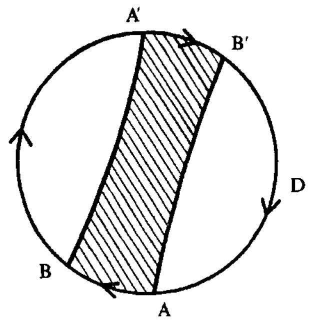


Fig. 4.

1.3. *On knots*

This is assuredly the most curious and most perplexing chapter of algebraic topology. One may also say of it that while it has borrowed enormously from the rest of algebraic topology it has returned very scant interest on this "borrowed" capital. It is, however, full of problems *sui generis* with some of the simplest, in formulation, as yet unsolved. In this respect it resembles considerably number theory.

Our main reason for placing "something about knots" in this early location, is the impossibility to give more than a faint notion of this topic in a reasonable space. We shall therefore merely indicate a very few salient points and refer for knots to the highly interesting and thorough exposition given in the recent monograph *Introduction to Knot theory* by R.H. Crowell and R.H. Fox, Ginn & Co., Boston, 1963. (Hereafter referred to as CF.) This book contains also a good guide to the literature, an extensive bibliography and a wealth of figures. For the few points to be discussed here, no better source of references can be found. It is not possible to touch knot theory at any point without utilizing many advanced topological concepts. For most of these brief indications will be found in Sections 3 and 4.

DEFINITION 1.1. A knot is merely a graph in 3-space $\mathfrak{E}_3$ which as a point-set is the homeomorph of a circle.

As a graph then the knot K consists of a finite set of points $A_1, A_2, \ldots, A_n$, which are joined, consecutively, by arcs (A_n joined to A_1). The arcs may be assumed differentiable (the complications à la sophisticated Jordan curves are avoided).

We will assume that K, as a Jordan curve, *is oriented*. K designates the knot with a definite orientation; $-K$ will denote it with the opposite orientation.

Let K, K_1 be two knots in the same $\mathfrak{E}_3$. We consider them as *equivalent*: $K \simeq K_1$, whenever there exists a homeomorphic deformation of $\mathfrak{E}_3$ into itself under which K_1 goes into K.

To illustrate the perversity of knots, Trotter has proved recently (by an infinity of rather simple examples) that there exist knots K not $\simeq -K$ (thus solving a long outstanding problem).

A *knot invariant* is a knot character which is the same for all equivalent knots.

The central problem of knot theory is to find a collection of knot invariants which guarantee that if they are the same for two knots K, K_1 then $K \simeq K_1$ (K and K_1 are assumed imbedded in the same $\mathfrak{E}_3$). Although this central problem has been attacked, in our century at least, by many very eminent mathematicians, it is doubtful if we are nearer to a solution than a century ago.

The most important invariant of a knot K is the group of paths $\Pi(\mathfrak{E}_3 - K)$ of its complement. However, Trotter's example shows that this is not a "decisive" invariant.

The following "knottists" J.W. Alexander and R.H. Fox, will be mainly mentioned. Alexander attacked knots in the twenties, Fox belongs to the forties to date. We owe to Alexander two noteworthy but related sources of invariants: Alexander matrices and Alexander polynomials. Both center around the concept of projection of a knot onto a plane. The projection is a plane graph whose sides may intersect, but one may organize the situation so that (a) the self intersections are never nodes of the graph; (b) they are always double points with distinct tangents.

The mere penetration of Knot theory requires a formidable amount of modern algebra, far more than I can go into here. A little of it is, however, indispensable even for a bare description of a few main concepts.

Let $G = \{g\}$ be a multiplicative group. Let J be the ring of integers. With G there is associated the group ring JG, defined as the set of mappings $\nu : G \to J$ such that $\nu(g) = 0$ except for at most a finite set of $g's \in G$. Addition and multiplication in JG are defined by

$$(\nu_1 + \nu_2)g = \nu_1 g + \nu_2 g; \qquad (\nu_1\nu_2)g = \sum(\nu_1 h)(\nu_2 h^{-1} g)$$

for any $\nu_1,\ \nu_2$ of JG and any g of G. One may easily verify that JG is a ring under these operations; also that if n is any integer then $(n\nu)g = n(\nu g)$.

The *free calculus*, introduced by Fox, yields a most powerful technique for calculating knot invariants. The basis of this calculus is this definition of a *derivative* D as the unique linear extension to JG of any mapping $\mathrm{D} : G \to JG$ which satisfies for $g_1,\ g_2$ in G:

$$\mathrm{D}(g_1 g_2) = \mathrm{D}g_1 + g_1 \mathrm{D}g_2.$$

For further information see CF, Chapter VII.

The weight of algebra in Knot theory is best indicated by this: in CF out of eight chapters five are on pure algebraic questions (mainly general group theory). I should like to recommend to the advanced reader and to any one interested in new and up to date problems, the *Guide to the literature* at the end of CF.

1.4. *On braids*

If one severs a knot at one joint one obtains a braid (Emil Artin 1925). The group of braids (Artin) is defined as:

$$\sigma_1, \sigma_2, \ldots, \sigma_n; \quad \sigma_1\sigma_{i+1}\sigma_i = \sigma_{i+1}\sigma_i\sigma_{i+1}, \quad i = 1, 2, \ldots, n-1;$$
$$\sigma_i\sigma_j = \sigma_j\sigma_i, \quad (i-j) \neq 1.$$

These groups have been completely classified by Artin – they do not offer the complications of Knot groups.

2. Riemann and Riemann surfaces. Construction. Number of integrals of first kind. The work of Scorza

2.1. *Puiseux's theorem*

This is really the initial place of algebraic topology. Not that Riemann himself thought of it that way, but this will cause no argument.

The problem attacked by Riemann, (around 1850) among many others, was the nature, as geometry, of a complex plane algebraic curve

$$F(x, y) = 0, \tag{2.1}$$

where F is a complex irreducible polynomial. Much was known about the geometry of real curves – since the period was post-Plücker – but "as geometry" complex curves remained obscure. Important information was contained in the *Theorem of Puiseux*. But even this theorem did not offer any information in the large about the curve F. However, we shall need the material provided by Puiseux.

Let $x = a$ be a value for which the roots of $F = 0$ in y remain finite. Let $y_1(x)$ be such a root. As x describes a small circle around $x = a$ in its complex plane the root $y_1(x)$ varies continuously and at all times say around $x = x_0$ on C it is holomorphic around x_0 as a function of x. Hence, as x describes C once $y_1(x)$ returns to a value which is still a root of F in y but not necessarily the same root $y_1(x)$. Let it return to a different root $y_2(x)$, etc. There arises a set of say q roots $y_1(x), \ldots, y_q(x)$ which are circularly permuted as x describes C. This implies that these q roots are represented by q series in powers of $(x - a)^{1/q}$ or as

$$y(x) = b + \alpha(x - a)^{p_1/q} + \beta(x - a)^{p_2/q} + \cdots. \tag{2.2}$$

The q roots of the circular system may be jointly represented by a unique series in t:

$$x = a + t^q, \quad y = b + \alpha t^{p_1} + \beta t^{p_2} + \cdots, \quad |t| < \rho \tag{2.3}$$

where $q,\ p_1,\ p_2, \ldots,$ have no common factor.

Since the number of values $x = a$ with true circular representations is finite the points corresponding to $0 < |t| < \rho$, are *ordinary* points of the curve. That is: (a) to any such point there corresponds only one solution $y(a)$; (b) the corresponding $q = 1$.

We have assumed that a is finite. The points at infinity are taken care of by the standard transformation

$$x' = \frac{1}{x}, \qquad y' = \frac{y}{x}.$$

If m is the degree of F in y there are at most m such points and hence at most m series $\{x'(t), y'(t)\}$. The transformation just introduced merely means that the true space of the curve F is a projective plane. It was well known (as an analytical artifice) to the mathematicians preceding Riemann: Abel, Jacobi, Plücker, Weierstrass, and many others.

The set $\pi = \{$pair of series in t, point $t = 0$, number $\rho\}$ is called a *place* of F and $t = 0$ is the *center* of the place. The number ρ, the *convergency radius* of the series, is the *extension* of the place. It is agreed that the change in ρ, provided that it does not reach another singular place, does not affect π. (Singular place is one which is the center of several distinct places, or of a single place around which more roots than one $y_j(x)$ are permuted.)

(The concepts related to places have been clearly set down by Hermann Weyl in the monograph: *Die Idee der Riemannschen Fläche*, Springer, Berlin (1913).) However, there are well founded reasons to believe that "places" were no strangers for Riemann. The main reason would go something like this. Let (a, b) be a center of several places $\pi_1, \ldots, \pi_s$. In his construction of the Riemann surface, Riemann always represents the π_h by s distinct points.

2.2. *Construction of the Riemann surface*

Suppose that m is the degree of F in y. Take the sphere S of the variable x and mark on it two diametral points A and B so disposed that no great circle through them contains more than one of the *critical* points $a_1, a_2, \ldots, a_n$ which are centers of singular places. Mark on S arcs of great circles Aa_h and cut S along these arcs. Choose now one sphere S_h for each root $y_h(x)$ of F. Mark on S_h the cuts Aa_h which do permute $y_h(x)$. The complement Ω_h of the cuts on S_h is a 2-cell and the value $y_h(\xi)$ at $\xi \in \Omega_h$ is uniquely determined by the value $y_h(\xi(B))$. We look now at the 2-cell Ω_h and at its boundary the polygon π_h. Let all the π_h be positively oriented, that is let the Ω_h all be oriented in the same way. It follows at once that if A'_{hj} and A''_{hj} are the two sides of a cut A_{hj} then in that cut their orientations are opposite.

Suppose then that A_{hj} is a cut permuting y_k with y_j. Then there correspond to it two cuts A'_{hk} in Ω_k and A''_{hj} in Ω_j and those two are oppositely oriented. Hence, if we bring them back into coincidence, and similarly for all permuted pairs y_r, y_s the result is a closed surface $\Phi(F)$: the Riemann surface of the curve F. The construction has obtained these fundamental consequences:

(a) $\Phi(F)$ is covered by a finite collection of 2-cells $E_1, \ldots, E_m$, one for each root $y_j(x)$ of $F = 0$.

(b) If the polygons E_c, E_d have a common side then they are oppositely oriented relative to it.

(c) (less evident) Each point P of $\Phi(F)$ has a neighborhood in the surface which is a union of closed polygonal regions each making up a 2-cell.

(d) The surface $\Phi(F)$ is connected. This is a ready consequence of the irreducibility of the polynomial F. For if $\Phi(F)$ is not connected the roots $y_1(x), \ldots, y_m(x)$ may be divided into at least two collections say $y_1, \ldots, y_r$ and $y_{r+1}, \ldots, y_m$ whose elements are not permuted under the variation of x. Hence the symmetric functions of the y_h, $h \leqslant r$, are meromorphic in x and so satisfy a relation $F_1(x, y) = 0$, where F_1 is like F, but of smaller degree in y, and hence it is a proper factor of F. Since this contradicts the irreducibility of F, the surface $\Phi(F)$ is connected.

(e) Property (b) implies that $\Phi(F)$ is orientable.

CONCLUSION 2.1. *The preceding properties imply that $\Phi(F)$ is an orientable compact two dimensional manifold (in the sense of modern topology).*

We notice also:

(f) Under an appropriate definition of place-continuity the collection of places $\{\pi\}$ is turned into a surface homeomorphic with $\Phi(F)$.

The statement just made implies the following important result:

THEOREM 2.2. *The Riemann surface is a birational invariant.*

For the places have birational character and hence this holds also for their surface.

CHARACTERISTIC. A particular case of a very general property (Euler–Poincaré characteristics) asserts the following property: Let the polygons of the decomposition of $\Phi(F)$ consist of α_2 polygons, with α_1 sides and α_0 vertices then *whatever this decomposition* we have the relation

$$X(\Phi) = \alpha_0 - \alpha_1 + \alpha_h = 2 - 2p. \tag{2.4}$$

This is a classical formula due to De Jonquière.

The number p is the well known *genus* of the curve F. It will be shown later that the *characteristic has topological character. Hence the genus p is a topological invariant of the Riemann surface and therefore of the curve F.*

A direct calculation of $X(\Phi)$ is of interest. Let $\beta_0,\ \beta_1,\ \beta_2$ be the analogues of the α for Φ. Evidently if α_i are the same numbers for a 2-sphere, then from Euler's result

$$\alpha_0 - \alpha_1 + \alpha_2 = 2.$$

Also $\beta_1 = m\alpha_1,\ \beta_2 = m\alpha_2$. But for each place with q permuting roots $y(x)$ we lose $q - 1$ vertices. Hence if $N = \sum(q-1)$ then $\beta_0 = m\alpha_0 - N$. From this follows

$$\beta_0 - \beta_1 + \beta_2 = 2m - N = 2 - 2p.$$

Hence this formula due to Riemann

$$N = 2(p + m - 1). \tag{2.5}$$

2.3. *Topological models of a surface*

After Riemann, in the latter part of the 19th century, his surfaces, or more generally their topological type was deeply studied by a number of geometers (Klein, Clifford and others). Clifford showed that a surface of genus p was homeomorphic to a 2-sided disk with p holes. This model is identical to a sphere with p-handles (Fig. 5).

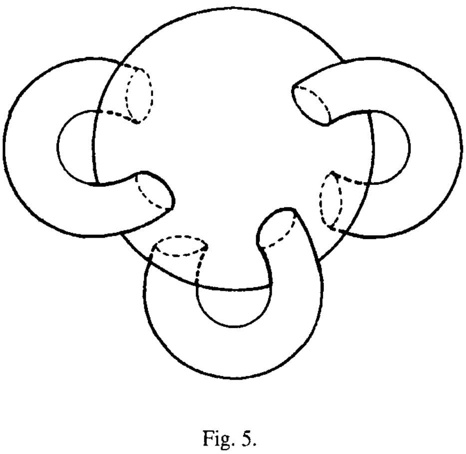


Fig. 5.

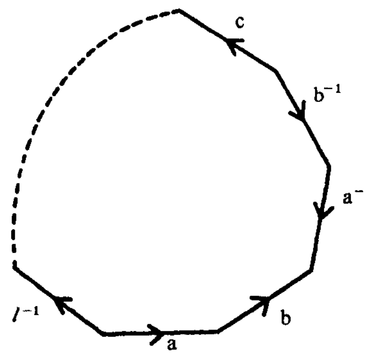


Fig. 6.

From the Clifford model one may obtain with little difficulty the most significant model of all: a polygonal region with sides matched in a certain way (Fig. 6). Draw on a plane a $4p$-sided regular polygonal region whose boundary polygon Π is to be described so that the successive sides are labelled (with their orientations)

$$a, b, a^{-1}, b^{-1}, c, d, c^{-1}, d^{-1}, \ldots, e^{-1}$$

($4p$-sides). Let the 2-cell bounded by Π be Ω.

The labels are such that for instance d^{-1} means d described in the opposite direction. Let now all the $4p$ vertices be brought into coincidence, and match for instance d with d^{-1} so that d^{-1} is merely d described in the opposite way.

From the new polygonal boundary, still called Π, say to the Clifford model is but a step and conversely. Hence the new model is a general model for a surface of genus p.

2.4. *Analytical application*

One might get the impression from what precedes that the Riemann surface is a pure geometrical instrument without further ado. This would be entirely misleading. For Riemann, like all his mathematical contemporaries was strongly under the influence of the theories created and developed by Cauchy. His surfaces show this plainly: it is at least through analysis that he obtained some of his most beautiful results. However, in expounding them I shall not endeavor to follow in Riemann's footsteps and shall not hesitate to utilize later results especially if they come under "early algebraic topology".

Consider then a function $f(z)$ on the Riemann surface $\Phi(F)$ which is uniquely defined on $\Phi(F)$ or perhaps on a region $\mathcal{R} \subset \Phi(F)$. We assume this property: If P is a point of $\mathcal{R}$ there is a place π of center P and parameter t ($|t| < \rho$) whose points are all in $\mathcal{R}$. On Π the function $f(t)$ is holomorphic in t. One defines f as holomorphic in $\mathcal{R}$, whenever it is holomorphic throughout $\mathcal{R}$.

2.5. *Extended Cauchy theorem*

THEOREM 2.3. *Let $\mathcal{R}$ be a 2-cell with boundary Π. Let $f(z)$ be holomorphic at all points of $\overline{\mathcal{R}}$ ($\mathcal{R}$ plus Π). Then*

$$\int_{\Pi} f(z)\,\mathrm{d}z = 0.$$

Let L be a line dividing $\mathcal{R}$ into two similar regions $\mathcal{R}_1$ and $\mathcal{R}_2$ with boundaries Π_1 and Π_2. Then

$$\int_{\Pi} = \int_{\Pi_1} + \int_{\Pi_2}.$$

It is sufficient therefore to prove the theorem for Π_1 and Π_2.

Consider the original decomposition of $\Phi(F)$ into polygons (one for each sheet of the surface). Let A be a vertex of one of the polygons. The set U consisting of all the open polygons and edges together with A, with vertex A – called the *star* of A, written St A, is an open set of $\Phi(F)$. In fact it is a place of center A and say parameter t ($t = 0$ at A). The collection $\mathcal{U} = \{U\}$ of all St A, is a *finite open covering* of $\Phi(F)$. We recall that such a covering has a Lebesgue number $\lambda(\mathcal{U}) > 0$ with the property that if a set H on $\Phi(F)$ is of diameter $< \lambda(\mathcal{U})$ then H is contained in some set U.

Now upon carrying the subdivision process far enough we shall obtain sets $\overline{\mathcal{R}}_0$ all of diameter $< \lambda$, hence each contained, with its boundary Π_0 in a set U_0 of $\mathcal{U}$. Let t be the uniformizing parameter of U_0. In U_0 the function f is holomorphic in t. Hence by Cauchy's theorem

$$\int_{\Pi_0} f(t) = 0$$

and this implies the theorem.

In the preceding proof it has been implicitly shown that an integral

$$\int_{z_0}^{z} f(z)\,\mathrm{d}z$$

of a holomorphic function $f(z)$ at all points of a path μ in a region of holomorphism of the function is well defined.

Let $Q(z)$ be a rational function on the surface $\Phi(F)$: a function represented at all points of the curve by a *rational* function

$$S(x, y) = \frac{A(x, y)}{B(x, y)},$$

where locally the rational function is always represented by a convergent power series $t^k(\alpha + \beta t + \cdots)$, k a positive integer. The integral along *any* path of $\Phi(F)$

$$u = \int S(x, y)\,\mathrm{d}x$$

is then uniquely defined and represents a holomorphic function on the entire Riemann surface. Such an integral is said to be *of the first kind*. We refer to it briefly as (ifk).

Let γ be a closed path on $\Phi(F)$. The value

$$\int_\gamma \mathrm{d}u = \int_\gamma S(x, y)\,\mathrm{d}x = \omega$$

is then uniquely defined and called a *period* of u.

Going back to the model of a $4p$ sided polygonal plane region plus its boundary aba^{-1}, $b^{-1}cd\ldots$, set

$$\int_{a_\mu} \mathrm{d}u = \omega_\mu, \qquad \int_{b_\mu} \mathrm{d}u = \omega_{p+\mu}, \qquad \mu \leqslant p.$$

Take now two (ifk) u_1 and u_2 and define their periods as

$$\omega_{i\mu}, \omega_{p+\mu}, \qquad i = 1, 2.$$

Then we have this all important

THEOREM 2.4 (Theorem of Riemann)

$$\sum \begin{vmatrix} \omega_{1\mu}, & \omega_{1,p+\mu} \\ \omega_{2\mu}, & \omega_{2,p+\mu} \end{vmatrix} = 0.$$

PROOF. The proof is very simple.

Since u_1 and u_2 are holomorphic throughout $\Phi(F)$ we have

$$\int_\Pi u_1\,\mathrm{d}u_2 = 0.$$

This integral is the sum of p terms each of the same type as the sum

$$\int_a + \int_b + \int_{a^{-1}} + \int_{b^{-1}}. \tag{$*$}$$

The first and third term combine to

$$\int_a \big[u_1(P) - \big(u_1(P + \omega_{12})\big)\big]\,\mathrm{d}u_2 = -\omega_{21}\omega_{12}.$$

The second and fourth term combine like

$$\int_b \big[u_1(Q) - \big(u_1(Q) - \omega_{11}\big)\big]\,\mathrm{d}u_2 = \omega_{11}\omega_{22}.$$

Hence the sum $(*)$ is

$$\begin{vmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{vmatrix}.$$

The μ-th set of four terms has the sum

$$\begin{vmatrix} \omega_{1\mu} & \omega_{1,p+\mu} \\ \omega_{2\mu} & \omega_{2,p+\mu} \end{vmatrix}.$$

Hence

$$\int_{\Pi} u_1 \, \mathrm{d}u_2 = \sum_{\mu=i}^{p} \begin{vmatrix} \omega_{1\mu} & \omega_{1,p+\mu} \\ \omega_{2\mu} & \omega_{2,p+\mu} \end{vmatrix} = 0.$$

This proves Riemann's equality. □

Let now $u = u' + \mathrm{i}u''$ be a nonconstant (ifk) and let $\omega'_\mu, \omega''_\mu$, $0 < \mu \leqslant 2p$, be the respective periods of the *real* integrals u', u''. By Cauchy's inequality over $\Phi(F)$ we have

$$\int_{\Pi} u' \, \mathrm{d}u'' > 0.$$

Hence if we reason as before we obtain Riemann's inequality:

$$\sum \begin{vmatrix} \omega'_\mu & \omega'_{p+\mu} \\ \omega''_\mu & \omega''_{p+\mu} \end{vmatrix} > 0, \quad \mu \leqslant p. \tag{2.6}$$

CONSEQUENCE 2.5. *There are at most p linearly independent* (*ifk*) *mod constants.*

For if say there were $p+1$: $u_1, \dots, u_{p+1}$ then there would exist a linear combination

$$u = \lambda_1 u_1 + \cdots + \lambda_{p+1} u_{p+1}$$

such that every $\omega_\mu = 0$, $\mu \leqslant p$. That is $\omega'_\mu = \omega''_\mu = 0$, which contradicts the inequality.

THEOREM 2.6 (Digression). *There are exactly p linearly independent* (*ifk*) *modulo constants.*

That is one may find p linearly independent $\{\mathrm{d}u_h\}$, u_h is an (ifk) but no more.

PROOF. We have already seen that the number p' in question is $\leqslant p$. There remains to prove that $p' \geqslant p$.

There are two distinct approaches to this property:

(a) A proof by Riemann using highly complicated analytical properties of the well known theorem of existence of potential functions. See Hermann Weyl loc. cit.

(b) A proof of a more algebraic nature based upon a reduction of singularities theorem of much more geometric nature, due in part to Max Nöther (around 1870) which states:

THEOREM 2.7. *An irreducible plane curve F may always be birationally transformed into a plane curve G whose only singularities consist of a finite number of double points with distinct tangents.*

We have already seen that the genus p has birational character. It is also evident from the definition of the (ifk) that each of them has individual birational character. Hence the number p' of linearly independent (ifk) mod constants has the same character. Therefore to study their linear dependence we may freely replace the curve F by the curve G. That is we may assume that F has only the singularities just ascribed to G. This is the procedure that we shall follow.

Consider the most general curve of degree $m-3$ passing through all the double points: *adjoint* of degree $m-3$, also called *canonical* curve.

The curves of degree $m-3$ have $((m-1)(m-2))/2$ arbitrary coefficients. Those passing through the δ double points satisfy that many linear equations. Hence they have at least

$$\frac{(m-1)(m-2)}{2}-\delta$$

arbitrary coefficients. Now from an earlier Riemann formula

$$N=2(p+m-1)$$

since N is the class of F and it has only double points with distinct tangents

$$N=m(m-1)-2\delta. \tag{2.7}$$

Thus

$$p=\frac{(m-1)(m-2)}{2}-\delta.$$

This expression is actually the classical definition of the genus by Plücker.

Let us ask now for the *dimension* μ of the system of adjoints of degree $m-3$. Since they are merely the curves of degree $m-3$ through the double points

$$\mu\geqslant\frac{(m-1)(m-2)}{2}-\delta=p.$$

Now given such an adjoint Q_{m-3} we may write the integral

$$\int\frac{Q_{m-3}(x,y)\,\mathrm{d}x}{F'_y}$$

and we prove easily that it is an (ifk). It follows that the

$$\mu=p'\geqslant p.$$

But we have already proved that $\mu\leqslant p$. Hence $\mu=p$. This proves the Theorem 2.6. □

Let then $u_1,\dots,u_p$ be a system of p linearly independent (ifk). Form their period matrix

$$\Omega=[\omega_{j\nu}],\quad j=1,2,\dots,p,\ \nu=1,2,\dots,p.$$

Let $\eta = \lambda_1 u_1 + \cdots + \lambda_p u_p$, η_ν, $\eta_{p+\nu}$, $\nu \leqslant p$. Owing to Riemann's inequality the η cannot all be zero, whatever the choice of the λ's. Hence

$$[\omega_{j\mu}], \quad j, \ \mu \leqslant p$$

is of rank p. We may therefore apply a linear transformation such that this matrix becomes a unit matrix. That is

$$[\omega_{j\mu}] = \big[1, [\tau_{j\mu}]\big], \quad \mu \leqslant p.$$

The corresponding (ifk), written usually v_j are the *normal* (ifk).

Let

$$\tau = \tau' + \mathrm{i}\tau''.$$

From Riemann's equality and inequality we infer at once that:

(a) the matrix τ is symmetrical;

(b) it is the matrix of a positive definite quadratic form

$$\sum \tau_{jk} x_j x_k.$$

2.6. *Scorza's theory of Riemann matrices (1915)*

The preceding results have been strongly generalized and at Scorza's hand given rise to a very interesting new theory. We will say a few words about it.

The basic scheme of Scorza was *not* to take special bases for the cycles and the (ifk). We take then p linearly independent (ifk) and $2p$ independent one-cycles $\gamma_1, \ldots, \gamma_{2p}$ and write down their period matrix as a $p \times 2p$ matrix Ω_1. We then define

$$\Omega = \begin{bmatrix} \Omega_1 \\ \overline{\Omega}_1 \end{bmatrix}.$$

A more or less simple calculation shows then that the Riemann equality and inequality combined are equivalent to the existence of a unimodular skew symmetric matrix C ($|C| = 1$) such that $\mathrm{i}^{2p}\Omega' C \Omega = M$ is of the form

$$\mathrm{i}^{2p}\begin{bmatrix} 0 & A \\ A^* & 0 \end{bmatrix},$$

where A is a $p \times p$ matrix, $A^* = (\overline{A})'$, $|A| \neq 0$, so that M is a Hermitian positive definite matrix.

So far we only have a "clever" reformulation of Riemann. Scorza's departure is this:

DEFINITION 2.8. A Riemann matrix is a $p \times 2p$ matrix of type $\begin{bmatrix} \Omega_1 \\ \overline{\Omega}_1 \end{bmatrix}$ such that there exists a skew-symmetric *rational* matrix C such that

$$\Omega' C \Omega = \begin{bmatrix} 0 & A \\ A^* & 0 \end{bmatrix}.$$

No condition is placed on A. Whenever

$$\mathrm{i}^{2p} \begin{bmatrix} 0 & A \\ A^* & 0 \end{bmatrix} = M$$

Ω is said to be a *principal* matrix.

Given a Riemann matrix Ω there may be many matrices C which merely satisfy the definition (no Hermitian matrix condition imposed). The number k of linearly independent matrices C is the *singularity index* of Ω (Scorza had $1 + k$ where we have k, but the latter yields much simpler formulas).

Still another index h: *multiplication* index was introduced by Scorza, when the only condition imposed on C is that C need not be skew symmetric. Both indices have highly important applications in the theory of algebraic varieties.

3. Henri Poincaré and algebraic topology

3.1. *Poincaré: the founder of algebraic topology*

Presentday topology consists of two distinct parts: point set topology and algebraic topology. The first has mainly been the prerogative of Poland plus a strong American component: the school of R.L. Moore (of Austin, Texas). At all events, I shall only deal with algebraic topology.

The enormous impetus given by Poincaré to our field deserves to call him its founder. His contribution is contained in his paper Analysis Situs (1895) together with its five complements (till 1909), two on applications to algebraic surfaces.

Incidentally Poincaré did not say "topology" but "Analysis Situs", a beautiful but awkward term at best. Since the midtwenties "topology" has been generally adopted (much earlier I believe in Germany).

My purpose in this section is to develop Poincaré's basic concepts, but as seen by a modern: with algebra, especially group theory, in evidence. No doubt Poincaré himself, had he lived long enough, would have adopted this mode of exposition. Where simpler proofs than his have appeared, I do not hesitate to outline them. It must be said that simplifications have largely been due to the injection in topology of group theory by Emmy Noether (through Alexandroff).

In conformity with modern usage I generally omit the term "dimensional" and say: n-space, n-manifold, etc. for n-dimensional space, manifold, etc.

3.2. *Manifolds in the sense of Poincaré*

The whole of Poincaré's first Analysis Situs paper is devoted to manifolds. However, as is often the case with him, he is never too precise about what meaning he attaches to the term (in French, variété). I have therefore endeavored to extract a more precise meaning from his description.

Let $\mathfrak{E}_r$ denote a real Euclidean r-space referred to coordinates $x = (x_1, x_2, \ldots, x_r)$. By an *absolute* n-manifold $M_n \subset \mathfrak{E}_r$ $(n < r)$ I shall understand a compact, connected subset of $\mathfrak{E}_r$ without boundary, represented by the equations

$$f_p(x) = 0, \tag{3.1}$$

where the $f_p(x)$ are of class C^k, $k \geqslant 2$, and Jacobian rank $r-n$ in some bounded set $\supset M_n$. It is then known that any point ξ of M_n has in M_n a neighborhood $U(\xi)$ which is an n-cell differentiably parametrized by n local coordinates $u_1, \ldots, u_n$ with the condition that if two such neighborhoods say $U(\zeta)$, $U'(\zeta)$ overlap at ζ with respective parameters u_k, u'_k then each set is differentiable in the other with a Jacobian say

$$J = \frac{\partial(u)}{\partial(u')} \neq 0$$

and continuous at ζ.

Notice that compactness of M_n implies that it has a finite open covering $\{U(\zeta)\}$. If the Jacobians J have a fixed sign over M_n then M_n is *orientable*, otherwise *nonorientable*.

One may equally define M_n directly as possessing a finite open covering by parametric n-cells $\{U\}$ with the above overlapping property. This is the modern definition of "differentiable manifold". However, while Poincaré indicates its equivalence with the definition by the system (3.1), it is the latter upon which he always falls back.

I called "absolute" the manifolds just defined. This mention, however, will usually be omitted.

Suppose that M_n is orientable so that the Jacobians have a fixed sign. We may then orient M_n by choosing a given order of the parameters u_k in some $U(\zeta)$ and use that ordering, modulo an even permutation, as determining the Jacobian sign and hence the orientation of M_n. One refers then to $U(\zeta)$ as *indicatrix* of M_n.

EXAMPLE 3.1 (*Some examples of absolute* M_n). In $\mathfrak{E}_3$, a sphere, a torus, in $\mathfrak{E}_4$ a Riemann surface are examples of orientable M_n. On the other hand a projective plane is a nonorientable M_2.

Relative or open manifolds. In an M_n let M_p be a connected and compact subset contained in an open subset W of M_n. Thus W is a neighborhood of M_p in M_n. Set $V = U \cap M_p$. The collection $\{V\}$ is a finite open covering of M_p and $M_p \subset W$. The set $\overline{M}_p - M_p = \partial M_p$ is the *boundary* of M_p. We will assume that every point ζ of M_p has a neighborhood $V(\zeta)$ parametrized by p parameters $v_1, \ldots, v_p$ with the same overlapping property as for M_n. Orientability, indicatrices, etc. are defined as for M_n.

An additional hypothesis is

∂M_p consists of a finite set of closures of manifolds M^h_{p-1}.

Let $\zeta \in M^h_{p-1}$ and let $v_1, \ldots, v_{p-1}$ be local parameters for ζ on M^h_{p-1}. Since ζ is a point of a parametric p-cell of M_p, whose intersection with M^h_{p-1} contains a small parametric $(p-1)$-cell X_{p-1}, one may choose the parameters x_j of the latter so that together with one local parameter v of M_p at ζ, they make up a set of p local parameters of M_p at ζ. We shall use this property in a moment.

We refer to M_p as an *open* p-manifold.

EXAMPLE 3.2. Let S_3 be a sphere in $\mathfrak{E}_4$. A solid cube in S_3 is an open M_3. Here ∂M_3 consists of the surface of the cube. The faces of the cube are manifolds $M_2 \subset \partial M_3$. Together with the edges and vertices of the cube they make up ∂M_3.

3.3. *Boundary relations. Homologies*

The situation remaining the same write for the present M_{p-1} for M_{p-1}^h. Suppose that $\varepsilon_p(v, x_1, \ldots, x_{p-1})$ and $\varepsilon_{p-1}(x_1, \ldots, x_{p-1})$, ($\varepsilon_p$ and $\varepsilon_{p-1} = \pm 1$) represent indicatrices for M_p and M_{p-1}. The product $[M_p : M_{p-1}] = \varepsilon_p \varepsilon_{p-1} = \pm 1$ is the *incidence number* of M_p and M_{p-1}.

More generally if M_p^j and M_{p-1}^h are all oriented p- and $(p-1)$-manifolds in M_n then one defines the incidence number $[M_p^j : M_{p-1}^h]$ as 0 or ± 1 : 0 when M_{p-1}^h is not in ∂M_p^j, and ± 1 according to the preceding rule when M_{p-1}^h is in ∂M_p^j.

Call for the present (temporarily) *p-chain of* M_n a finite expression

$$c_p = \sum m_j M_p^j.$$

(The felicitous term "chain" is due to Alexander.) I define a chain-boundary ∂c_p under the rule

(a) $\partial M_p^j = \sum m_j [M_p^j : M_{p-1}^h] M_{p-1}^h$; $\partial M_0 = 0$ (M_0 is a point);

(b) $\partial c_p = \sum m_j \partial M_p^j$;

(c) if in the last sum M_{p-1}^j occurs with a total coefficient μ_h we define

$$\partial c_p = \sum \mu_h M_{p-1}^h.$$

Following Poincaré, if one is not interested in the special ∂c_p at the right then one expresses it by a *homology*

$$\sum \mu_h M_{p-1}^h \sim 0.$$

Such homologies do combine like linear equations. We also note:

DEFINITION 3.3. A chain c_p such that $\partial c_p = 0$ is called a *p-cycle*.

One proves that

∂M_p is a $(p-1)$-cycle; hence every ∂c_p is a $(p-1)$-cycle; boundary cycle.

In operator symbolism

$$\partial \partial = 0. \tag{3.2}$$

A set of p-cycles $\gamma_p^1, \ldots, \gamma_p^s$ is *independent* whenever they satisfy no homology. The maximum number of independent p-cycles is the p-th *Betti number* $R_p^d(M_n)$.

REMARKS 3.4. (I) R_p^d has no topological pretension since it depends strictly upon the differential structure of M_h. No such distinction was ever made by Poincaré.

(II) The notation $[:]$ is taken from Tucker's thesis (Princeton, 1931) and will be widely utilized later.

(III) The notion of *cobordism*, developed by Thom, and in full vogue nowadays finds its origin in the ideas of Poincaré.

(IV) Poincaré said "one or two sided (unilatère or bilatère)" where one says today "nonorientable or orientable", suggested by Alexander. His just criticism of Poincaré's terminology was that it referred really to a relationship with the ambient space, whereas orientability or nonorientability characterize an intrinsic property of the space (of the manifold M_n).

3.4. *Complex analytic manifolds*

These are the M_{2n} whose $2n$-cells are "complex analytic", that is parametrized by n complex variables $\{x_h \mid 1 \leqslant h \leqslant n\}$ with the condition that if $U(x)$ and $U(y)$ are two of the $2n$-cells overlapping at the point ζ then near ζ the complex variables y are holomorphic functions of the x.

Let $x_h = x'_h + ix''_h$. (The x', x'' are real.) Agree to orient $U(x)$ by naming the parameters in the order $(x'_1, x''_1, \ldots, x''_n)$. Then the Jacobians

$$\frac{\partial(x'_1, \ldots, x''_n)}{\partial(y'_1, \ldots, y''_n)}$$

are all *positive*. Hence *analytic manifolds are all orientable, and this in a unique manner.*

EXAMPLE 3.5. A nonsingular algebraic variety is always an orientable M_{2n}.

In M_{2n} the analytic manifolds M_{2p} are likewise oriented by the scheme just given. However, the arbitrary differentiable submanifolds have perfectly arbitrary orientations.

3.5. *Intersection of orientable manifolds*

Let M_p and M_{n-p} be orientable submanifolds of an orientable M_n. Let ξ be a common *isolated* intersection of the two submanifolds with parameters $\{x_h \mid 1 \leqslant h \leqslant p\}$ and $\{x'_j \mid 1 \leqslant j \leqslant n-p\}$, and such that $\{x_h; x'_j\}$ is a set of parameters for M_n at ξ.

Suppose now that $\varepsilon\{x_h\}$, $\varepsilon'\{x'_j\}$ and $\varepsilon_0\{x_h; x'_j\}$ all in their proper natural order, with $\varepsilon_0, \varepsilon, \varepsilon' = \pm 1$, are indicatrices of M_n, M_p, M_{n-p}. Then we assign to ξ the coefficient $\varepsilon_0\varepsilon\varepsilon' = \pm 1$ to be counted as *algebraic intersection* of M_p, M_{n-p} in M_n. Let ξ be described as a *simple intersection* of M_p and M_{n-p}.

Let M_p, M_{n-p} have only isolated intersections $\xi_1, \dots, \xi_s$ all simple with coefficient ε_h for ξ_h. By the *intersection number*, (M_p, M_{n-p}) is meant the sum

$$(M_p, M_{n-p}) = \sum \varepsilon_h. \tag{3.3}$$

Note that

$$(M_{n-p}, M_p) = (-1)^{p(n+1)}(M_p, M_{n-p}). \tag{3.4}$$

By approximations one may extend the meaning of (M_p, M_{n-p}) when ∂M_p and ∂M_{n-p} are disjoint. By a far from simple argument Poincaré proved:

THEOREM 3.6. *N.a.s.c. in order that $M_p[M_{n-p}] \sim 0$ is that $(M_p, M_{n-p}) = 0$ for every $M_{n-p}[M_p]$.*

REMARK 3.7. All the preceding results were obtained by Poincaré in his first paper Analysis Situs (§ 9). However, he had recourse to his first definition of a manifold together with a very subtle analytical argument.

The treatment which I have given is essentially parallel to that of chain intersections in a manifold, of my 1930 book on topology (LT), Chapter 4.

3.6. *Duality in manifolds*

Let now $\{M_p^h \mid 1 \leqslant h \leqslant R_p^d\}$ and $\{M_{n-p}^j \mid 1 \leqslant j \leqslant R_{n-p}^d\}$ be maximal independent sets relative to $\sim$ of M_p's and M_{n-p}'s of M_n. *Let ρ be the rank of the intersection matrix* $[(M_p^h, M_{n-p}^j)]$.

Applying Theorem 3.6 we find at once that $R_p^d = \rho = R_{n-p}^d$. This is the

THEOREM 3.8 (Duality theorem of Poincaré). *The Betti numbers $R_p^d(M_n)$ and $R_{n-p}^d(M_n)$ for an orientable M_n are equal.*

3.7. *Group of paths*

Let X be an arcwise connected metric space and let A be a given point of X. Let l be the directed segment $\alpha \leqslant x \leqslant \beta$, $\alpha < \beta$, and let ϕ map $l \to X$ so that $\phi(\alpha) = \phi(\beta) = A$. The image $\lambda = \phi(l)$ is a loop from A to A. Take the collection $\Lambda = \{\lambda\}$ with the following conventions: (a) if λ is homotopic to A in X write $\lambda = 1$; (b) λ described in the opposite sense is written λ^{-1}; (c) if ψ maps l in a second loop λ' then λ followed by λ' is a loop written $\lambda'\lambda$. Under these conventions Λ is a group $g(A)$. If B is a second point of X and $\mu = BA$ a directed arc from B to A the operations of $g(B)$ may be represented by $\{\mu^{-1}\lambda\mu\}$ where λ is any operation of g. Hence the groups $g(A)$ and $g(B)$ are *similar*. Upon identifying the operations λ and $\mu^{-1}\lambda\mu$, for all points $B \in X$, there results an abstract group $\pi(X)$, the *Poincaré group*, or *group of paths* of X. It is generally noncommutative. *It is also* (obviously) a *topological invariant of the space X*. In the ulterior investigations of Poincaré this group plays a very important role. For a reason to appear in a moment its general designation is $\pi_1(X)$, and it is also called first homotopy group of the space X.

3.8. *Homotopy groups and homotopy type of Hurewicz*

The group π_1 has been generalized (around 1935) in a very fortunate way. Let X, A be as before and let S_n be an n-sphere (generally $n > 1$) on which a certain point P is designated as fixed. Let ψ be a map $S_n \to X$ such that $\psi P = A$. The collection of the maps (ψ, A, P) may be made into a group, more or less as done by Poincaré for π_1. The only point, not obvious, is the mode of combination of these operations. Let me merely say that ψ_1 and ψ_2 are combined *additively*, as the combination is commutative, except for the Poincaré group π_1. The new groups are freed from dependence upon A and P and called n-th *homotopy groups of* X, written $\pi_n(X)$. This explains the π_1 designation for the Poincaré group.

Hurewicz groups have occupied a central position in modern algebraic topology. Although they are commutative, they do not have the rather simple properties of homology groups. This has greatly enhanced their importance.

Homotopy type. This is another noteworthy concept introduced by Hurewicz. Two topological spaces X, Y are of the same homotopy type whenever there exist mappings $\phi : X \to Y$ and $\psi : Y \to X$ such that $\psi\phi$ is homotopic to the identity as a mapping $X \to X$ and $\phi\psi$ is homotopic to the identity as a mapping $Y \to Y$. This is not quite homeomorphism, but the closest approach to it and assuredly much more elastic. This is why it has been in much favor among modern topologists.

3.9. *Examples*

In Analysis Situs Poincaré constructed eight examples of 3-spaces by matching appropriate faces of a cube (first four examples) or of a regular octahedron. His purpose was to obtain explicit 3-manifolds whose Betti numbers and groups π_1 could be computed. The second example is to be rejected as not corresponding to an M_3.

Of particular interest is his fifth example for his reason. Poincaré desired to settle the question whether Betti numbers alone were sufficient to characterize an M_n, $n > 2$. The examples in question enabled him to answer in the negative. For he obtained a whole family of 3-manifolds with the same Betti numbers but different groups π_1 and hence topologically distinct. In fact a careful study of these manifolds have produced $R_0 = R_3 = 1$, $R_1 = R_2 = \{1, 2, 3\}$ and yet there are an infinity of distinct groups π_1 (see Analysis Situs, p. 83).

3.10. *Complexes*

Soon after Poincaré's first Analysis Situs paper the Danish mathematician Heegard criticized his approach, more particularly for having missed torsion. In the Introduction to his first Complement Poincaré answered in part Heegard, but perhaps did not realize that his general "homology" description failed to cover a variety of cases. It was also clear to him that his general method was far from suitable for deriving, for example, his very general formula for the characteristic.

The upshot was that he introduced an entirely new approach to algebraic topology: the concept of *complex* and the highly elastic algebra going so naturally with it.

While Poincaré's complexes were formally only applied by him to manifolds, they have a far broader range. Moreover, his complexes were made up of quite general cells. It has been found more and more expedient to base everything on simplicial complexes, and their easy proofs.

Simplexes. Take $(n+1)$ linearly independent points (vectors) in $\mathfrak{E}_{n+p}$, $p \geqslant n$, say $A_0, A_1, \ldots, A_n$. The set of points

$$A = k_0 A_0 + k_1 A_1 + \cdots + k_n A_n; \quad 0 < k_n < 1, \ \sum k_h = 1,$$

constitutes an n-simplex σ_n. It is, and will always be, assumed oriented, by the order of naming the A_h, modulo an even permutation.

By replacing $n-p$ of the "< 1" by "$= 0$" one obtains a p-face σ_p of σ_n with a suitable orientation. Given σ_{n-1} let ε_n, $\varepsilon_{n-1} = \pm 1$ be such that $\varepsilon_n\{A_{i_0} \ldots A_{i_n}\} = \sigma_n$, and $\varepsilon_{n-1}\{A_{i_0} \ldots A_{i_{n-1}}\} = \sigma_{n-1}$. Then $\varepsilon_n \varepsilon_{n-1} = \pm 1 = [\sigma_n : \sigma_{n-1}]$ is the *incidence number* of the two simplexes.

Simplicial complex. $K = \{\sigma\}$ is a finite collection of disjoint simplexes such that if $\sigma \in K$ then every face of $\sigma \in K$.

For any σ_p and σ_{p-1} of K there is an incidence number $[\sigma_p : \sigma_{p-1}] = 0$ or ± 1, 0 when σ_{p-1} is not a face of σ_p, and ± 1 according to the above rule when σ_{p-1} is a face of σ_p.

I will now follow the modern treatment, rather than the very details contained in the second and third Complements. Let α_p denote the number of p-simplexes of K.

A p-chain is a linear integral expression

$$c_p = \sum m_h \sigma_p^h, \quad 1 \leqslant h \leqslant \alpha_p.$$

One defines a boundary $(p-1)$-chain of σ_p as

$$\partial \sigma_p = \sum [\sigma_p : \sigma_{p-1}^h] \sigma_{p-1}^h$$

and the boundary of c_p by linear extension as

$$\partial c_p = \sum m_h \partial \sigma_p^h.$$

It is then easily shown that

$$\partial\partial = 0. \tag{3.5}$$

A c_p with $\partial c_p = 0$ is a *p-cycle*. Hence:

Every ∂c_p is a $(p-1)$-cycle called a bounding cycle.

Evidently the collections C_p, Z_p, B_p of chains, cycles, bounding cycles are additive groups. Moreover

$$C_p \supset Z_p \supset B_p,$$

where each term is a subgroup of its predecessor.

From this follows that $H_p = Z_p / B_p$ is likewise an additive group: integral p-th homology group of K.

From a fundamental result of Frobenius, rediscovered by Poincaré (2nd Complement) we have:

THEOREM 3.9. *The group H_p has the following structures*:

$$H_p \simeq I_1 \oplus I_2 \oplus \cdots \oplus I_{R_p} \oplus T_p$$

where the I_h are infinite cyclic and T_p is finite. More precisely

$$T_p \simeq \Theta_1 \oplus \cdots \oplus \Theta_r$$

where the Θ_h are finite cyclic. If t_p^h is the order of Θ_h then t_p^h divides t_p^{h+1}.

The t_p^h are the *torsion* coefficients of Poincaré and R_p is the p-th *Betti number* of K.

By a fairly simple calculation one obtains the relation

$$R_p = \alpha_p - r_{p+1} - r_p \tag{3.6}$$

where r_h is the rank of the *incidence matrix*

$$\eta_p = \left[\sigma_p^j : \sigma_{p-1}^k\right].$$

Hence

THEOREM 3.10 (Theorem of Poincaré). *The characteristic $\chi(K) = \sum(-1)^p \alpha_p$ satisfies the relation*

$$\chi(K) = \sum(-1)^p R_p.$$

Barycentric subdivision. *A subdivision $K_1 = \{\zeta\}$* of K, with simplexes ζ, is defined by the condition that every $\sigma \in K$ is a union of ζ's. The barycentric type is particularly simple.

Let $n = \dim K$.

Let the *derived* of K to be defined, be denoted by K'. If $n = 0$ (a finite set of points) let $K' \equiv K$. If K has ν simplexes suppose that K' has been defined for $\nu - 1$.

Let σ be an n-simplex of K and let $K_1 = K - \sigma$. Thus K_1' is known. Call P the barycenter of σ. Join P by arcs to all the points of $(\partial\sigma)'$. Replacing σ by the resulting new simplexes, including P yields σ'. The orientations are defined by the condition that in $\overline{\sigma}$ they are determined so that

$$\partial(\sigma)' = (\partial\sigma)'.$$

Once K' is defined, one determines the derived sequence $K', K'', \ldots, K^{(n)}, \ldots$, by the condition $K^{(n+1)} = K^{(n)'}$. One proves then easily

$$\textit{Mesh}\, K^{(n)} \to 0 \textit{ with } 1/n. \tag{3.7}$$

This is the most important property of $\{K^{(n)}\}$.

Special case of manifolds. When K is an M_n one may construct a dual complex K_n^* which has the same Betti numbers and torsion coefficients as K_n itself but with complementary dimensions.

The construction of K_n^* is simple enough. Let K_n' be the first derived of K_n and let $\{\zeta\}$ be its simplexes. Given $\sigma_p \in K_n$ the simplexes ζ with a single vertex (centroid) of σ_p and all others exterior to σ_p make up an $(n-p)$-cell σ_{n-p}^* and $K_n^* = \{\sigma_{n-p}^*\}$.

The relation between K_n and K_n^* leads to Poincaré's famous duality relations

(a) *for Betti numbers*

$$R_p(M_n) = R_{n-p}(M_n) \tag{3.8}$$

(b) *for torsion numbers*

$$t_p^h = t_{n-p-1}^h$$

(for details see LT, Chapter 1).

REMARK 3.11. We recall again the origin of "homology". When two chains c_p, c_p' differed by a boundary ∂c^{p+1}, Poincaré wrote $c_p \sim c_p'$ or $c_p - c_p' \sim 0$. These relations, called homologies combined like linear relations. In other words they form groups: homology groups.

Various types of coefficients. While Poincaré only dealt with integral chains, cycles, etc., wide extensions were soon made to other types. I just mention: mod 2, Tietze; mod m (m prime) Alexander; rational coefficients, Lefschetz; (these are the same as Poincaré's: $\sim$ with division allowed: my later homologies $\approx$); any number system (real or complex) which is a field, Pontrjagin. These last led Pontrjagin to his famous duality: simultaneous in the complex and the coefficients.

3.11. *Subdivision invariance*

In the first Complement Poincaré dealt at length with subdivision and barycentric subdivision of a complex and proved that under them his Betti numbers, characteristic relation, torsion numbers, and for manifolds the manifold property and duality relations were subdivision invariant. He seems never to have attacked topological invariance.

Problems posed by Poincaré will be discussed at the end of Section 4.

4. Algebraic topology after Poincaré

4.1. *A touch of topological history*

After 1904 Poincaré turned his attention to some arduous problems suggested by his previous work. He attacked applications to algebraic geometry (see my note "A page of mathematical autobiography") and to dynamics, more particularly to the famous theorem of Poincaré–Birkhoff discussed below.

Three important events mark the period before 1910 and immediately after: (a) the introduction by Tietze (1909) of chain coefficients mod 2, the first departure beyond Poincaré; (b) the advent on the scene of the powerful figure of L.E.J. Brouwer the advocate par excellence of strict rigor. Curiously in his early years the Poincaré concepts played little role in his work; (c) the definition by Lebesgue of dimension for compact metric spaces. Finally the most salient features of the period before 1923 (I omit my own work on algebraic geometry) are the appearance of Oswald Veblen and J.W. Alexander at Princeton.

Beyond 1923 we find my extensive work on coincidences and fixed points together with their extensive and necessary ramifications; the related work of Hopf (Berlin); various contributions by Alexander notably on knots already mentioned; the contributions of Morse on critical points and applications to the calculus of variations; the research of Alexandroff (Moscow) on compact and dimension theory. This will take us more or less to 1930: roughly my intended terminal point.

4.2. *The Poincaré–Birkhoff theorem*

This is the last partly topological question that occupied Poincaré. In a long mémoire (Circolo di Palermo) he stated the theorem, exposed his unsuccessful endeavors to prove it and motivated his publication with the expressed hope that perhaps a younger man would be more successful. This hope was fulfilled with the solution of young Birkhoff which appeared in the Transactions (1912) soon after Poincaré's death! In a sense this marked the entrance of the US into the new world of topology.

The problem consists in this: – Let T be a topological mapping, area preserving, of a plane closed annular ring between two circles sending the two into opposite directions. To prove that T has at least one fixed point. Birkhoff's solution is not only brilliant but very short. It marks the beginning of his extensive work on celestial mechanics: his later research. Birkhoff not only proved the theorem but completed it by showing that if T is not area preserving then either there is a fixed point or else some Jordan curve in the ring surrounding the inner circle is mapped by T into its interior or else into its exterior. This is a strictly topological property – which is not the case for the theorem itself.

The initial theorem has many applications to dynamics, notably in the study of the various periodic solutions near one such solution.

4.3. *Henri Lebesgue and his definition of dimension*

Let X be a compact metric space and let $F = \{F_k\}$ be a finite closed covering of X. The *order* of F, written $\omega(F)$, is the least number of sets F_σ minus unity which have a common point. Lebesgue defines the dimension of X, written dim X, as the least order of F of mesh $< \varepsilon$ as $\varepsilon \to 0$ and this for all possible F. This is the first appearance of the concept of "order of a covering", found so useful later. This dimension was identified later with the Menger–Urysohn classic by Brouwer.

4.4. *The early work of L.E.J. Brouwer*

This work was done around 1910. One of his early contributions was a rather short proof of the Jordan curve theorem (the second accurate proof; the first was given several years earlier by Veblen). He also gave a proof of the invariance of regionality. That is if Ω_m, Ω_n are two Euclidean regions, with $m \neq n$, then they could not be homeomorphic. In more modern language assuming that Ω_m and Ω_n could correspond under an homeomorphism T_m of Ω_m and Ω_n, their local Betti numbers (defined later) would have to be equal. But those R_h of Ω_m are zero for $0 < h < m$, with $R_m = 1$; similarly for Ω_n: $R_h = 0$ for $0 < h < n$ and $R_n = 1$, which contradicts $m \neq n$.

The more striking result of Brouwer coming a lot closer to our topic is this:
– Let M_n, M'_n be two absolute orientable, manifolds. Let T be a mapping $M_n \to M'_n$. Assuming the two manifolds simplicial a suitable subdivision of M'_n has its n-simplexes covered the same algebraic number μ of times by images of those of M_n and μ is a topological invariant of the triple (M_n, M'_n, T). In terms of more modern topology the result is readily obtained. For if γ_n, γ'_n are the fundamental n-cycles of the manifolds: $T\gamma_n \sim \mu\gamma'_n$ in M'_n and μ is known to be a topological invariant. It is called degree of the mapping.

Noteworthy corollaries for mappings of spheres were obtained by Brouwer.

Many other striking topological results are due to Brouwer but we cannot deal with them here.

4.5. *Oswald Veblen as topologist*

He really began his work in the early part of the century. He was as much a rigorist as Brouwer, but operated first out of Chicago with E.H. Moore as mentor (under whom he took his doctorate). Moore was likewise given to full rigor, but less exclusively than the early Veblen. At any rate Veblen, perhaps under Moore's influence, or under the appearance of David Hilbert's (Göttingen): Über die Grundlagen der Geometrie, was early launched into geometry. For some years he studied polyhedra – source of his proof (first correct) of the Jordan curve theorem. He then launched into his major work: Projective geometry: 2 volumes, close to 1000 pages, first volume coauthored by J.W. Young. The second volume already shows leanings towards topology. This occupied him till about 1913. As a professor at Princeton he was fortunate to have as a disciple J.W. Alexander the outstanding topologist. Their collaboration led to a significant but short paper in the Annals of Mathematics of 1913 in which their aim – fully accomplished – was to present Poincaré's main ideas in Analysis Situs and complements, in strict rigorous manner. This led to Veblen's monograph "Analysis Situs" (Colloquium Lectures 1921; lectures given in 1916), which had the same objective as the short note, but with far more details. Noteworthy in it is a proof of the invariance of the homology groups for an n-complex (first for a 3-manifold is due to Alexander).

4.6. *J.W. Alexander as topologist*

As a mathematician and above all as topologist Alexander was distinguished by exceptional originality. At first he was attracted by the many questions left pending by Poincaré. Thus in 1915 when he was still a graduate student he gave the first proof of the topological invariance of the Betti numbers of a 3-manifold M_3. During a one-year visit to Italy he also showed that the algebraic invariant of Zeuthen–Segre had really topological character (proof by extension of the Riemann surface concept).

The invariance proof of the numbers R_3 introduced the first ideas of the future classical deformation theorem.

Thereupon World War I produced a 3-years interruption. In the early postwar period Alexander produced several noteworthy results. I mention particularly:

(a) Poincaré had already produced two orientable M_3's with equal Betti numbers but with different group of paths hence topologically distinct. Therefore homology groups were insufficient to characterize 3-manifolds. Alexander went further and produced a very simple example of two topologically distinct M_3's but with the same group of paths and therefore equal Betti numbers (and no torsion). Hence homology and group of paths identity were insufficient to distinguish two M_3's. The example is simple enough. Two solid tori (in $\mathfrak{E}_3$) could be identified at their bounding surfaces so that the characters just mentioned be the same. However, this could be done so that one obtains two topologically distinct M_3's (so-called lens spaces).

(b) The generalized Jordan curve theorem. One may presume a 2-sphere S_2 in $\mathfrak{E}_3$ has for (bounded) complement a 3-cell. Alexander gave an example where the complement has an infinite group of paths. On the other hand he proved also that if S_2 is analytical then the complement was effectively a 3-cell.

(c) A remarkable result of Alexander was his famous duality theorem (1922), the first beyond Poincaré. Given a complex immersed in an n-sphere S_n the Betti numbers satisfy:

$$R_p(S_n - K) = R_{n-p-1}(K) + \delta_{p0} - \delta_{p,n-1},$$

where the δ's are Kronecker indices: $\delta_{ii} = 1$, $\delta_{ij} = 0$ for $i \neq j$. As an application

$$R_0(S_n - K) = R_{n-1}(K) + 1$$

which expresses the number of components of $S_n - K$ (the number R_0) in terms of the Betti number $R_{n-1}(K)$.

Actually Alexander's result holds for any, say compact subset of S_n. It is also a special case of my, more general duality theorem proved several years later.

(d) From 1926 on Alexander dealt at considerable length with an improved organization of complexes and in particular obtained a new proof of the topological invariance of homology groups of a complex. Given two homeomorphic simplicial complexes K, K_1 he interprojected their derived sequences (using his deformation theorem). He then showed that the limit of the corresponding homology groups of the sequence is merely the corresponding ones of K, K_1 so that the two are the same. (It is actually not necessary to pass to the limit – one may show that

$$H(K) = H^{(p)}(K) = H^{(q)}(K_1) = H(K_1),$$

where H stands for homology groups of same type: equal dimension and same coefficient system.)

(e) *Singular Theory*. This scheme is actually implicit in part in Alexander's first topological invariance proof of a Betti number (for an M_3). I have since organized it into a highly elastic theory, which, together with a deformation theorem of Alexander (see LT) is applicable to a large number of topological invariance properties. I will say a few words about this singular theory.

Let X be an arcwise connected metric space and let ϕ map a rectilinear closed p simplex σ_p into X. The pair (ϕ, σ_p) is, by definition a *singular p-simplex* in X. One agrees, however, that if τ_p is another rectilinear p-simplex and f is a rectilinear homeomorphism $\tau_p \to \sigma_p$ then $(\phi f, \sigma_p) \equiv (\phi, \sigma_p)$.

Orientation of (ϕ, σ_p) is copied from that of σ_p. Hence if $\sigma_{p-1} \in \partial\sigma_p$ one defines $(\phi, \sigma_{p-1}) \in \partial(\phi, \sigma_p)$ with the same incidence number. Hence, if

$$c_p = \sum m_h(\phi_h, \sigma_p^h),$$

then

$$\partial c_p = \sum m_h \partial(\phi_h, \sigma_p^h).$$

The definitions of singular cycles, bounding cycles, homology groups is then automatic.

I merely mention that one may prove:

THEOREM 4.1. *The collection of singular p-cycles, is isomorphic with the special subcollection (identity, cycles of K).*

COROLLARY 4.2. *Since the singular cycle collection has obvious topological character this holds also for the homology groups of K.*

For Alexander's central contributions to knot theory see Section 1.

4.7. *Marston Morse: Critical point theory*

In the twenties and later Morse initiated his classical work based on the study of critical points of functions and applications, most particularly to the calculus of variations. The results for the period in question are developed in his Colloquium Lectures, vol. 18, 1934. The particular point of interest for us is Chapter 6. This volume contains also an extensive bibliography.

The results of Morse are far more general than what we describe, but it seems preferable for the short space at our disposal to lean more to clarity than to generality.

Let then $\mathbb{R}$ be a real (closed) bounded region on an analytical manifold referred to Euclidean coordinates $x_1, \ldots, x_n$. Consider also on $\mathbb{R}$ another analytic function $g(x) = g(x_1, \ldots, x_n)$, likewise analytic and such that on R: $a \leqslant g \leqslant b$, $a < b$. The *critical points* in $\mathbb{R}$ of g are its extreme points and the points where

$$\frac{\partial g}{\partial x_i} = 0, \quad 1 \leqslant i \leqslant n. \tag{4.1}$$

Assume that they are all isolated. Moreover, grant that at the critical points all the Hessians

$$\left|\frac{\partial^2 f}{\partial x_h \partial x_k}\right| \neq 0. \tag{4.2}$$

These are all simplifying assumptions, which Morse has abandoned.

Let $b_1, b_2, \ldots, (b_i = g(a_i))$ be the successive critical values. The problem dealt with by Morse is to find the variation of the Betti numbers of the region $b_1 \leqslant g < b_h$ with increasing h as g crosses b_h. This has been determined in terms of certain integers, the *type numbers* t_j which are defined in the following manner.

Corresponding to the critical point b_k the Hessian $H(b_k)$ determines a nondegenerate quadratic form

$$\phi(x) = \sum h_{jk} x_j x_k. \tag{4.3}$$

This quadratic form reduced to normal form has say m negative roots. The number t_h is the total number of critical points where (4.3) has m negative signs in its canonical form. As Morse showed the Betti numbers R_h of the region and the type numbers satisfy

$$\begin{aligned} t_0 &\geqslant R_0, \\ t_0 - t_1 &\leqslant R_0 - R_1, \\ \sum_{h=1}^{k} (-1)^h t_h &\geqslant (-1)^k \sum (-1)^h R_h, \\ \sum_{h=1}^{n} (-1)^h t_h &= \sum (-1)^h R_h. \end{aligned} \tag{4.4}$$

The last relation for $n = 2$ is due to Poincaré.

As given by Morse these relations were proved by him only for coefficients mod 2, but the proofs for integral coefficients or coefficients in a field is the same.

In his book Morse deals directly with the most general case but the proof for the simpler case is found in his paper.

In the same book Morse treats a great many applications, which cannot be discussed here as they usually involve a large amount of analytical technique, especially of the Calculus of Variations type. I merely mention by way of example:

(a) information about the *number* of normals to a variety V in Euclidean space from a point of the space; (b) the number of chords to V normal at both end points; (c) information about closed geodesics on V.

In all this research Morse rarely imposes analyticity and freely accepts C^1 or C^2 classes of functions. This of course adds considerably to the difficulties.

4.8. *The work of A.W. Tucker*

In his thesis (Princeton, 1932) Tucker algebraized the Poincaré scheme to the last degree, yet preserving a strong contact with algebraic topology. Briefly speaking, he considered a complex as a finite collection of unspecified elements and assigned "dimensions" from o to p. Let σ_q^h, $1 \leqslant h \leqslant \alpha_q$, be the q-elements. There were introduced incidence numbers $[\sigma^{q+1} : \sigma^q]$ under the sole condition that there takes place the general matrix relation

$$[[\sigma_{q+1}^h : \sigma_q^j]] \cdot [[\sigma_q^k : \sigma_{q-1}^l]] = 0.$$

One may define chains, their boundary relations, cycles and homology groups in the standard way. The boundary relation is given by

$$\partial \sigma_q^h = \sum [\sigma_q^h : \sigma_{q-1}^j] \sigma_{q-1}^j$$

and for a chain c_q by standard linear extension. This leads to the usual functional relation $\partial\partial = 0$. Betti and torsion numbers arise in the usual way. Briefly then the whole theory of complexes follows. The same holds for *manifolds* and their duality provided one specifies that every $\mathrm{St}\,\sigma$ has the homology groups of a point.

What attracted me most to Tucker's work is an extremely simple derivation of my fixed point formula (see the next chapter). Tucker's attack was not to be excelled for *single-valued* transformations. It did not seem to go over to multiple valued transformations. Here my early intersection method had the best of it.

4.9. *The work of Walter Mayer*

This author went to the extreme of abstraction. His first contribution was simply to take a finite sequence of additive groups: chain groups $G_0, G_1, \ldots, G_p$ with homomorphism $\tau_q : G_{q+1} \to G_q$ (boundary relations) satisfying $\tau_{q+1}\tau_q = 0$. One may then define the boundary subgroups $\tau_q G_{q+1} \subset G_q$, cycle subgroups, homology groups. I will not enter into a description save to say that Mayer's scheme has had quite a vogue later.

Another contribution of Mayer was most curious. Having defined the boundary operators – call them just τ for simplicity – he had the interesting idea of subjecting them to a relation $\tau^3 = 0$. It was proved later by Spanier (Michigan Thesis) that the resulting scheme was reducible to the standard type.

4.10. *Some open problems left by Poincaré*

One of the problems that evidently occupied Poincaré was to what extent the integral Betti numbers plus the group of paths π sufficed to characterize a manifold (let alone a complex). He actually showed by an example that a sphere S_3 and an M_3 with the same homology groups but with different π could be distinct. Furthermore, we have already observed that Alexander showed by an example that two M_3 with the same homology theory and same π need not be homeomorphic.

There the question has rested, except that nowadays one excepts much more, namely identity of all homotopy groups in addition to the identity of the homology groups. In fact whenever a new topological character is discovered one asks if it suffices to distinguish two given complexes. No such character has been discovered at the present time.

Let our discourse be limited to compact differentiable manifolds. An absolute M_n is *differentiable* whenever it admits a finite open covering $\mathcal{U} = \{U_h\}$ such that: (a) each U_h is parametrizable by variables x_{hj}, $1 \leqslant j \leqslant n$, (b) whenever U_h and U_j overlap say at a point P then about P the x_i^h are differentiable functions of the x_i^j with a nowhere zero Jacobian.

Now the question arises for a given M_n, with one system of differentials, is it unique? This has been answered in the negative, in 1956, by John Milnor, by exhibiting a 7-sphere S_7 with two *distinct* (unrelated) differential systems. This has been extended by Stallings and Smale up to S_5. Several authors have even computed for some of these spheres the exact number of "disjoint" differential systems.

One may also raise this question: given a polyhedron Π with *two* covering complexes K, K_1 (say simplicial), do they possess subdivision K^*, K_1^* with the same (algebro-geometric) structure? In 1960 Milnor gave an example which showed that in general this did not hold.

[Editor's note: these lectures have been only lightly edited: they give an idea of how one of the pioneers saw the early development of algebraic topology.]

This Page Intentionally Left Blank

CHAPTER 19

From Combinatorial Topology to Algebraic Topology

I.M. James

Mathematical Institute, University of Oxford, 24-29 St Giles, Oxford OX 1 3LB, UK

Topology, or analysis situs as it used to be called, is largely a creation of the twentieth century. The early history of the subject, up to end of the nineteenth century, has been carefully studied by J.-C. Pont [39] and others. However, twentieth century topology has not received the same attention as yet. It should be possible to investigate the development of the subject in the first half of the century, at least, in the same kind of way as has been done for the earlier period. The number of individuals involved is not great, the literature not extensive, and in many respects the subject was still at a pioneering stage. In the second half of the century, however, the subject has expanded so much and the pace of development has been so rapid that a different approach is called for. In this article I do not propose to say anything about this later period. And rather than attempt to discuss everything that happened in the earlier period I will describe what seem to me to be some of the most significant developments.

During the nineteenth century topology hardly existed as a separate discipline. After the last part of Poincaré's 'Analysis Situs' was published in 1905 it was clear that topology, as a subject, had arrived. Before long a few mathematicians started to specialize in this area of mathematics. Others, although their main research interests lay elsewhere, were not slow to recognize its importance. For example Hadamard, in a lecture at Columbia University in 1911, said

> 'Analysis situs is connected ... with every employment of integral calculus. It constitute(s) a revenge of geometry on analysis. Since Descartes, we have been accustomed to replace each geometric relation by a corresponding relation between numbers, and this has created a sort of predominance of analysis. Many mathematicians fancy that they escape that predominance and consider themselves as pure geometers in opposition to analysis; but most of them do so in a sense I cannot approve: they simply restrict themselves to treating exclusively by geometry questions which other geometers would treat, in general quite easily, by analytical means; they are of course, very frequently forced to choose their questions not according to their true scientific interest, but on account of the possibility of such treatment without intervention of analysis. I am even obliged to add that some of them have dealt with problems totally lacking any interest

HISTORY OF TOPOLOGY
Edited by I.M. James

whatever, this total lack of interest being the sole reason why such problems have been left aside by analysts.'

It is interesting to compare this with the following extract from an address (translated into English by Abe Shenitzer) given by Dehn at Frankfurt in 1928:

> 'As the last case we consider analysis situs or topology, the branch of mathematics that deals with the most general properties of the shape of a figure. It was developed only in the 19th century, and largely through the work of the Göttingen mathematician Riemann, who identified the topological core of many function-theoretic questions. At the end of the 19th century Henri Poincaré gave topology another strong impulse. At the present time there appear very many topological papers, but when it comes to fundamental problems we have hardly gone beyond Poincaré or, strictly speaking, Riemann – this in spite of the fact that such progress would be of great significance for, among other things, the theory of algebraic functions of two variables. Here the failure is not due to the fact that – as in number theory – the problems cannot be tackled, but to the fact that they are so intricate that the power of the human intellect, the ability to imagine different things at the same time, is not sufficient for mastery.'

In the early part of the twentieth century topology was regarded as having two main branches. The first, which was known as set theoretic analysis situs, treated spaces as sets of points. The other was known as combinatorial analysis situs, the term introduced by Dehn and Heegard in their article [24] of 1907 in the Enzyklopädie der Mathematischen Wissenschaften. Today the term 'combinatorial topology' is obsolescent; when it is used at all it seems to be thought of as equivalent to 'piecewise-linear topology'. In the early part of the century, however, the term 'combinatorial analysis situs' had a much broader meaning. Spaces were treated as being made up of cells, usually simplexes. The topology of the cells was regarded as well-understood; the interest lay in the way they were fitted together to form the space. Although this viewpoint was implicit in the work of Poincaré it was Dehn and Heegard who made it explicit. The story behind the writing of this important article is told in the biographies of the authors elsewhere in this volume.

In their article, Dehn and Heegard seem mainly concerned with trying to place some of the earlier work on a more satisfactory foundation. They divide their article into three parts: Complexes, Nexus, and Connexus. In the first they consider the intrinsic properties of complexes, as in the four colour problem. In the second they consider those which are invariant under certain transformations, as in the classification of surfaces. In the third they consider properties of a relative nature, for example, surfaces embedded in Euclidean space. They look back to Listing and Möbius; one feels that the full impact of Poincaré's work has not yet been felt.

However there are some significant pointers to the future. Notably they introduced the term 'homotopy'. Before long this came to mean simply 'continuous deformation', but originally it was used in a somewhat different sense, which still survived in Veblen's monograph 'Analysis Situs' [46] of 1922. In later volumes of the Enzyklopädie their article was superseded by one by Tietze and Vietoris [42], dated 1929. Tietze and Vietoris divided the subject into three branches, namely general topology, the theory of manifolds and combinatorial topology. By this time the term 'topology' has superseded the term 'analysis situs'. Let us see what had happened in the intervening period of over twenty years.

There was much that was unclear in Poincaré's mainly geometric reasoning, and its was some time before his often intuitive arguments had all been made secure. One of the first specialists in topology who started to build on Poincaré's researches was Tietze, who in a lengthy paper [14] of 1908 settled a number of outstanding questions. For example, he completed previous work on the relationship between the fundamental group of a space and what would soon be called the first homology group.

The fundamental group was studied intensively, in these early years, but otherwise it is difficult to find any results which might be regarded as belonging to homotopy theory until Brouwer, in the Netherlands, began to introduce entirely new ideas. In two miraculous years, between 1910 and 1912, he wrote an amazing series of papers, decisively advancing topology into a new era. His methods were unlike anything seen previously. Simplicial approximation, for example, was said to be akin to witchcraft. Although Brouwer himself was always more interested in the foundations, and published little on topology later, the influence of his innovative methods can scarcely be exaggerated. In contrast to that of Poincaré his expository style was exceedingly rigorous. He showed no interest in homology theory, going to great lengths to avoid using it, and yet Brouwer has a claim to be regarded as the founder of homotopy theory.

Naturally, other mathematicians who were studying topology were attracted to Blaricum, the village outside Amsterdam where Brouwer lived and held court. Those who came to work there at various times included Aleksandroff, Freudenthal, Hopf, Hurewicz, Reidemeister and Vietoris, each of whom made a great contribution to the development of the subject.

Meanwhile on the other side of the Atlantic research in topology was not slow to develop. At Austin, Texas, R.L. Moore was starting to build what became the American school of point-set topology. At Harvard the senior Birkhoff succeeded in settling the main problem posed by Poincaré in the final instalment of his Analysis Situs. Other American universities, especially Princeton, were also becoming known for research in the new subject.

Until early in the twentieth century, when Woodrow Wilson became President of the University, Princeton could not have been described as an important academic institution. However, in a determined effort to make it so, Wilson recruited some young men of exceptional quality, one of whom was Veblen. Veblen in turn recruited a strong team to the mathematics department, including some of his own former students, one of whom was Alexander. He also played an important part in the appointment of Lefschetz and other able mathematicians, and in the establishment of Fine Hall, where the facilities were of an unusually high standard.

While still only a graduate student, Alexander had seen how to use Brouwer' methods to establish the topological invariance of homology, another of the questions which Poincaré had left open. The story is told in detail by Dieudonné in I.3 of [25]; it might be added that Alexander's paper, which treated only the three-dimensional case, received only a perfunctory review from Blaschke in the Forschritte. Alexander used the singular approach, pioneered by Dehn and Heegard, but, unfortunately, he left much unclear. However, the essential point is that Alexander saw that the Hauptvermütung could be avoided by using a method which is essentially that by which the homotopy invariance of homology groups would be proved today. Later he and others clarified and generalized the argument, and produced alternative arguments, but it took almost thirty years before all the misconceptions were cleared up.

As well as the research papers which appeared in a number of journals the issues of the Jahresbericht in this period contain other articles which help to give a picture of how topology was seen in those days, such as the historical note by Feigel [27] and the semi-expository papers by Kneser [34] and by Van der Waerden [45]; the last of these contains a valuable bibliography. It is also interesting to read what was said at the 1932 International Mathematical Congress in Zurich [22], the first at which topologists had much of an opportunity to try to explain what they were doing. Alexander gave a lecture which included the following passage:

> 'Broadly speaking, we may say that analysis situs, or topology, deals with the properties of geometrical figures that remain invariant when the figures are subject to arbitrary continuous transformations. There are, however, several distinct kinds of analysis situs, because there are several distinct ways of interpreting the physical notion of continuity in mathematical language. For example, there is what we call point-theoretical analysis situs, which is different in spirit as well as in content from the sort of analysis situs originally proposed by Leibnitz. This branch of the science is essentially an outgrowth of function theory, whereas what Leibnitz had in mind was a new and independent type of mathematics, especially designed to avoid the complications of function theory and to deal directly with the purely qualitative aspects of geometrical problems. No doubt combinatorial analysis situs is more nearly a development of Leibnitz's original idea.'

> 'The vogue for point theoretical analysis situs seems to be due, in large part, to the predominating influence of analysis on mathematics in general. Nowadays we tend, almost automatically, to identify physical space with the space of three variables and to interpret physical continuity in the classical function theoretical manner. But the space of three real variables is not the only possible model of physical space, nor is it a satisfactory model for dealing with certain types of problems. Whenever we attack a topological problem by analytic methods it almost invariably happens that to the intrinsic difficulties of the problem, which we can hardly hope to avoid, there are added certain extraneous difficulties in no way connected with the problem itself, but apparently associated with the particular type of machinery used in dealing with it.'

Menger, speaking later in the same Congress, emphasized that while it was good for some purposes, other methods were necessary to obtain a proper understanding of the topology, even in the case of compact metric spaces. For a long time the combinatorial point of view was to predominate in Western Europe and the set theoretical in Eastern Europe.

Until the late twenties homology was always discussed in terms of Betti numbers and torsion coefficients. That was the case, for example, in Veblen's influential textbook, although he remarks that the abelianized fundamental group 'may well be called the homology group'. The first mention of Betti groups in print occurs in Vietoris' paper [15] of 1927. However, this important conceptual development seems to have occurred a little before this. Emmy Noether, early in 1925, gave a talk at a meeting of the Göttinger Mathematische Gesellschaft in which she showed how to replace the theory of elementary divisors for modules over the integers by the structure theorem for abelian groups. This is reported on page 104 of the 1926 Jahresbericht. Now Aleksandroff relates, in his autobiography [20]:

> 'In the middle of December (1925) Emmy Noether came to spend a month in Blaricum. This was a brilliant addition to the group of mathematicians around Brouwer. I remember a dinner at Brouwer's in her honour during which she explained the definition of the Betti groups of complexes, which spread around quickly and completely transformed the whole of topology.'

In his Memorial address of 1935 for Emmy Noether Aleksandroff gave a somewhat different version:

> 'In the summers of 1926 and 1927 she went to the courses on topology which Hopf and I gave at Göttingen. She rapidly became oriented in a field which was completely new to her, and she continually made observations, both deep and subtle. When in the course of our lectures she first became acquainted with a systematic construction of combinatorial topology, she immediately observed that it would be worthwhile to study directly the groups of algebraic complexes and cycles of a given polyhedron and the subgroup of the cycle group consisting of cycles homologous to zero. This observation now seems self evident. But in those years (1925–1928) this was a completely new point of view, which did not immediately encounter a sympathetic response on the part of many authoritative topologists. Hopf and I immediately adopted Emmy Noether's view in this matter, but for some time we were among the small number of mathematicians who shared this viewpoint. These days it would never occur to anyone to construct combinatorial topology in any way other than through the theory of abelian groups; it is thus all the more fitting that it was Emmy Noether who first had the idea of such a construction. At the same time she noticed how simple and transparent the proof of the Euler–Poincaré formula becomes if one makes systematic use of the concept of Betti group. Her remarks in this connection inspired Hopf completely to rework his original proof of the well known fixed point formula, discovered by Lefschetz in the case of manifolds and generalized by Hopf to the case of arbitrary polyhedra. Hopf's work 'Eine Verallgemeinerung der Euler–Poincaréschen Formel', published in Göttingen Nachrichten in 1928, bears the imprint of these remarks of Emmy Noether.'

The notion of degree, originally due to Kronecker, takes a particularly simple form in the case of maps of a sphere into itself, and Hopf had shown that such maps are classified by the degree. It was also known that every map of a sphere into a sphere of higher dimension was homotopic to a constant. Reputedly Lefschetz held the opinion that the same would be true in the case of a sphere of lesser dimension, as is certainly the case for maps into a circle. This conjecture, if that is what it amounted to, was demolished by Hopf in a paper which constitutes a landmark. In this paper [7] of 1930 in the Annalen he considered maps of a 3-sphere into a 2-sphere, the simplest case where the classification was unknown. He found a map, the celebrated Hopf map, for which the preimages of points were circles, and for which the preimages of distinct points were linked with coefficient unity. He showed that this linking coefficient could be defined for maps generally, and that homotopic maps have the same coefficient. In this way he was able to show that the number of homotopy classes is infinite. The history of this important paper is described by Samelson elsewhere in this volume. Hirosi Toda and George Whitehead, in their respective accounts [43, 48] of the first fifty years of homotopy theory, take Hopf's paper as their starting point.

Five years after it appeared another topologist, with a background in set-theoretic topology rather than geometry, was publishing the first fruits of research which would open up entirely new lines of investigation. This was Hurewicz, who had made his reputation in Vienna by completing the development of modern dimension theory initiated by Brouwer and Menger. When Menger left Vienna to join Brouwer in Amsterdam, Hurewicz followed him, and joined the group of topologists who had gathered around Brouwer. However, the master himself had turned away from topology long before, and was only publishing work on the foundations. It would be interesting to know more about what Hurewicz was working on in the years following his arrival in the Netherlands when there was a lengthy period during which he hardly published at all.

It is now sixty years since Hurewicz published his four famous Beiträge (usually rendered as research notes) in the Proceedings of the Royal Scientific Academy of Amsterdam [9]. These notes, which total 35 pages, contain many of the seminal ideas of homotopy theory. For example, Hurewicz introduced the concept of homotopy type, in which spaces are classified with respect to homotopy equivalence rather than topological equivalence. However, the notes are mainly concerned with the higher homotopy groups.

To quote Veblen's Analysis Situs [46] again

> 'Whether there exist generalizations of the fundamental group, and whether, in particular, these generalizations can be made in such a way as to bear a relation like the one just described to the n-dimensional Betti numbers and coefficients of torsion is a problem on which nothing has yet been published.'

The first public statement of the definition of these invariants was given by Čech at the 1932 International Congress in Zurich, although the idea had been around for some time. Čech showed that they were commutative, unlike the fundamental group, and in the discussion which followed his talk it seems to have been held by Aleksandroff and others that therefore they could not give any information which was not already given by homology. Discouraged by this Čech did not pursue the idea further, and it was left to Hurewicz to follow up his initiative and to establish some of their fundamental properties.

In these four research notes Hurewicz acknowledges the assistance of Freudenthal and Hopf right from the start, and it seems rather surprising that it is only in the second of the notes that he refers to Čech's talk, given at a session of the Congress presided over by Brouwer himself and attended, in all probability, by Aleksandroff and Hopf. However, Čech's definition was not the same as Hurewicz's and it may not have been immediately clear that they were equivalent.

In the notes Hurewicz not only introduced – or reintroduced – the higher homotopy groups but he opened up the possibility of calculating them through what became known as the Hurewicz theorem, which provides a fundamentally important connection between the homotopy groups and the homology groups. In the last of the notes he drew attention to the special properties of aspherical spaces, where all the higher homotopy groups vanish, and thereby initiated the discipline which became known as homological algebra.

By this time it was already clear that the problem of calculating the homotopy groups of spheres posed a special challenge. Rather than try and describe the subsequent development of homotopy theory generally I will now focus on this particular problem. First Hopf generalized and extended his earlier work in 1935 with another key paper [8] in which he introduced the concept of fibration. He described the families of fibrations where the fibres are spheres, the total spaces are spheres, and the base spaces are projective spaces. As special cases there are the fibrations where the base spaces are projective lines, therefore also spheres, and it is these which are usually called the Hopf fibrations. The map of the 3-sphere to the 2-sphere he considered in [7] is one of these special cases, as is the corresponding map of the 7-sphere to the 4-sphere and of the 15-sphere to the 8-sphere. Within the next few years the concept of fibration was further developed by Eckmann, by Ehresmann and Feldbau, and by Hurewicz and Steenrod, working more or less independently. Full details will be found elsewhere in this volume.

The special cases studied by Hopf turned out to be of exceptional interest. Using the same idea as in his 1930 paper Hopf assigned a numerical invariant, soon to be known as the Hopf invariant, to each map of a $(2n-1)$-sphere into an n-sphere. After a preliminary deformation, if necessary, the invariant is just the linking number of the preimages. For odd values of n Hopf showed that the Hopf invariant is always zero. For even values of n he constructed maps with any given even integer as Hopf invariant. He raised the question of whether, for particular values of n, there existed maps with odd Hopf invariant, and showed that this was indeed the case for $n = 1, 2, 4$ and 8, using the fibrations mentioned above. It was to be over twenty years before Hopf's question could be answered completely.

The next major advance in homotopy theory was due to Hopf's student Freudenthal, who in 1937 introduced [4] the concept of Einhängung or suspension. For a space X, the suspension SX is the join of X with a pair of points. A similar construction applies to maps and homotopies. In the case of a sphere the suspension is a sphere of one dimension higher. By this method Freudenthal constructed a homomorphism

$$\pi_{r+n}(S^n) \to \pi_{r+n+1}(S^{n+1}),$$

now called the Freudenthal suspension, and was able to show that this is injective for $r < n - 1$, surjective for $r < n$. The method he used to prove this was very geometric in character; it has been summarized in rather more modern language by Dieudonné [25]. In fact Freudenthal also showed that when $r = 2n$ the kernel of the suspension consists precisely of the elements of Hopf invariant zero. Freudenthal's paper is labelled 'I. Grosse Dimensionen'. Apparently the sequel was withdrawn after Hopf found an error in it, but unfortunately an announcement [5] had already appeared, in which Freudenthal asserted the existence of maps of Hopf invariant one from the $(2n - 1)$-sphere into the n-sphere for all even values of n. As the American mathematician George Whitehead showed not long afterwards [16] this is untrue, but it was not until 1958 that Adams succeeded in proving that the only values of n with this property were those found by Hopf.

The Freudenthal theorems constituted a major advance. They showed that if one continues to suspend, all the homotopy groups beyond a certain point are isomorphic, so that there is essentially just the one group to consider, the stable group of the r-stem. The groups before that stage is reached are called the non-stable groups. When $r = 0$, for example, the stable range is reached immediately and the stable group of the 0-stem is just the cyclic infinite group Z. When $r = 1$, however, the first group is 0, the second Z and the stable group $Z/2Z$. Maps of the 3-sphere into the 2-sphere suspend trivially if the Hopf invariant is even.

Techniques for calculating homotopy groups were still in their infancy. A new line of attack was developed by Pontryagin, involving what would nowadays be called the framed cobordism groups of smooth manifolds. Although the method is important its early applications were not free of error. Thus Pontryagin himself claimed the stable group of the 2-stem to be trivial; in fact it is $\mathbf{Z}/2\mathbf{Z}$, as shown later by Pontryagin and by G.W. Whitehead [17]. Also Pontryagin's student Rokhlin announced that the stable group of the 3-stem was $\mathbf{Z}/12\mathbf{Z}$. In fact the correct value is $\mathbf{Z}/24\mathbf{Z}$, as shown by Barratt and Paechter in Oxford, by Toda in Osaka and by Rokhlin himself, all about the same time (1952).

Let us now cross the Atlantic again and see what had been happening at Princeton during the thirties. Even before the foundation of the Institute for Advanced Study in 1930 Princeton was beginning to rival the best European universities as a centre for mathematical research, notably in topology. In 1932 Veblen migrated from the University to the Institute, where he was largely responsible for the selection of its early mathematics faculty. This included Einstein, Morse, von Neumann, and Weyl, who where later joined by Alexander and

Whitney, amongst others. Many of the mathematicians who visited the Institute came from Europe and not a few stayed on in the United States. The links between the Institute and the University were quite close, particularly where the visitors were concerned, and taken together the two institutions provided a centre of excellence in mathematics which outshone any others in the United States. As the international situation worsened and leading European universities such as Göttingen suffered an eclipse, Princeton went from strength to strength.

Shaun Wylie has kindly given me his recollections of Princeton in those days. After explaining that he went there in 1934 for three years on Henry Whitehead's advice he goes on:

> 'At that time the University Mathematics Department and the Institute for Advanced Study cohabited in Fine Hall, and I had only the vaguest idea of who belonged to what. The topological luminaries were Veblen and Alexander, of the Institute, and Lefschetz, of the University. Veblen had recently published his book 'Analysis Situs' but did not in my time lecture on Topology. Alexander at least twice announced a series of lectures and each time abandoned it fairly early on. There were, however, satisfying lectures from Lefschetz and from Tucker, and after tea there were one-off seminar presentations, generally by research students.
>
> Lefschetz's lectures were highly instructive. Of course he knew what mattered and what it was all about, but was bad at detail. There was a great deal of audience participation (which he was entirely happy with) and details were hammered out democratically. People learnt a lot. Lefschetz also contributed personally to audience participation at the seminars; he asked frequent questions, sometimes pretending not to understand and sometimes to illuminate.
>
> Tucker was quite different. His lectures were quietly elegant, and he communicated (at least to me) a relish for the subject. Lefschetz passed me on to him, and he shepherded me helpfully through my research.
>
> At that time homology was central, cohomology being mentioned as an algebraic stepsister. Lefschetz had established his duality theorem for manifolds, and Alexander his for polyhedra embedded in spheres; and Pontryagin's duality theorem was around. People knew about the fundamental group and covering spaces. The available homology theories for general spaces were singular homology and Čech homology; and there were rumours that Alexander was proposing something based on what he called gratings. I do not remember anyone being actively involved in Analytic Topology at Princeton.
>
> During my time there (1934–1937) the great new excitements were the cohomology product and Hurewicz's higher homotopy groups. Among the advances (but less seminal) was Reidemeister's combinatorial invariant.
>
> Princeton was a splendid place to be. Čech was there for a year and Hurewicz. Among the graduate students were Dowker and Steenrod and Wallman. Most people turned up for tea and were ready to talk; and the lecturers were highly available.'

In the textbooks of the thirties, such as that of Aleksandroff and Hopf [21], the exposition of homology theory is a mixture of algebraic and geometric arguments. In the next ten or twenty years it became accepted that it was better to separate out these two types of argument. The concept of chain complex was not new but it became standard practice to develop the homology theory of chain complexes before dealing with the geometry, of which there tended to be less and less. An essential ingredient in this process is the notion of exact sequence, another of the ideas of Hurewicz. This first appeared in an abstract [10] of 1941 in which he rather incidentally introduced the notion of exact sequence. Essentially the same notion appears in Eckmann's 1941 paper [2] on fibrations. The idea was taken up by Henri Cartan [1] in his set of axioms for the cohomology of locally compact spaces and by Eilenberg and Steenrod [3] who used it in their axiomatic approach to homology. It is interesting that, writing in 1945, they thought it necessary to comment 'At first glance this axiom may seem strange even to one familiar with homology theory'.

Hurewicz could not have realized that the idea of exactness would come to be seen as practically indispensable. In the abstract he only refers to the special case of the cohomology sequence for a pair of spaces, states the result without proof, and does not use the term 'exactness'. As a result the general application of this extremely useful notion was held up while those who wanted to use it waited in vain for a proper account to appear. In the end it was Kelley and Pitcher [12] who provided this, but it is not entirely clear who invented the term 'exact sequence' as distinct from the concept.

Of course set-theoretic topology was developing at the same time as combinatorial topology. For example, the theory of absolute neighbourhood retracts, described elsewhere in this volume, was in some ways a rival to combinatorial homotopy theory. In Poland the set-theoretic tradition was particularly strong, as the following extract from a note by Peter Hilton on a visit he made in 1955 well illustrates.

> 'Kuratowski had been very largely responsible for reviving Polish mathematics after the devastation wrought by the second World War, and his influence was immense. As a consequence, topology was one of the most active fields of mathematical research in Poland at that time. Borsuk was undoubtedly the leader of the Warsaw school of topology – and remained so until his death; generally speaking his influence was very positive indeed, since he was a wonderfully inventive mathematician with superb geometrical insight and intuition, but there was one surprising consequence of his dominance in the field, as I will explain.
>
> The structure of academic life and the traditions of Eastern Europe ensure that the influence of leading scholars in any field is very strong and sharply focused. Thus Borsuk's students (and Kuratowski's) would continue to work on problems within the domain of special interest to their teachers long after they had creased to have any formal relationships with them. Thus not only would their research areas continue to reflect their teachers' special concerns, but so would their methods. Now Borsuk was not comfortable with algebraic methods in topology; and his mathematical taste communicated itself unmistakeably to his students, his 'school'. Let me give two examples.
>
> Borsuk had, just before the outbreak of war, invented the *cohomotopy group* of a space X. Now at the time of my visit Borsuk had recently published a paper on his new idea of the dependence of maps. Among his results was one concerning the dependence of maps of an n-dimensional polyhedron K into S^n, in which the n-th cohomology group of K appeared in the statement of the criterion. I was very much interested by Borsuk's idea of the dependence of maps; and on studying his paper I was able to point out that this particular result could be extended to polyhedra K of dimension up to $2n - 2$, provided the role of the n-th cohomology group was replaced by that of the n-th cohomotopy group. 'Yes, you are probably right' said Borsuk 'but, unfortunately, I never really understood the cohomotopy groups'.
>
> My second example is closely related. Borsuk had asked whether, when two maps $f, g : X \to Y$ are dependent on each other, it must follow that f is homotopic to ug,

> for some self-equivalence u of Y. I was fortunate to find a counterexample during my stay in Warsaw. A year later the Rumanian topologist Tudor Ganea visited Warsaw and found an erstwhile student of Borsuk working on the quoted question of Borsuk. 'But surely', said Ganea, 'Hilton answered that question in the negative?' 'Yes, replied the Polish mathematician', but Hilton's answer was algebraic'.'

It should not be forgotten, however, that two of the leading figures in the development of algebraic topology, Eilenberg and Hurewicz, came from Poland and were originally set-theoreticians. Also algebraic topology has been strongly represented in Poland for many years now.

As we have seen, algebraic topology evolved from combinatorial topology during the late twenties. However, as far as I can discover the first appearance in print of the term 'algebraic topology' is not until the end of 1936 when Lefschetz, in an address which he gave at Duke University began:

> 'The assertion is often made of late that all mathematics is composed of algebra and topology. It is not so widely realized that the two subjects interpenetrate so that we have an algebraic topology as well as a topological algebra.'

Lefschetz went on to use the term as the title of his Colloquium volume of 1942 and today it is standard terminology. The term homotopy theory does not seem to have become accepted until after the second world war. Although the two terms are often used interchangeably, because the methods of homotopy theory tend to be algebraic in nature, there are parts of algebraic topology, such as fixed-point theory, where homotopy-theoretic methods are used but which are not part of homotopy theory itself. The term combinatorial topology seems obsolescent. Although combinatorial homotopy theory, as developed by J.H.C. Whitehead, is in widespread use the terminology is not.

When a subject is developing as rapidly as topology was during the first half of the twentieth century it is hardly surprising that at first there were only a few successful attempts to organize the material in the form of a textbook. Veblen's Analysis Situs of 1922, based on lectures he gave in 1916, has already been mentioned. This was the first to give an introduction to combinatorial topology, especially homology theory, and became a standard work. Kerékjártó's Vorlesungen über Topologie [33], which appeared the following year, was more concerned with the set-theoretic and geometric side of the theory. Lefschetz' Topology [36] of 1930 was to some extent an up-date of Veblen; his earlier Borel tract [35] of 1924 was more concerned with his own work. Reidemeister's Einführung in die Kombinatorische Topologie [40], which appeared in 1932, was mainly concerned with the fundamental group and covering spaces.

In the mid-thirties, however, two books were published which were of lasting importance. The first was Seifert and Threlfall's Lehrbuch der Topologie [41] of 1934, which gave an admirable account of the more geometric theory. The second, which appeared the following year, was the first (and only) volume of Alexandroff and Hopf's Topologie [21], which, after providing the student with all the relevant algebra and general topology, went on to give a beautiful account of homology theory. The introduction provides a valuable historical overview.

There were also several projected books on topology which, unfortunately, never appeared. One was the sequel to [33], which was intended to deal with higher-dimensional topology. For some years there was correspondence about this between Kerékjártó, Kneser and Reidemeister. In the end, after Kneser had done quite a lot of work on it, the project was abandoned. In the case of [21] the plan was for two more volumes, but neither of these appeared. For many years Hurewicz was preparing a textbook on homotopy theory, but the rapid development of the subject meant that there was always new material he felt he must include. The work seems to have ground to a halt soon after the end of the war, and the typescript perished in a fire not long after Hurewicz' death. Another abandoned project was the second volume of Eilenberg and Steenrod's Foundations of Algebraic Topology. This is described in some detail in the preface to the first (and only) volume.

For the history of algebraic topology up to (but not including) the time of Poincaré it would be difficult to improve on the monograph [39] of J.-C. Pont, now unfortunately out of print. The little-known article 'Topologie' [31] by Guy Hirsch provides a well-informed overview of the development of topology generally, including algebraic topology of course, until about twenty years ago. Several accounts have been published which describe the development of algebraic topology or, more specifically, homotopy theory, over a particular period of years, although they are in the nature of historical surveys, taking the reader through the literature but not adding much historical background. Thus both Hirosi Toda [43] (in Japanese) and George Whitehead [48] have treated the half-century 1930–1980, while Hans-Werner Henn and Dieter Puppe [30] extended their treatment to the century 1890–1990. To these accounts must be added Dieudonné's book 'A History of Algebraic and Differential Topology 1900–1960' [25] which contains a great deal of material, and his long article 'Une brève histoire de la topologie' [26], which in some ways serves as a summary of the book. In [37] Lefschetz gives a rather more personal view of the development of the subject up to 1935; part of this is reprinted in the present volume. There are in addition a number of articles on particular topics, such as MacLane's on the cohomology of groups [38]; the second section of the list of references includes a selection of these.

Bibliography

Primary literature

[1] H. Cartan, *Méthodes modernes en topologie algébrique*, Comment. Math. Helv. **18** (1945), 1–15.

[2] B. Eckmann, *Zur Homotopietheorie gefaserter Räume*, Comment. Math. Helv. **14** (1941–1942), 141–192.

[3] S. Eilenberg and N.E. Steenrod, *Axiomatic approach to homology theory*, Proc. Nat. Acad. Sci. USA **31** (1945), 117–120.

[4] H. Freudenthal, *Über die Klassen der Sphärenabbildungen*, Compos. Math. **5** (1937), 299–314.

[5] H. Freudenthal, *Neue Erweiterungs- und Ueberführungssatze*, Proc. Akad. Wessensch. Amsterdam **42** (1939), 139–140.

[6] H. Hopf, *Eine Verallgemeinerung der Euler–Poincaréschen Formel*, Göttingen Nachrichten (1928), 127–136.

[7] H. Hopf, *Über die Abbildungen der dreidimensionalen Sphäre auf die Kugelfläche*, Math. Ann. **104** (1931), 637–665.

[8] H. Hopf, *Über die Abbildungen von Sphären auf Sphären niedriger Dimension*, Fundam. Math. **25** (1935), 427–440.

[9] W. Hurewicz, *Beiträge zur Topologie der Deformationen*, Proc. Akad. Wetensch. Amsterdam; I. *Höherdimensionalen Homotopiegruppen* **38** (1935), 112–119; II. *Homotopie- und Homologiegruppen* **38** (1935), 521–528; III. *Klassen und Homologietypen von Abbildungen* **39** (1936), 117–126; IV. *Asphärische Räume* **39** (1936), 215–224.

[10] W. Hurewicz, *On duality theorems*, Bull. Amer. Math. Soc. **47** (1941), 562–563.

[11] W. Hurewicz and N.E. Steenrod, *Homotopy relations in fibre spaces*, Proc. Nat. Acad. Sci. USA **27** (1941), 60–64.

[12] J.L. Kelley and E. Pitcher, *Exact homomorphism sequences in homology theory*, Ann. of Math. **48** (1947), 682–709.
[13] L.S. Pontryagin, *Classification of continuous transformations of a complex into a sphere*, C. R. Doklady **19** (1938), 361–363.
[14] H. Tietze, *Über die topologischen Invarianten mehrdimensionaler Mannigfaltigkeiten*, Monatsh. für Math. und Phys. **19** (1908), 1–118.
[15] L. Vietoris, *Über den höheren Zusammenhang kompakter Räume und eine Klasse von zusammenhangstreue Abbildungen*, Math. Annalen **97** (1927), 454–472.
[16] G.W. Whitehead, *On the homotopy groups of spheres and rotation groups*, Ann. of Math. **43** (1942), 634–640.
[17] G.W. Whitehead, *The $(n+2)$-nd homotopy group of the n-sphere*, Ann. of Math. **52** (1950), 245–247.

Secondary literature

[18] P.S. Aleksandrov, *Die Topologie in und um Holland in den Jahren 1920–1930*, Nieuw Archief voor Wiskunde **17** (1969), 109–127.
[19] P.S. Aleksandrov, *Poincaré and topology*, Russian Math. Surveys **27** (1) (1972), 157–168.
[20] P.S. Aleksandrov, *Pages from an autobiography*, Russian Math. Surveys **34** (6) (1979), 267-302; **35** (3) (1980), 315–358.
[21] P.S. Aleksandrov and H. Hopf, *Topologie*, Springer, Berlin (1935).
[22] J.W. Alexander, *Verhandlungen des Internationalen Mathematiker Kongresses Zurich*, Saxer, ed., Fussli, Zurich and Leipzig (1932).
[23] M. Bollinger, *Geschichtliche Entwicklung des Homologiebegriffs*, Arch. Hist. Exact. Sci. **9** (1972), 84–170.
[24] M. Dehn and P. Heegaard, *Analysis Situs*, Enzyklopädie der mathematischen Wissenschaften III, AB 3, Teubner, Leipzig (1907).
[25] J.A. Dieudonné, *A History of Algebraic and Differential Topology 1900–1960*, Birkhäuser, Basel (1989).
[26] J.A. Dieudonné, *Une Brève Histoire de la Topologie*, Development of Mathematics 1900–1950, Pier, ed., Birkhäuser, Basel (1994).
[27] G. Feigl, *Geschichtliche Entwicklung der Topologie*, Jahresber. Deutscher Math. Vereinig. **37** (1928), 273–280.
[28] H. Freudenthal, *L'algébre topologique, en particulier les groupes topologiques et de Lie*, Actes XII Congres Internat. d'Histoire des Sciences, Paris (1982).
[29] H. Freudenthal, *Topologie in den Nederlanden: das erste Halbjahrhundert*, Nieuw Arch. Wiskunde III Ser. **26** (1978), 22–40.
[30] H.-W. Henn and D. Puppe, *Algebraische Topologie*, Ein Jahrhundert Mathematik 1890–1990, Deutsche Math. Vereinig. (1992), 673–716.
[31] G. Hirsch, *Topologie*, Abrégé d'Histoire des Mathematiques 1700–1900, Dieudonné, ed., Hermann, Paris (1978).
[32] H. Hopf, *Ein Abschnitt aus der Entwicklung der Topologie*, Jahresber. Deutsche Math. Vereinig. **68** (1966), 182–192.
[33] B. von Kerékjártó, *Vorlesungen über Topologie*, Springer, Berlin (1923).
[34] H. Kneser, *Die Topologie der Mannigfaltigkeiten*, Jahresber. Deutsche Math. Vereinig. **34** (1925), 1–14.
[35] S. Lefschetz, *L'Analysis Situs et la Géométrie Algébrique*, Gauthiers-Villars, Paris (1924).
[36] S. Lefschetz, *Topology*, Amer. Math. Soc., Providence, RI (1930).
[37] S. Lefschetz, *The early development of algebraic topology*, Bol. Soc. Bras. Matem. **1** (1970), 1–48.
[38] S. MacLane, *Origins of the cohomology of groups*, L'Enseignement Mathematique **24** (1978), 1–29.
[39] J.-C. Pont, *La Topologie Algébrique des Origines à Poincaré*, Presses Universitaire de France, Paris (1974).
[40] K. Reidemeister, *Einführung in die Kombinatorische Topologie*, Friedr. Vieweg & Sohn, Braunschweig (1932).
[41] H. Seifert and W. Threlfall, *Lehrbuch der Topologie*, Teubner, Leipzig (1934).
[42] H. Tietze and L. Vietoris, *Beziehungen zwischen den verschieden Zweiger der Topologie*, Enzyklopädie der Mathematischen Wissenschaften III, AB 13, Teubner, Leipzig (1914–1931), 141–237.
[43] H. Toda, *Fifty years of homotopy theory*, Iwanami-Sugaku **34** (1982), 520–582.
[44] R. Van den Eynde, *Historical evolution of the concept of homotopic paths*, Arch. Hist. Exact Sci. **45** (1992), 127–188.

[45] B.L. Van der Waerden, *Kombinatorische Topologie*, Jahresber. Deutsche Math. Vereinig. **39** (1929), 121–139.
[46] O. Veblen, *Analysis Situs*, Amer. Math. Soc. Coll. Publ. 5, New York (1922).
[47] H. Weyl, *Analysis situs combinatorio*, Rev. Math. Hisp. Amer. **5** (1923), 209–218, 241–248, 278–279; **6** (1924), 33–41.
[48] G.W. Whitehead, *Fifty years of homotopy theory*, Bull. Amer. Math. Soc. **8** (1983), 1–29.
[49] H. Whitney, *Moscow 1935: Topology Moving Toward America*, A Century of Mathematics in America, Duren, ed., Amer. Math. Soc., Providence, RI (1988).

This Page Intentionally Left Blank

CHAPTER 20

$\pi_3(S^2)$, H. Hopf, W.K. Clifford, F. Klein

H. Samelson*

Department of Mathematics, Stanford University, Stanford, CA 94305, USA
E-mail: samelson@gauss.stanford.edu

In 1931 there appeared the seminal paper [2] by Heinz Hopf, in which he showed that $\pi_3(S^2)$ (the third homotopy group of the two-sphere S^2) is nontrivial, or more specifically that it contains an element of order ∞. (His language was different. Homotopy groups had not been defined yet; E. Czech introduced them at the 1932 Congress in Zürich. Interestingly enough, both Paul Alexandroff and Hopf persuaded him not to continue with these groups. They had different reasons for considering them as not fruitful; the one because they are Abelian, and the other (if I remember right) because there is no mechanism, like chains say, to compute them. It was not until 1936 that W. Hurewicz rediscovered them and made them respectable by proving substantial theorems with and about them.)

There are two parts to the paper: The first one is the definition of what now is called the Hopf invariant and the proof of its homotopy invariance. The second consists in the presentation of an example of a map from S^3 to S^2 that has Hopf invariant 1 and thus represents an element of infinite order of $\pi_3(S^2)$; it is what is now called the Hopf fibration; the inverse images of the points of S^2 are great circles of S^3. Taking S^3 as the unit-sphere $|z_1|^2 + |z_2|^2 = 1$ in $\mathbb{C}^2$, these circles are the intersections of S^3 with the various complex lines through the origin.

Hopf knew the example from non-Euclidean Geometry and puzzled for years over the question whether it is an "essential" map, i.e. one that is not homotopic to 0, until (so he told me once) one day in 1927 or so, while he was walking along the Spree river in Berlin, the idea "Any two of these circles are linked in S^3" came to him; the rest is history.

This note is concerned only with the second part, the example – where did it come from? Hopf, on p. 655 of the paper (Selecta p. 53), calls it a Clifford parallel congruence, and in a footnote refers to p. 234 of Felix Klein's (posthumous) book [5] on non-Euclidean Geometry. On looking up the reference one finds the very brief statement that these "parallels" had been introduced by Clifford in a talk to the British Association in 1873. (It is not stated which British Association is meant, and there is no reference to any publication.) A little later in the book the Clifford parallels are described with the help of quaternions; and earlier in the book they had been introduced by geometric considerations as families of

*Support by NSF grant DMSD91-02078 is acknowledged.

HISTORY OF TOPOLOGY
Edited by I.M. James

lines in the three-dimensional space of non-Euclidean geometry with spherical metric, that allow a continuous family of motions shifting each line in itself.

In the preface to [5] one learns that Hopf played a substantial role in the preparation of the book, particular in the part that had to do with his own research. It seems quite clear that it was he who put in the reference to Clifford and the description of his parallels. But where did he get it from? First I looked at Clifford's collected works [1], but I could not find any paper that had to do with the parallels (but see below). I started looking through the literature, through earlier books on non-Euclidean Geometry and on Projective Geometry in general, to no avail. I looked through the references to Clifford in the Enzyclopaedie der Mathematischen Wissenschaften (there are very many and I could not check all of them, but they all seemed to refer to other things).

Finally Felix Klein came to the rescue, with two publications. One, [6], is a set of notes of a course on non-Euclidean Geometry that he had given in 1890. It is not exactly a book, although it has a hard cover. It is a dittoed ("autographed") copy of handwritten notes, carefully prepared by one of the listeners (this was Klein's way at that time of making his lectures available to the world at large). The other one is a paper, [7], which amplifies the last few pages of the lecture notes. In both Klein tells of a visit that he made to England in 1873. At that visit he went to a meeting of the British Association for the Advancement of Science in Bradford. He met the young Clifford there, listened to a talk by him, and discussed the content with him afterwards, together with R.S. Ball and others. Unfortunately, he says, the talk was never published; only the title of the talk was published in the Report on the meeting, as "A Surface of Zero Curvature and Finite Extent". Clifford had become interested in elliptic geometry and had found certain interesting congruences in elliptic space (projective 3-space $\mathbb{R}P^3$ with elliptic metric) or in the 3-sphere S^3 (a congruence is a 2-parameter family of disjoint projective lines (or great circles) that covers the space). Each of Clifford's congruences has the property that there exists a one-parameter group of rigid motions of the space that shifts each line of the congruence along itself. In fact there exist two such families of congruences, say "left" and "right". (Each line in space belongs to a left and also to a right congruence.) By taking a line in a left congruence and moving it along the lines of a right congruence by the one-parameter group of motions associated with the latter, Clifford constructed a surface whose induced metric is flat and which thus has Gaussian curvature 0 and which is of finite extent (read compact); it is in fact clearly a torus. This was the first example of what became known as Clifford–Klein space forms. (The name was introduced by Killing in [3], p. 257, [4], p. 314, to denote those space forms, i.e. spaces of constant curvature, that are different from the prime examples sphere and projective space [positive curvature], hyperbolic space [negative curvature], Euclidean space [curvature zero]; the latter are the cases distinguished by free mobility – the isometry group is transitive on the orthonormal frames. As Klein puts it in [7], p. 559, respectively, 367: Just try to turn Clifford's surface around one of its points.)

Klein derives all the formulae needed; he says that he does not know how Clifford proceeded. As pointed out by him, the best way to understand the congruences is probably with quaternions, for the three-sphere formed by the unit-quaternions: A left [respectively, right] congruence consists of the orbits on S^3 under left [respectively, right] multiplication by the elements of a one-parameter group $\cos t + \sin t \cdot u$ with any unit-quaternion u; in other words, the right [respectively, left] cosets of the subgroup. With $u = i$ this is precisely the Hopf fibration, and this must be where Hopf became acquainted with it.

Klein makes a point of saying how glad he was to be able to present these very interesting results of Clifford to the mathematical world, particularly since Clifford died a few years after their meeting prematurely; as noted, Clifford's talk was published by title only; there are only very brief indications of the matter in some of his papers ([1], items XX, XXVI, XLI, XLII, XLIV).

Thus one might wonder: Where would $\pi_3(S^2)$ be today, if Klein had not gone to the meeting of the BAAS in 1873 or if he had not listened to Clifford's talk?

As an appendix we reproduce, with I.M. James's permission, a letter from Hopf to Hans Freudenthal which throws some light on the timing of Hopf's result; the letter was communicated to James by W.T. van Est who has the original.

Princeton, N.J., 30 Murray Place, den 17. August 1928.

Lieber Herr Freudenthal!

Für den Fall, dass Sie sich noch für die Frage nach den Klassen der Abbildungen der 3-dimensionalen Kugel S^3 auf die 2-dimensionale Kugel S^2 interessieren, möchte ich Ihnen mitteilen, dass ich diese Frage jetzt beantworten kann: es existieren unendlich viele Klassen. Und zwar gibt es eine Klasseninvariante folgender Art: x, y seien Punkte der S^2; dann besteht bei hinreichend anständiger Approximation der gegebenen Abbildung die Originalmenge von x aus endlich vielen einfach geschlossenen, orientierten Polygonen $P_1, P_2, \ldots, P_a$ und ebenso die Originalmenge von y aus Polygonen $Q_1, Q_2, \ldots, Q_b$. Bezeichnet v_{ij} die Verschlingungszahl von P_i mit Q_j, so ist $\sum_{i,j} v_{ij} = \gamma$ unabhängig von x, y und von der Approximation und ändert sich nicht bei stetiger Änderung der Abbildung. Zu jedem γ gibt es Abbildungen. Ob es zu einem jeden γ nur eine Klasse gibt, weiss ich nicht. Wird nicht die ganze S^2 von der Bildmenge bedeckt, so ist $\gamma = 0$. Eine Folgerung davon ist dass man die Linienelemente auf einer S^2 nicht stetig in einen Punkt zusammenfegen kann.

Es bleiben noch eine Anzahl von Fragen offen, die mir interessant zu sein scheinen, besonders solche, die sich auf Vektorfelder auf der S^3 beziehen und mit analytischen Fragen zusammenhängen (Existenz geschlossener Integralkurven). Wenn Sie sich dafür interessieren, so schreiben Sie mir doch einmal. Meine Adresse ist bis 20. Mai die oben angegebene, im Juni und Juli: Göttingen, Mathematisches Institut der Universität, Weender Landstrasse.

Mit den besten Grüssen, auch an die übrigen Bekannten im Seminar,

Heinz Hopf.

Translation:
Princeton, N.J., 30 Murray Place, Aug 17 1928

Dear Mr. Freudenthal!

In case you are still interested in the question of the [homotopy] classes of maps of the 3-sphere S^3 onto the 2-sphere S^2 I want to tell you that I now can answer this question: there exist infinitely many classes. Namely there is a class invariant of the following kind: let x, y be points of S^2; then for a sufficiently decent approximation of the given map the counter image of x consists of finitely many simple closed oriented polygons $P_1, P_2, \ldots, P_a$ and likewise the counter image of y consists of polygons $Q_1, Q_2, \ldots, Q_b$. If v_{ij} denotes the linking number of P_i and Q_j, then $\sum_{i,j} v_{ij} = \gamma$ is independent of x, y and of the approximation and does not change under continuous change of the map. For every γ there exist maps. Whether to every γ there is only one map, I do not know. If the whole S^2 is not covered by the image, then γ is $= 0$. A consequence is that one cannot sweep the line elements on S^2 continuously into a point.

A number of questions that seem interesting to me remain open, in particular those that have to do with vector fields on S_3 and are related to analytic questions (existence of closed integral curves). If you are interested in this, then do write me. My address till May 20 is the one given above, in June and July: Göttingen, Mathematical Institute of the University, Weender Landstrasse.

With the best wishes, also to the other acquaintances in the seminar,

Heinz Hopf.

(Note: Freudenthal was Hopf's first student, in Berlin. Hopf told me once that Freudenthal was the "easiest" doctoral student he ever had. One day Freudenthal came to Hopf and said: "Dr. Hopf, I would like to have you as my thesis adviser. And here is my thesis.". It was Freudenthal's work on the ends of topological spaces and groups, in which he proved that a topological group (with suitable conditions, e.g., a connected Lie group) has at most two ends.)

Bibliography

[1] W.K. Clifford, *Mathematical Papers (1882)*, Chelsea, New York (1968).

[2] H. Hopf, *Über die Abbildungen der dreidimensionalen Sphäre auf die Kugelfläche*, Math. Annalen **104** (1931), 637–665; *Selecta Heinz Hopf*, Springer, Berlin (1964), 38–63.

[3] W. Killing, *Über die Clifford–Kleinschen Raumformen*, Math. Annalen **39** (1891), 257–278.

[4] W. Killing, *Einführung in die Grundlagen der Geometrie*, Part 4, Paderborn (1893).

[5] F. Klein, *Nicht-Euklidische Geometrie*, Springer, Berlin (1928).

[6] F. Klein, *Nicht-Euklidische Geometrie, I*, Vorlesung gehalten während des Wintersemesters 1889–1890, ausgearbeitet von Fr. Schilling, Göttingen (1893).

[7] F. Klein, *Zur Nicht-Euklidischen Geometrie*, Math. Annalen **37** (1890), 544–572; *Gesammelte Mathematische Abhandlungen I*, Springer, Berlin (1921), 353–383.

CHAPTER 21

A History of Cohomology Theory

William S. Massey

Department of Mathematics, Yale University, New Haven, CT 06520-8283, USA

1. Introduction

Today we take cohomology for granted and even teach it in a beginning course on algebraic topology. But this was not always so. In fact, it took approximately forty years after the introduction of homology theory in 1895 by Poincaré before cohomology theory appeared on the scene. Undoubtedly one reason for this was the fact that the early algebraic topologists were not much interested in homology groups *per se*. Rather, they seemed to be more interested in such things as the Betti numbers and torsion coefficients of finite complexes, their incidence matrices, etc. As long as this point of view held sway, there was not much point in introducing cohomology groups. There were, however, several precursors of cohomology before 1935. We will now consider some of these.

1.1. *The chains on a dual subdivision of a manifold*

Assume M^n is a closed, orientable n-manifold with a given triangulation. By barycentrically subdividing, and amalgamating these smaller simplices in a new way, it is possible to define what is called the *dual subdivision* of the original triangulation (the process is nicely described in Chapter X of Seifert and Threlfall [46]). This dual subdivision has the property that its k-cells can be put in 1–1 correspondence with the $(n - k)$-simplices of the original triangulation in such a way that the intersection number of an oriented k-cell and its dual oriented $(n - k)$-simplex is $+1$, while its intersection number with any other $(n-k)$-simplex is 0. Thus there is defined a pairing of the integral k-chains of the dual subdivision and the integral $(n - k)$-chains of the original triangulation to the additive group of integers. The boundary operator of the dual subdivision and the boundary operator of the original triangulation are adjoint linear operators with respect to this pairing (up to a plus or minus sign).

The modern reader will recognize that the k-chains of the dual subdivision play the same role as $(n-k)$-dimensional cochains of the original triangulations. This process was used to prove the version of the Poincaré duality theorem that was common before the introduction of cohomology groups; see Seifert and Threlfall [46], loc. cit. for details.

HISTORY OF TOPOLOGY
Edited by I.M. James

1.2. *Lefschetz's pseudo-cycles*

Unfortunately, the dual subdivisions just described exist only for orientable manifolds. To get something analogous for an arbitrary finite simplicial complex, Lefschetz in his 1930 book [37] introduced what he called "pseudo-cycles". Given a finite simplicial complex K, one can consider it as a subcomplex of a triangulated n-sphere S^n for sufficiently large values of n. One can then consider the dual subdivision of S^n, as described above. The pseudocycles are certain cycles on this dual subdivision in a certain neighborhood of K in S^n, the details can be found in [37].

Although these pseudocycles are frequently referred to as forerunners of the notion of cocycles, they suffer several obvious disadvantages. Most importantly, they are cycles on another space which is not uniquely associated with the given complex K.

1.3. *Intersection theory of cycles in a manifold*

This theory, which was introduced by Lefschetz and Alexander in the middle 1920's, is a precursor of cup products in cohomology. The basic idea is very simple and has great appeal to one's geometric intuition. In an oriented n-manifold, two cycles, of dimensions p and q, respectively, should intersect in a cycle of dimension $p + q - n$, provided they are in "general" position and the intersection is non-empty. If they are not in general position, one should be able to replace them by a pair of homologous cycles which are in general position. Finally, this operation should lead to a multiplication of homology classes,

$$H_p(M^n) \times H_q(M^n) \to H_{p+q-n}(M^n),$$

which is a topological invariant of the given orientable manifold M^n. The multiplication of homology classes thus defined is associative and commutative (up to a plus or minus sign). This program was actually carried out in more or less detail by the topologists of the era from about 1925 to 1940; but the details are long and tedious. See, for example, [26, 37].

1.4. *De Rham's theorem (see [17])*

This famous theorem is usually stated today in terms of cohomology: The cohomology groups of a smooth manifold (with real coefficients) may be computed by using exterior differential forms as cochains. But cohomology groups had not yet been defined in 1931 when De Rham's paper was published, so he was forced to state his theorem in terms of homology and the integration of differential forms over smooth chains. When De Rham wrote this paper, cohomology groups defined using exterior differential forms were practically staring him in the face, and he could have gone ahead and made the definition with very little additional effort. But this probably seemed pointless to him at the time, given the state of algebraic topology in 1931.

In a rather brief paper [18] published in 1932, De Rham outlined a proof that the product of closed differential forms gives rise to the same information about a manifold as the intersection theory of cycles. Re-interpreted from a modern point of view, he showed that the product of cohomology classes determined by the product of closed differential forms corresponds under Poincaré duality to the product of homology classes determined by the intersection of cycles.

Apparently the first modern statement and proof of De Rham's theorem (in terms of cohomology) was in mimeographed notes of lectures by H. Cartan [12, 13]. According to André Weil, [67] he communicated this proof to Cartan in 1947.

1.5. *Hopf's umkehrunghomomorphismus*

Simple examples show that if $f : M_1 \to M_2$ is a continuous map from one orientable manifold to another, then the induced homomorphism

$$f_* : H_*(M_1) \to H_*(M_2)$$

cannot preserve the multiplication of homology classes defined by intersection theory. This statement is true even if M_1 and M_2 are of the same dimension. If M_1 and M_2 are of different dimensions, then the intersection of a p-dimensional cycle and a q-dimensional cycle will obviously have different dimensions in M_1 and M_2.

This deficiency was repaired by Hopf in a rather novel way in 1930 [27]. Assuming that the manifolds M_1 and M_2 are of the same dimension, Hopf showed how to associate with the continuous map $f : M_1 \to M_2$ a homomorphism

$$\varphi : H_p(M_2) \to H_p(M_1), \quad p \geqslant 0,$$

going in the *opposite* direction; hence the name, which means "reverse homomorphism". The homomorphism φ *does* preserve intersection products, i.e.

$$\varphi(u \cdot v) = (\varphi u) \cdot (\varphi v),$$

for any $u \in H_p(M_2)$ and $v \in H_q(M_2)$. This reverse homomorphism φ is a precursor of the homomorphism induced by the continuous map f on cohomology groups.

In modern terms, Hopf's umkehrhomomorphismus corresponds under Poincaré duality to the homomorphism f^* induced by f on cohomology; and intersection of homology classes corresponds to cup product of cohomology classes.

1.6. *The struggle to find more general and natural statements of the duality theorems of Poincaré and Alexander*

Today we usually state these duality theorems by saying that certain homology and cohomology groups are isomorphic. Before 1930 these theorems were usually stated as equalities between certain Betti numbers and torsion coefficients. For example, in Seifert and Threlfall [46], the Poincaré duality theorem is stated as follows: the kth Betti number of a closed, orientable n-manifold is equal to the $(n - k)$th Betti number; the torsion coefficients of dimension k are equal to those of dimension $n - k - 1$. Once the group theoretic point of view came to dominate homology theory, there must have been a natural impulse to try to reformulate these duality theorems as some kind of relation between various homology groups. This eventually culminated in two papers published by L. Pontrjagin in the *Annals of Mathematics* in 1934. In the first paper [40] he stated and proved his famous duality theorem for the case of compact groups and discrete groups. Then in the second paper [41] he applied this to give the desired general statement of the Alexander duality theorem. The first step in this general statement is to define for any compact space X and any compact topological group G the appropriate homology groups $H_k(X; G)$. These are the homology groups in the sense of Vietoris or Čech; they are compact topological groups. Let $\widehat{G}$ denote the character group of G; it is a discrete abelian group. Then the desired general statement of the Alexander duality theorem is the following: For any compact subset X of the n-sphere S^n, the (reduced) homology groups

$$\widetilde{H}_k(X; G) \quad \text{and} \quad \widetilde{H}_{n-k-1}\big(S^n - X; \widehat{G}\big)$$

are the character groups of each other. The product of an element $u \in \widetilde{H}_k(X; G)$ and $v \in \widetilde{H}_{n-k-1}(S^n - X; \widehat{G})$ is the "linking coefficient" of these two homology classes.

Similarly, the Poincaré duality theorem can be stated as follows: For any closed, orientable n-manifold M, the homology groups

$$H_k(M; G) \quad \text{and} \quad H_{n-k}\big(M; \widehat{G}\big)$$

are character groups of each other. The product between a homology class of $H_k(M; G)$ and one of $H_{n-k}(M; \widehat{G})$ is the "intersection coefficient" of these two homology classes.

As we will see later, the cohomology group $H^k(X; \widehat{G})$ is the character group of the homology group $H_k(X; G)$. Thus the introduction of cohomology groups made it unnecessary to consider homology groups with compact coefficients. However topologists were so impressed by Pontrjagin's results that for several years after the introduction of cohomology theory, homology with compact coefficients still appeared in various books and papers. For a well-known example of this, see [24, Chapter IX] (the fact that homology groups with compact coefficients are superfluous was pointed out by H. Cartan in his review of this book in *Mathematical Reviews*).

In the remainder of this chapter we will discuss the development of cohomology theory, including cup products and primary cohomology operations. We will *not* discuss sheaf theoretic cohomology, cohomology with local coefficients, spectral sequences, or extraordinary cohomology theories, such as K-theory. We will try to use modern terminology and notation throughout, for the benefit of the reader; the terminology and notation in the original papers were often quite different.

In a short chapter such as this it is impossible to take notice of every paper on cohomology theory published during the period under consideration. We hope that the authors of papers which are not discussed will understand our reasons.

The reader should note that biographies of many of the mathematicians discussed here are provided in other chapters of this book.

The author acknowledges with gratitude the kind assistance of Paul Lukasiewicz, librarian of the Yale Mathematics Library, in helping him find various references, etc.

2. The first papers on cohomology theory. The 1935 Moscow conference

An International Conference on Topology was held in Moscow, September 4–10, 1935. One of the American participants was Hassler Whitney, then only three years past his Ph.D. degree. Fifty three years later he published his rather vivid reminiscences of this conference [75]. Seldom in the history of mathematics has a conference occurred at such a propitious time, or marked the initiation of so many new basic lines of research. Hurewicz introduced his homotopy groups and described some applications. Hopf and Whitney lectured about vector fields and sphere bundles, thus starting in motion the study of fibre bundles. And Alexander and Kolmogoroff independently introduced cohomology theory, along with cup products. The official proceedings of the conference were published in *Math. Sbornik*, Vol. 43 (1936), pp. 619–793.

According to Whitney, Kolmogoroff spoke before Alexander; he described his theory of products in the cohomology of a simplicial complex. When he had finished, Alexander announced that he also had essentially the same results; both had papers in press. Each of them published two papers, and their papers were amazingly similar. Indeed, if it were not for the fact that all the evidence indicates otherwise, the casual reader of these papers would be inclined to suspect that they were written in collaboration!

Alexander published his two papers in Vol. 21 (1935) of the *Proceedings of the National Academy of Sciences* (see [4, 5]). These papers were submitted to the *Proceedings* before the Moscow conference, on July 8, 1935. Each is a very brief announcement of results, only two pages long. Kolmogoroff's two papers were published in Vol. 43 (1936) of *Math. Sbornik* (see [31, 32]). No dates of submission are given; however in a footnote to the first paper, Kolmogoroff says he reported some of his results at an international conference on tensor analysis in May, 1934. Kolmogoroff's paper are over twice as long as Alexander's, 6 pages and 5 pages, respectively, but still they are essentially announcements of results, with few detailed proofs.

The first paper of each author is concerned with what we would call today "finite cell complexes". It is assumed that the reader is familiar with such concepts as the chains of a cell complex, the boundary operator, homology groups, etc. In each case, the author then "dualizes" this procedure, to describe what we would today call cochains, the coboundary operator, and cohomology groups. Thus the major point of each of these two papers was concerned with what are essentially algebraic formalisms. Both authors pointed out that for any finite complex K, and any compact abelian group G, the homology group $H_r(K; G)$ and the cohomology group $H^r(K; \widehat{G})$ are character groups of each other. Indeed, in view of the above mentioned results of Pontrjagin on duality theorems, this fact must have been one of the main motivations leading these authors to define cohomology groups. Kolmogoroff also includes in this paper statements of the Poincaré and Alexander duality theorems using both homology and cohomology groups, such as is common today.

While these first papers were concerned with defining cohomology groups for finite cell complexes, these authors' second papers were more ambitious. It was the authors' aim to define cohomology for general spaces, and to introduce cup products into cohomology theory.

When faced with the question of defining cohomology groups for general spaces, most topologists today would instantly think of singular cohomology. In 1935 this was not a likely option, because singular homology theory as such was not well developed. The singular simplexes which were used then included certain degenerate singular simplexes which gave rise to elements of order two in the singular chain groups. Thus the singular chain groups were not free abelian groups. We will come back to this question later when we discuss singular cohomology theory.

Another option might have been to dualize the Čech homology theory, thus defining Čech cohomology theory. For some reason, neither Alexander nor Kolmogoroff chose to do this. Perhaps they were not sufficiently familiar with the properties of direct and inverse limits (which then were relatively new) to carry out the dualization. Instead, they chose to dualize the Vietoris homology theory.

First we will discuss Alexander's second paper [5]. Like Vietoris, Alexander limited his considerations to compact metric spaces. On such a space C, he defined a p-cochain to be a skew-symmetric function $f(x_0, \ldots, x_p)$ of $p+1$ points in C with values in a given abelian coefficient group A. Such a function is said to be *locally zero* if there exists a number $\varepsilon > 0$ such that f vanishes whenever the distances between any two of the points $x_0, \ldots, x_p$ is less than ε. Presumably Alexander wanted to factor out the subgroup of all p-cochains which are locally zero, but he was not quite precise about this. It is at this stage that the topology of C enters the picture. The coboundary of a p-cochain is defined as a multiple of the usual coboundary

$$\delta f(x_0, \ldots, x_{p+1}) = \sum (-1)^i f(x_0, \ldots, x_{i-1}, x_{i+1}, \ldots, x_{p+1}).$$

With this definition, one can define cocycles, coboundaries, and the cohomology group $H^p(C; A)$ as usual. Alexander asserts that if A is the character group of the compact abelian group B, then $H^p(C; A)$ is the character group of the Vietoris homology group $H_p(C; B)$.

Next, he wishes to define a product in this cohomology theory. Therefore he assumes that the coefficient group A is a ring and defines the product of a p-cochain f and a q-cochain g to be a $(p+q+1)$-cochain defined by the following peculiar formula:

$$(f \times g)(x_0, \ldots, x_{p+q+1}) = \frac{1}{(p+1)!(q+1)!} \sum (-1)^{N(\alpha)} f(x_{\alpha(0)}, \ldots, x_{\alpha(p)}) \times g(x_{\alpha(p+1)}, \ldots, x_{\alpha(p+q+1)}),$$

where the sum is over all permutations α of the integers from 0 to $p+q+1$, and $N(\alpha)$ is 0 or 1 according as the permutation is even or odd.

The reader will immediately recognize that there are several things wrong with this formula. First of all, the factor in front of the summation sign requires that A should be an algebra over the rational numbers, which was not assumed initially. Secondly, the dimension of the product is wrong, it is $p+q+1$ rather than $p+q$. Alexander goes on to assert that the product of cochains is commutative, up to a $\pm$ sign; this would require that the ring A should be commutative, which is not mentioned. He gives the following coboundary formula:

$$\delta(f \times g) = (\delta f) \times g = \pm f \times (\delta g)$$

from which he concludes that the cocycles constitute an ideal in the ring of cochains, and the coboundaries form an ideal in the ring of cocycles. Thus the product of cohomology classes is well-defined. Alexander goes on to assert that the cohomology ring thus defined is a stronger invariant than the cohomology groups alone. However, he gives no examples to show this.

In a footnote to his second paper [32], Kolmogoroff stated that Alexander erroneously assumed that the product of cocycles can be non-zero! In another footnote, he says that at the Moscow conference he learned that Alexander had found another product similar to the one he (Kolmogoroff) has in his paper. In footnote number 7 of [6], Alexander states that "The definition of the product announced by the author in the second of the two *Proceedings* notes ... is not the significant one, as was noticed by him while the note was in press. His revised definition was equivalent to Kolmogoroff's."

This second paper of Alexander appears to have been rather hastily written, with little attention to detail. In addition to the difficulties with the product he defined, there are other obscurities and/or errors which we have not cited.

In Kolmogoroff's second paper [32], he first considers finite simplicial complexes. He is concerned with introducing products in the cohomology *with rational coefficients* of such a complex. First he defined the following product of a p-cochain f and a q-cochain g:

$$(f \cdot g)(v_0, \ldots, v_{(p+q+1)}) = \frac{1}{2} \sum f(v_{\alpha(0)}, \ldots, v_{\alpha(p)}) \cdot g(v_{\alpha(p+1)}, \ldots, v_{\alpha(p+q+1)}),$$

where the summation is over all permutations α of the integers from 0 to $p+q+1$. Thus the product is again a cochain of dimension $p+q+1$. The reader will note that this formula is similar to that given by Alexander. After various manipulations with this formula, he concludes that the product of two cocycles is always zero! One wonders why he bothered to introduce such a product.

To get something of significance, he now introduces a second product of cochains. As before, let f be a p-cochain, and g a q-cochain. The new product is defined as follows:

$$[f, g](v_0, \ldots, v_{p+q}) = \frac{1}{4(p+q+1)} \sum f(v_{\alpha(0)}, \ldots, v_{\alpha(p)}) g(v_{\alpha(p)}, \ldots, v_{\alpha(p+q)}),$$

where again the sum is over all permutations α of the integers from 0 to $p+q$. This new product has dimension $p+q$, as we know it should have. He now derives the following coboundary formula for this product:

$$(p+q+1)\delta[f, g] = (p+1)[\delta f, g] + (-1)^p(q+1)[f, \delta g]$$

which is the formula we are familiar with today, except for the extraneous numerical factors.

With this formula one can define the product of rational cohomology classes as usual, and hence define the rational cohomology ring of a simplicial complex.

Kolmogoroff next devotes a couple of paragraphs to defining the rational cohomology ring of a *locally* compact space. Here he refers to a *Comptes Rendus* note [33] for some of the relevant definitions. His definition of a p-cochain on such a space is essentially the same as that of Alexander, but he imposes a couple of extra conditions on his cochains. These extra conditions are complicated, and it is difficult to comprehend the reason for introducing them. Probably they were introduced to insure that his cochains would have compact supports, in the language of today. Like Alexander, he factors out the subgroup of cochains which are locally zero.

He then points out that the formulas for the product of cochains on a simplicial complex apply equally well to these cochains on a locally compact space, hence one can define the rational cohomology ring of such a space.

In the last paragraph of [32], the author considers products in the rational cohomology ring of a triangulable closed orientable n-manifold, M. As mentioned above, Kolmogoroff pointed out in his first paper that Poincaré duality for such a manifold can be stated as an isomorphism $H^p(M; \mathbb{Q}) \approx H_{n-p}(M; \mathbb{Q})$. He now asserts that the products he has just defined in cohomology correspond under Poincaré duality to the products in homology defined by intersection of cycles, up to a constant multiple which depends only on the dimensions of the two cohomology class involved. No proof, or hint of a proof, is given.

It should be pointed out that Kolmogoroff published three additional *Comptes Rendus* notes on cohomology in 1936 [34–36]. One theorem in these papers asserts that if R is a compact metric space, G is a compact abelian group, and $\widehat{G}$ is its character group, then the Vietoris homology group $H_p(R; G)$ and the cohomology group $H^p(R; \widehat{G})$ are character groups of each other. As was mentioned above, this result probably motivated Kolmogoroff's definition of cohomology groups for a general space. He also considers relative cohomology groups $H^p(R, Q)$, where R is a locally compact space and Q is a closed subset. Since he has defined cohomology groups with compact supports, he can prove that the relative group $H^p(R, Q)$ is isomorphic to the group $H^p(R - Q)$. Then he asserts that if $H^p(R) = 0 = H^{p-1}(R)$, the groups $H^{p-1}(Q)$ and $H^p(R - Q)$ are isomorphic – a result we would derive today by using the exact cohomology sequence. Indeed, it is very plausible that thinking about the proof of this result led Hurewicz to introduce the exact cohomology sequence in 1941. Another result is the Poincaré duality theorem for *open*, triangulable, orientable manifolds in the form $H_p(M^n) \approx H^{n-p}(M^n)$; again, this depends on the fact that his cohomology groups have compact support. Finally, he puts these last two isomorphisms together to get the Alexander duality theorem.

In his reminiscences, Whitney described his reaction to the Moscow lectures of Alexander and Kolmogoroff and their definition of cup products as follows:

> "From the reputation of these two mathematicians, there must be something real going on; but it was hard to see what it might be. I digress for a moment to say what happened to this product. Within a few months, E. Čech and I both saw a way to rectify the definition. We each used a fixed ordering of the vertices of a simplicial complex K, and defined everything in terms of this ordering." [75]

It is not difficult to see why Whitney and the other participants at the Moscow conference must have been mystified when Kolmogoroff and Alexander wrote down their definitions of a product of cochains. These definitions were pure ad hoc formulas, presented with no motivation. It is hard to guess how Alexander and Kolmogoroff arrived at them. It must have seemed like numerology or magic. It was not until several years later that Lefschetz gave a geometric motivation for cup products.

That J.W. Alexander, a well-established algebraic topologist, should be one of the founders of cohomology theory is not surprising. But as Hassler Whitney remarked in his reminiscences of the Moscow conference, "Kolmogoroff [was] an unlikely person at the conference." Andrei Nikolaevich Kolmogoroff was justly more reknown for his work in other branches of mathematics, such as probability and analysis. In 1936 and 1937 he published ten short papers on algebraic topology; six of these are listed in the bibliography of the present paper. There is a lengthy obituary of Kolmogoroff together with discussions of his work in various areas of mathematics in Vol. 22 (1990) of the *Bulletin of the London Mathematical Society*.

In 1936 Alexander published a more lengthy and corrected version of his second *Proceedings* paper (see [6]). He adopted the correct definition of Čech and Whitney for cup products in a simplicial complex, based on the use of a choice of a fixed linear order of the vertices. Much of the paper is concerned with proving that the resulting product of cohomology classes is independent of the order chosen.

It is interesting to note that exact sequences of groups and homomorphisms first made their appearance in the context of Kolmogoroff's cohomology theory for locally compact spaces. In 1941 Witold Hurewicz published a brief abstract (of a paper that was never published) with the rather cryptic title "On Duality Theorems" (see [29]). In this abstract, Hurewicz described the exact cohomology sequence of a pair consisting of a locally compact space and a closed subspace. He made explicit reference to the work of Kolmogoroff. Of course Hurewicz could not have guessed that some day exact sequences would become ubiquitous in algebraic topology and related branches of mathematics. The exact homology sequence of a pair consisting of a space and subspace did not appear in the literature until about four years later, just before the end of World War II, almost simultaneously on both sides of the Atlantic: in America, in a note of Eilenberg and Steenrod [25], and in France, in a paper by H. Cartan [11].

3. The more comprehensive treatment of cohomology and cup products by Čech and Whitney

Unfortunately, neither Alexander nor Kolmogoroff gave a full and complete exposition of cohomology and cup products after the Moscow conference. In fact, after 1936 Kolmogoroff turned to other branches of mathematics in his research. Alexander published a few more papers on cohomology after the 1936 paper just mentioned; but most of them treated cohomology in terms of rather abstract concepts, and apparently had little influence on his contemporaries.

This gap in the literature was quickly filled by E. Čech and H. Whitney. Čech was the first of these two authors to publish on cohomology, with a paper which appeared in the *Annals of Mathematics* in 1936 [16]. He confined his treatment of cohomology to finite simplicial complexes. In order to define cup products, he first defines what he called "an auxiliary construction", which is an operation assigning to q-cocycle and a p-simplex a $(p+q)$-chain; this operation is required to satisfy several conditions. He then proves that auxiliary constructions always exist. Using such an auxiliary construction, cup products of cocycles are defined. Different auxiliary constructions give rise to different products, but the cohomology class of the product is unique, independent of the choice of the particular auxiliary construction, and only depends on the cohomology classes of the cocycles one started with. One particular auxiliary construction gives rise to the usual formula for cup product of cochains based on a choice of a chosen linear order of the vertices. Čech then quickly proves the basic properties of cup products of cohomology classes, i.e. bilinearity, associativity, and commutativity.

In a similar manner Čech uses an auxiliary construction to define the cap product of a cycle and a cocycle, and proves that this gives rise to a product of homology classes with cohomology classes, which has all the expected properties. This is the first appearance of the cap product in the mathematical literature.

Čech now uses this cap product to prove the Poincaré duality theorem for oriented, triangulated, closed combinatorial manifolds. The Poincaré duality isomorphism is defined by means of the cap product with the fundamental top-dimensional homology class of the manifold. The proof, which operates in one fixed triangulation of the manifold, is rather lengthy. Finally Čech proves that in an orientable manifold of the type he is considering, the Poincaré duality isomorphism takes the cup product of cohomology classes to the intersection product of homology classes.

This paper of Čech has a somewhat different character from the publications of Kolmogoroff and Alexander which we have mentioned previously (except for Alexander [6]). The definitions are precise, and the theorems are given complete proofs. Although the discussion is limited to simplicial complexes, it must have been much more accessible to topologists of the 1930's. The main thing that is missing is a discussion of the homomorphism on cohomology groups induced by a continuous map. Also, there is no motivation for the definitions that Čech makes.

Whitney's publication followed rather quickly after Čech's. An outline of his results appeared in Vol. 23 of the *Proceedings of the National Academy of Sciences* in 1937 [71] and the complete paper was published in the *Annals of Mathematics* in 1938 [74]. Whitney covered all the ground which had been covered by Čech, but he had several additional results.

Whitney worked with cell complexes which are more general than simplicial complexes. What he called "a complex admitting a product theory" is essentially a finite cell complex such that each *closed* cell has the homology groups of a point. For such a complex, he shows how to define the cup product of two cochains, and the cap product of a chain and a cochain. These products are required to satisfy several conditions. They always exist for such complexes, but they are not unique. However, the products obtained on passage to homology and cohomology classes are unique. The proofs of existence and uniqueness are somewhat involved.

Whitney was apparently the first to explicitly define the homomorphism induced on cohomology groups by a continuous map, and to explicitly state how the cup and cap products behave under such homomorphism. He also proved that the products defined in the homology and cohomology of a cell complex are actually topological invariants of the underlying space of the complex. Finally, he gave the formulas for cup and cap products in a product space in terms of the cup and cap products in the factors.

To summarize, this paper of Whitney is even more comprehensive than the preceding paper of Čech. The definitions are clear and precise, and the proofs are completely given. It is surprisingly modern in tone, compared to all the earlier papers on cohomology. Apparently Whitney introduced the modern terms "coboundary", "cocycle", "cohomology", "cup product" and "cap product". Since he was dealing only with finite complexes, he identified chains and cochains; in fact, the term "cochain" never appears in this paper.

The main difficulty with this paper is that the proof of existence and uniqueness of products in a cell complex is unmotivated. It is difficult to guess how Whitney found his proofs.

This question of a satisfactory motivation for the definition of cup and cap products was finally settled in Lefschetz's 1942 Colloquium Volume [38]. The reader is referred to pp. 38–41 of an article by Steenrod [58] for an excellent description of Lefschetz's ideas on this subject using the diagonal map $d: X \to X \times X$ of any space X into the product space $X \times X$. This idea of Lefschetz has since been used in several of the most influential textbooks on algebraic topology to introduce the reader to cup and cap products. It should be pointed out that De Rham used the diagonal map in the case where X is a differential manifold for a somewhat similar purpose in 1932; see [18].

4. Early applications of cohomology theory

The papers we have discussed put cohomology groups and cup products on a firm foundation. Undoubtedly many contemporary topologists were impressed, but before they put forth the effort to learn this new theory, they wanted to be convinced that it was good for something. In other words, they would have liked to see evidence that some problems could be solved more easily with cohomology theory that with homology theory. Fortunately, examples of such problems were soon forthcoming. We will consider some of these.

4.1. *The classification of maps of an n-dimensional complex into an n-sphere*

In 1932, H. Hopf published a beautiful paper discussing the homotopy classification of continuous maps of an n-dimensional complex K into an n-sphere S^n [28]. The statements of the theorems in this paper were in terms of the homomorphism induced by a given map $f: K \to S^n$ on the n-dimensional homology groups; in general, it is necessary to use many different coefficient groups for these homology groups. In 1937 Whitney gave a new treatment of this homotopy classification problem using cohomology instead of homology (see [73]). The final result is very simple to state: The homotopy classes of maps $f: K \to S^n$, (where K is an n-dimensional complex) are in 1–1 correspondence with the elements of the integral cohomology group $H^n(K)$. The correspondence is established by assigning to each such map f the element $f^*(d_n)$, where d_n is a generator of the infinite cyclic cohomology group $H^n(S^n)$.

To see why Whitney's version of this theorem is superior, recall that if K is a *finite* n-dimensional complex, the integral cohomology group $H^n(K)$ is isomorphic to the direct sum of the integral homology group $H_n(K)$ (which is a free abelian group) and the torsion sub-group of $H_{n-1}(K)$. The only way the torsion subgroup of $H_{n-1}(K)$ can come into play in the homomorphism $f_*: H_n(S^n) \to H_n(K)$ is to use different coefficient groups for homology. Not only is the precise statement of the theorem simpler using cohomology, but Whitney's proof is shorter and simpler than that of Hopf.

4.2. *The theory of obstructions to extensions and homotopies of continuous maps*

The first more or less systematic exposition of this theory was by Eilenberg [22]. However, the basic idea was used in papers of Whitney, Pontrjagin, and perhaps others about this time, without ever calling the cocycles they defined "obstructions". It is difficult to imagine anybody trying to fit the ideas involved into homology theory; it is clearly a cohomological theory.

4.3. *Characteristic classes of sphere bundles and more general fibre bundles*

At the 1935 Conference in Moscow, Hopf lectured on the work of his student, E. Stiefel. Stiefel's paper appeared the next year [61]. In modern language, Stiefel defined certain homology classes in a differentiable manifold which are the Poincaré duals of the Stiefel–Whitney characteristic classes of the tangent bundle. His method was to construct representative cycles for these homology classes by a very geometric process.

Whitney gave a talk at the Moscow conference entitled "Sphere spaces" (sphere spaces are now called "sphere bundles"), see H. Whitney [70]. These two talks, and the corresponding papers, marked the start of work on the general subject of fibre bundles. The most important invariants of fibre bundles are usually various characteristic classes, which are always cohomology classes.

4.4. *Pontrjagin's classification theorem for mappings of a 3-dimensional complex K into a 2-sphere*

In 1941 in the dark days of World War II, L. Pontrjagin published a paper giving the homotopy classification for continuous maps of a 3-dimensional complex into a 2-sphere (see [43]). He had previously announced the results of this paper in a brief note without any proofs (see [42]). (Unfortunately, this 1938 note contains an erroneous statement about the homotopy classification of maps of an $(n + 1)$-complex into an n-sphere for the case $n > 2$.) The statement of Pontrjagin's classification theorem is in terms of cohomology, and requires the use of cup products; there is no way to avoid these cup products.

In spite of these examples of the possible advantages of cohomology over homology in certain situations, some topologists were hesitant to use the new cohomology theory. Probably the main reason was the difficulty of relating cohomology classes to one's geometric intuition. By contrast, most algebraic topologists felt that they had a good geometric intuition about cycles, homologous cycles, etc.

5. Cohomology theory for general spaces

The papers we have discussed so far developed cohomology theory and cup products for finite simplicial complexes and a more general type of finite cell complexes. Also, Alexander and Kolmogoroff independently introduced a cohomology theory for more general spaces in which the p-cochains were skew-symmetric functions of $p + 1$ points of the space with values in the coefficient group.

It would seem that it would not have been difficult to "dualize" the definition of Čech homology groups for compact spaces, and thus define Čech cohomology groups. Probably the first person to do this in a published paper was Steenrod in his thesis [51]. The exact timing here is rather interesting. Steenrod published a preliminary announcement of the results of his thesis in 1935 in the *Proceedings of the National Academy of Sciences* [50].

This was the same volume of the *Proceedings* in which Alexander's first two short notes on cohomology appeared [4, 5], but Steenrod's paper was submitted before those of Alexander, and appeared earlier in that volume of the *Proceedings*. Presumably Steenrod was not aware of the contents of Alexander's two short notes when he submitted his paper. In any case, there is no hint of cohomology in Steenrod's preliminary announcement. On the other hand, when the thesis was actually published a year later in the *American Journal of Mathematics*, Steenrod was well aware of Alexander's two short notes; he explicitly refers to [4] in his bibliography, and he discusses the Čech cohomology theory of a compact space (although he called it the "dual homology theory"). In the introduction to his thesis, he stresses the future importance of cohomology theory.

Čech cohomology theory based on infinite coverings of a non-compact space was introduced by C.H. Dowker. He published a brief announcement of his results in 1937 in the *Proceedings of the National Academy of Sciences* [20]. Unfortunately, he did not get around to publishing the details of his work until a decade later (see [21]). In these papers, Dowker extended the Hopf classification theorem mentioned previously to a theorem giving a homotopy classification of maps of a non-compact n-dimensional topological space into an n-sphere.

Singular cohomology theory did not arrive on the scene until 1944 (see Eilenberg [23]). Eilenberg described the difficulties with the earlier singular homology theory as follows in the introduction to his paper:

> "The best treatment of the singular homology theory so far has been given by Lefschetz. He defines a singular simplex in a space X as a pair (s, T), where s is an oriented simplex and $T : s \to X$ is a continuous mapping. If $B : s \to s'$ is a barycentric map of s onto another oriented simplex of the same dimension as s, then
>
> $$(s, T) \equiv \pm(s', TB^{-1}), \qquad (*)$$
>
> where the sign is $+$ or $-$ according as B preserves or reverses the orientation. Following a suitable definition of boundary and incidence numbers, Lefschetz arrives at what he calls the "total singular complex" $S(X)$ of the space X. In this closure finite complex homologies, cohomologies, and products can be constructed.
>
> The main difficulty with using the complex $S(X)$ is that it is not a bona fide abstract complex. Unfortunately, relation $(*)$ causes elements of order 2 to appear in the group of chains, while in an abstract complex the group of chains ought to be free. There is the possibility of leaving out the elements of order 2 as degenerate, but this would make the use of the complex $S(X)$ cumbersome."

Eilenberg's solution of the problem posed by these difficulties is now well-known and standard: he omitted the equivalence relation $(*)$ in the above quotation. A singular n-simplex is now a continuous map $T : s \to X$, where the proto-type standard n-simplex s is fixed once for all. There is no question of comparing this singular n-simplex with a map of another n-simplex s' into X. This change leads to a much larger total singular complex $S(X)$, but now there are no elements of order two in the chain groups; they are free abelian groups.

Parallel to this new definition of the singular chain complex of a space, Eilenberg introduced a new chain complex for a simplicial polyhedron. This new chain complex is now usually called the "ordered chain complex" of the given polyhedron, in contrast to what was then the more usual chain complex, which is now called the "oriented chain complex" of the given polyhedron. Eilenberg gives no clue as to his thinking, but these two new chain complexes are closely related. It is difficult to imagine somebody thinking of one of them without thinking of the other.

The oriented chain complex of a simplicial polyhedron had been more or less standard (although not under that name) since the earliest days of algebraic topology. It has great appeal to one's geometric intuition, and is closely connected with the integration of differential forms. By contrast, the ordered chain complex of a simplicial polyhedron has little geometric appeal, and it is surprising that it gives the correct homology groups. How did Eilenberg hit on this new idea? He gives no motivation whatsoever in his paper. In any case, it greatly simplified the exposition of homology and cohomology theory for succeeding generations of mathematicians.

Once Eilenberg had this new singular chain complex, it was a routine matter to define singular cohomology groups. These new chain complexes also simplified the introduction of cup products, because it was no longer necessary to choose an ordering of the vertices. The vertices of a simplex were already ordered.

The type of cohomology theory that Alexander and Kolmogoroff had proposed (for compact metric and locally compact spaces respectively) was modified and given a clear and complete exposition by the late Edwin H. Spanier in his thesis [49]. Spanier considers cohomology groups for an arbitrary topological space X with coefficients in an arbitrary abelian group G. A p-cochain is a function $f(x_0, x_1, \ldots, x_p)$ of $(p+1)$ points of X with values in the group G. Unlike Alexander and Kolmogoroff, Spanier allows arbitrary functions, they need not be skew-symmetric. This corresponds to Eilenberg's innovation in the definition of singular homology and cohomology: one uses ordered p-simplexes rather than oriented p-simplexes. One has to factor out by the subgroup of p-cochains which are locally zero, as did Alexander and Kolmogoroff.

In the introduction to his paper, Spanier acknowledges that his definition followed a suggestion of A.D. Wallace. The exact nature of Wallace's suggestion is not made clear, but one presumes that it was to drop the condition of skew-symmetry for the cochains.

Spanier gives a comprehensive treatment of the resulting cohomology theory. He proves that this theory satisfies the Eilenberg–Steenrod axioms for a cohomology theory, and that it is "continuous" under passage to inverse limits for compact Hausdorff spaces; as a consequence, this cohomology theory agrees with the Čech type cohomology theory on compact Hausdorff spaces. Finally, he proves directly that on locally finite simplicial complexes, his cohomology theory is the same as the simplicial theory based on infinite cochains. In an appendix, cup products are defined. Recall that Alexander in [6] had to linearly order all the points of a topological space in order to define the cup product of two of his skew-symmetric cochains. That is unnecessary with the definitions Spanier uses.

This cohomology theory that Spanier described in his thesis is now generally known as the "Alexander–Spanier Theory" (at least in America). Perhaps this is due to the fact that Spanier makes no reference to the work of Kolmogoroff in his thesis. In view of the historical record, it is clear that it should be called the "Alexander–Kolmogoroff–Spanier Theory".

In 1948, the cohomology groups defined by Spanier and the general Čech type cohomology groups (using infinite open coverings) were shown to be isomorphic for a rather general class of topological spaces (see Hurewicz, Dugundji and Dowker [30]).

As we mentioned earlier, the cohomology groups Kolmogoroff defined for a general locally compact Hausdorff space were based on cochains with compact support. Kolmogoroff did not consider any other kind of cohomology groups for general spaces. On the other hand, the cohomology groups for general spaces defined by Eilenberg, Dowker and Spanier were definitely not based on cochains with compact support. Apparently the first person to simultaneously consider cohomology of two types – that based on cochains with compact support and that based on cochains with arbitrary support – was H. Cartan in lectures he gave at Harvard University in the spring of 1948. Mimeographed notes of these lectures and of lectures he gave at the École Normale Superieur in Paris were widely circulated and very influential (see [12, 13]). In these notes, he explicitly introduced the two types of singular cohomology and the two types of Alexander–Kolmogoroff–Spanier cohomology. Cartan acknowledged receiving inspiration from two later papers by J.W. Alexander [7, 8] and from the work of J. Leray. This is the only case the author knows of reference being made to these later papers of Alexander concerned with what he called gratings.

6. The Pontrjagin squaring operation

In 1942 Pontrjagin published a brief note concerned with the problem of determining the third homotopy group of a simply connected space in terms of homology and cohomology invariants of the space (see [44]). Because World War II was going on at this time, this paper did not receive the attention it deserved until several years later. It was noteworthy because it introduced the first cohomology operation other than cup products, namely, the Pontrjagin squaring operation.

Before Pontrjagin got around to writing up a detailed account [45] of his results, the subject was taken up by J.H.C. Whitehead (see [68, 69]). Since the Pontrjagin squaring operation is not very well known, we will give a more detailed exposition of Pontrjagin's results, following the papers of J.H.C. Whitehead.

Pontrjagin's first step in defining his squaring operation was to define a new product of cochains, which was later called the cup-1 product. Let K be a simplicial complex; as in the Čech–Whitney definition of cup products, we must choose a linear ordering of the vertices of K. If f is an integral p-cochain and g is an integral q-cochain, $f \cup_1 g$ is an integral $(p + 1 - 1)$-cochain defined as follows:

$$\begin{aligned}&(f \cup_1 g)(v_0, \dots, v_{p+q-1}) \\ &\quad = \sum_{j=0}^{p-1} (-1)^{(p-j)(q+1)} f(v_0, \dots, v_j, v_{j+q}, \dots, v_{p+q-1}) g(v_j, \dots, v_{j+q}),\end{aligned}$$

where the vertices $v_0, v_1, \dots$ must be written in the chosen order. This product satisfies the following coboundary formula:

$$\delta(f \cup_1 g) = (-1)^{p+q-1}\big(f \cup g - (-1)^{pq} g \cup f\big) + \delta f \cup_1 g + (-1)^p f \cup_1 \delta g.$$

Now define

$$\mathcal{P}(f) = f \cup f + f \cup_1 \delta f.$$

It can now be proved that if f is a cocycle modulo $2r$, then $\mathcal{P}(f)$ is a cocycle modulo $4r$, for any positive integer r. Moreover, the cohomology class of $\mathcal{P}(f)$ only depends on the cohomology class of f. Thus $\mathcal{P}$ defines a mapping

$$\mathcal{P}: H^p(K;\mathbb{Z}_{2r}) \to H^{2p}(K;\mathbb{Z}_{4r}),$$

which J.H.C. Whitehead called the Pontrjagin square. It is *not* a homomorphism; instead, it satisfies the following two conditions:

$$\begin{aligned} \mathcal{P}(u+v) &= \mathcal{P}(u) + \mathcal{P}(v) + u \cup v, \\ \mathcal{P}(-u) &= \mathcal{P}(u). \end{aligned}$$

For this reason, it is called a quadratic map. It is a natural cohomology operation, in the sense that it commutes with the homomorphisms induced by continuous maps. Later on it was proved that if the degree p is odd, $\mathcal{P}$ is a composition of other cohomology operations. This is not true if p is even.

In his 1942 note Pontrjagin gave some motivation for his squaring operation. In more modern terms, such a motivation can be described as follows (for the sake of simplicity, we will consider only the case $r = 1$ in the above formulas. The case where $r > 1$ is entirely similar).

Consider the short exact sequence of coefficient groups:

$$0 \to \mathbb{Z}_2 \to \mathbb{Z}_4 \xrightarrow{\rho} \mathbb{Z}_2 \to 0.$$

This gives rise to a corresponding long exact sequence of cohomology groups:

$$\cdots \to H^n(K;\mathbb{Z}_2) \to H^n(K;\mathbb{Z}_4) \xrightarrow{\rho} H^n(K;\mathbb{Z}_2) \xrightarrow{\beta} H^{n+1}(K;\mathbb{Z}_2).$$

The so-called "Bockstein homomorphism" β satisfies the following formula vis-a-vis mod 2 cup products:

$$\beta(u \cup v) = (\beta u) \cup v + u \cup (\beta v).$$

Using commutativity of cup products and the fact that we are computing mod 2, it follows that

$$\beta(u \cup u) = 0.$$

Thus by exactness, there exists a mod 4 cohomology class x such that

$$\rho(x) = u \cup u.$$

The Pontrjagin square $\mathcal{P}(u)$ is a way of making a "natural" choice for the cohomology class x. As a matter of fact, this is one of the basic properties of the Pontrjagin square: $\rho[\mathcal{P}u] = u \cup u$.

We will now try to explain the connection of the Pontrjagin squaring operation with the 3-dimensional homotopy group. Let

$$h_n : \pi_n(K) \to H_n(K), \quad n = 1, 2, 3, \ldots,$$

denote the Hurewicz homomorphism. For a simply connected space K, h_2 is an isomorphism, and it can be proved that $h_3 : \pi_3(K) \to H_3(K)$ is an epimorphism. Thus the main problem is to determine the kernel of h_3. One way to construct elements in the kernel of h_3 is as follows. Let $\eta \in \pi_3(S^2)$ denote the homotopy class of the Hopf map $S^3 \to S^2$; it is a generator of the infinite cyclic group $\pi_3(S^2)$. Given any element $\alpha \in \pi_2(K)$, we can compose it with η to form the element $\alpha \circ \eta \in \pi_3(K)$; obviously $\alpha \circ \eta$ belongs to the kernel of the homomorphism h_3. J.H.C. Whitehead's results show that we can obtain a set of generators of the kernel of h_3 by this process; but actually they prove a much more precise result than this. The situation is complicated by the fact that the operation $\alpha \to \alpha \circ \eta$ is a mapping $\pi_2(K) \to \pi_3(K)$ which is *not* a homomorphism; instead we have the formula

$$(\alpha + \beta) \circ \eta = \alpha \circ \eta + \beta \circ \eta + [\alpha, \beta]$$

for any elements $\alpha, \beta \in \pi_2(K)$ (the square brackets denote the Whitehead product). We also have

$$(-\alpha) \circ \eta = \alpha \circ \eta.$$

Thus this mapping $\pi_2(K) \to \pi_3(K)$ is also a quadratic mapping.

The exact sequence mentioned in the title of [69] is a natural long exact sequence defined for any simply connected space X which has the Hurewicz homomorphisms $h_n : \pi_n(X) \to H_n(X)$ as every third homomorphism in the sequence. Whitehead defined a new sequence of groups $\Gamma_n(X)$, $n > 0$, and wrote his exact sequence as follows:

$$\cdots \to H_{n+1}(X) \xrightarrow{b_{n+1}} \Gamma_n(X) \xrightarrow{i_n} \pi_n(X) \xrightarrow{h_n} H_n(X) \xrightarrow{b_n} \cdots.$$

Unfortunately, the new groups $\Gamma_n(X)$ are very difficult to determine in general. However, since X is assumed to be simply connected,

$$\Gamma_2(X) = \Gamma_1(X) = 0$$

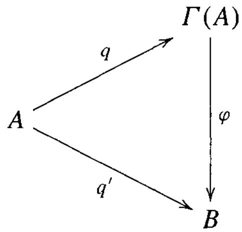
and J.H.C. Whitehead showed that the structure of $\Gamma_3(X)$ may be described as follows.

If A and B are abelian groups (written additively), a function $q : A \to B$ is called a *quadratic map* if it satisfies the following two conditions:

(1) The map $(a, b) \to q(a + b) - q(a) - q(b)$ is a bilinear map $A \times A \to B$.

(2) $q(-a) = q(a)$ for all $a \in A$.

We had two naturally occurring examples of quadratic maps in the preceding paragraphs. For any abelian group A, J.H.C. Whitehead pointed out that one can define another abelian group $\Gamma(A)$ and a quadratic map $q : A \to \Gamma(A)$ which is a universal object for quadratic maps from A to any other abelian group. To be precise for any abelian group B and any quadratic map $q': A \to B$, there exists a unique homomorphism $\varphi: \Gamma(A) \to B$ such that the following diagram is commutative:



As usual with universal mapping problems, $\Gamma(A)$ is unique up to isomorphism. If A is cyclic of order m, then $\Gamma(A)$ is also cyclic, and is of order m or $2m$ according as m is odd or even. If A is infinite cyclic, then so is $\Gamma(A)$. Finally we have the following rule for direct sums:

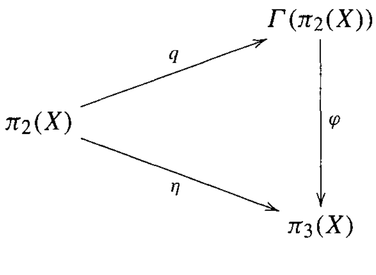
$$\Gamma(A \oplus B) = \Gamma(A) \oplus \Gamma(B) \oplus (A \otimes B).$$

Thus we can easily determine the structure of $\Gamma(A)$ for any finitely generated abelian group A.

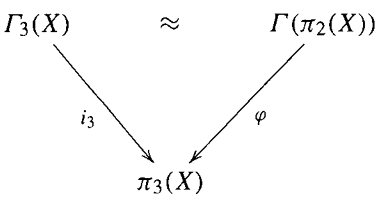
In view of the above described quadratic map

$$\eta: \pi_2(X) \to \pi_3(X),$$

there is a natural homomorphism $\varphi: \Gamma(\pi_2(X)) \to \pi_3(X)$ such that the following diagram is commutative:



Whitehead showed that for simply connected spaces X, the group $\Gamma_3(X)$ is naturally isomorphic to $\Gamma(\pi_2(X))$, and the isomorphism makes the following diagram commutative:



In other words, we can substitute $\Gamma(\pi_2(X))$ for $\Gamma_3(X)$ and the homomorphism φ for i_3 in J.H.C. Whitehead's exact sequence. The result is the following:

$$\cdots \to H_4(X) \xrightarrow[b_4]{} \Gamma(\pi_2(X)) \xrightarrow[\varphi]{} \pi_3(X) \xrightarrow[h_3]{} H_3(X) \to 0.$$

Thus the principal remaining problem is to determine the homomorphism b_4. J.H.C. Whitehead showed that this homomorphism is determined by the Pontrjagin squaring operators in the space X. In order to accomplish this, Whitehead showed that for any abelian group A, the definition of the Pontrjagin square can be extended so it is a quadratic map

$$\mathcal{P}: H^n(X; A) \to H^{2n}(X; \Gamma(A)).$$

(Pontrjagin originally defined it for the case where A is a cyclic group of even order.) Then the homomorphism b_4 is determined by the following sequence of homomorphisms (we have abbreviated $H_2(X)$ and $\pi_2(X)$ to H_2 and π_2, respectively):

$$\mathrm{Hom}[H_2, \pi_2] \approx H^2(X; \pi_2) \xrightarrow{\mathcal{P}} H^4(X; \Gamma(\pi_2)) \to \mathrm{Hom}[H_4(X), \Gamma(\pi_2)].$$

Recall that the Hurewicz homomorphism $h_2: \pi_2 \to H_2$ is an isomorphism. J.H.C. Whitehead's result is that the element h_2^{-1} of the left most group in the above sequence is sent to the element b_4 in the right most group in the sequence. Thus the homomorphism b_4 is determined by the Pontrjagin squaring operation $\mathcal{P}$. It follows that the homotopy group $\pi_3(X)$ is also determined up to a group extension by $\mathcal{P}$. Actually, Whitehead also showed that if X is a finite, simply connected polyhedron, the group extension can also be determined.

Whitehead gave the following interesting example of two simply connected 4-dimensional CW-complexes which have isomorphic cohomology rings (with any coefficients) but are not of the same homotopy type. Let $\Sigma\mathbb{RP}^2$ denote the suspension of the real projective plane and $K_0 = \Sigma\mathbb{RP}^2 \vee S^4$. Using the results of J.H.C. Whitehead quoted above, it is readily seen that $\pi_3(\Sigma\mathbb{RP}^2)$ is a cyclic group of order 4; let K_2 denote the space obtained by adjoining a 4-cell to $\Sigma\mathbb{RP}^2$ by a map representing twice a generator of $\pi_3(\Sigma\mathbb{RP}^2)$. Then K_0 and K_2 have isomorphic homology and cohomology groups, and all cup products of positive degree cohomology classes are zero, with any coefficients. But they can be distinguished by the Pontrjagin square

$$\mathcal{P}: H^2(K_i; \mathbb{Z}_2) \to H^4(K_i; \mathbb{Z}_4), \quad i = 0, 2.$$

For K_0, this Pontrjagin square is zero, but it is non-zero for K_2. Also, the third homotopy groups are different.

After this initial work by Pontrjagin and J.H.C. Whitehead, the Pontrjagin square found various other applications. The most spectacular of these was the proof by W.T. Wu that the Pontrjagin classes reduced mod 4 of a differentiable manifold are invariants of the homotopy type of the manifold (see [79]). Another application was a simplified proof by the present author of the following theorem of M. Mahowald: If a closed, connected, non-orientable n-manifold (n even) is embedded differentially in $\mathbb{R}^{2n}$, the reduction mod 4 of the twisted Euler class of the normal bundle is the same for all embeddings (see [39]). In general, the twisted Euler class of the normal bundle of an embedding of a non-orientable manifold varies with the embedding. For $n \geqslant 4$, this theorem is best possible.

7. Steenrod squares and reduced pth-powers

After the Pontrjagin square, the next natural cohomology operations to be discovered were the Steenrod squares. Steenrod's original paper on this subject was published in 1947 (see [52]). In this paper, Steenrod was concerned with the problem of the homotopy classification of mappings of an $(n+1)$-dimensional complex into an n-dimensional sphere. As we have just seen, Pontrjagin introduced his squaring operation to solve the case $n = 2$ of this problem. As Steenrod mentions in the introduction of his paper, Pontrjagin in 1938 and H. Freudenthal in 1939 had announced (without proof) results on this homotopy classification problem which were incorrect.

Steenrod defined his new cohomology operations by means of rather *ad hoc* cochain formulas. Let K be a simplicial complex. For any p-dimensional cochain u and q-dimensional cochain v on K he defined a $(p+q-i)$-dimensional cochain $u \cup_i v$, called the cup-i product of u and v. For $i = 0$, this is just the usual cup product of Čech and Whitney; for $i = 1$, it is the operation on cochains defined by Pontrjagin. In general, the cup-i product is defined by a formula similar to that for the cup-1 product, but for $i > 1$ it is more complicated. Steenrod proved that this product satisfied the following coboundary formula:

$$\begin{aligned} \delta(u \cup_i v) &= (-1)^{p+q-i} u \cup_{i-1} v + (-1)^{pq+p+q} v \cup_{i-1} u + \delta u \cup_i v \\ &\quad + (-1)^p u \cup_i \delta v. \end{aligned}$$

Using this formula, Steenrod was able to prove that if u is a p-dimensional cocycle mod 2, then $u \cup_i u$ is a $(2p-i)$-dimensional cocycle mod 2 whose cohomology class only depends on the cohomology class of u. Thus there is defined an operation

$$Sq_i : H^p(K; \mathbb{Z}_2) \to H^{2p-i}(K; \mathbb{Z}_2)$$

which is a homomorphism. This operation commutes with the homomorphism induced by a continuous map of one space into another, and is a new invariant of topological spaces. Later on it was found more convenient to denote these operations by the notation

$$Sq^n : H^p(K; \mathbb{Z}_2) \to H^{p+n}(K; \mathbb{Z}_2),$$

i.e. the superscript "n" denotes the amount by which it increases the degree.

Steenrod used this new cohomology operation to give his solution of the homotopy classification problem for maps of an $(n+1)$-dimensional complex into an n-sphere. Other topologists quickly saw that this new cohomology operation offered interesting possibilities for research and jumped on the band wagon. Particularly striking were the results announced in 1950 in Vol. 230 of the *Comptes Rendus de l'Academie des Sciences de Paris* by René Thom and the Chinese mathematician Wu Wen Tsun (then visiting in France; see [62, 63, 76, 77]). In these notes, Wu gave his well-known formulas for the result of applying a Steenrod squaring operation to a Stiefel–Whitney class of a sphere bundle and his formulas for computing the Stiefel–Whitney classes of a closed manifold in terms of the Steenrod operations on the mod 2 cohomology of the manifold. Thom introduced the Thom space of a sphere bundle, and the Thom class of such a bundle; then he showed that the Stiefel–Whitney classes of the bundle are determined by applying Steenrod squares to the Thom class. These two authors published complete proofs of their announced results a few years later (see [64, 78]). Additional striking results were announced by J.P. Serre in 1952 [47]. In the late 1940's and the early 1950's, one of the most pressing problems in algebraic topology was to determine the homology and cohomology of Eilenberg–Mac Lane spaces. Serre completely solved this problem in 1952 for cohomology with mod 2 coefficients; his methods relied heavily on the properties of the Steenrod squaring operations (of course they also depended on the clever use of spectral sequences of fiber spaces). Complete proofs were published in [48]. In this paper, Serre also gave for the first time the precise, general definition of a (first order, natural) cohomology operation with arbitrary coefficient groups. In addition, he pointed out that such cohomology operations are in a natural, one-to-one correspondence with cohomology classes in Eilenberg–Mac Lane spaces.

Soon after Steenrod described his new squaring operations, other mathematicians proved some of their basic properties. One of the first examples of this was the discovery by H. Cartan of his well-known formula

$$\mathcal{S}q^n(x \cup y) = \sum (\mathcal{S}q^i x) \cup (\mathcal{S}q^j y),$$

where the sum is over all pairs (i, j) such that $i + j = n$ (see [14]). Two years later José Adem, Steenrod's Mexican Ph.D. student, published an announcement of the relations on iterated Steenrod squares which are now referred to as "the Adem relations" (see [1]).

As explained above, Steenrod defined his squaring operations by means of the cup-i products, $i = 0, 1, 2, \ldots$. At the time, it seemed possible that one should be able to derive additional invariants from the cup-i products in addition to the Steenrod squares. Unfortunately, not much ever came of this line of research. Another observation was that the Steenrod squares were associated with mod 2 cohomology; was it possible that there existed "cubing" operations associated with mod 3 cohomology, fifth power operations associated with mod 5 cohomology, etc.? This line of research was pursued by Steenrod himself soon after the discovery of the squaring operations. Steenrod described his trials and errors which eventually led to success at the end of a lecture on the work of Lefschetz, [58]. Recall that Lefschetz clarified and provided motivation for the cochain formulas for cup products of Alexander, Kolmogoroff, Čech and Whitney by considering the diagonal map $K \to K \times K$ for any complex K (see Section 3 above). After fruitless experimenting with various formulas for possible cup-i products, Steenrod realized that he must try to generalize this idea of Lefschetz, and apply it to this new problem. To generalize from squaring operations to nth power operations, it is necessary to replace $K \times K$ by the product K^n of n factors K. In the case of the squaring operation, it is necessary to consider the action of a cyclic group of order two on $K \times K$, where the action is by interchanging the two factors. In the general case, one must consider the action of the symmetric group of degree n on K^n, operating by permuting the factors. In addition, it is necessary to bring into the picture an acyclic complex on which the symmetric group of degree n operates freely. For details, we refer the reader to the lecture notes by Steenrod and Epstein [59]. One can follow the evolution of Steenrod's ideas by perusing the series of papers he published, starting with a very brief announcement at the International Congress of Mathematicians in Cambridge, Mass. in 1950, and culminating in the above mentioned lecture notes by Steenrod and Epstein, published in 1962. This series of research papers is noteworthy in that it was almost entirely a solo effort by Steenrod himself. The only case where he had a collaborator was one paper in 1957 that he wrote jointly with his former student, Emery Thomas.

In May, 1951 Steenrod gave a series of lectures at the Collège de France in Paris on his new cohomology operations. This served to publicize them to the mathematical community in France, with the result that other mathematicians quickly found applications of these operations. Perhaps the most notable example of these applications was the work of A. Borel and J.P. Serre; their results were announced in 1951 [9] and complete details were published in 1953 [10]. These authors showed how to compute the reduced pth power operations in the mod p cohomology of any of the classical Lie groups or their classifying spaces. Then they made various applications of these calculations. For example, they proved that for $n \geqslant 8$, the n-sphere does not admit an almost complex structure, no matter what the differentiable structure. Also, they determined the p-primary components of certain homotopy groups of the classical groups using these calculations.

The relations satisfied by the iterated reduced power operations were soon determined by José Adem; his results were announced in 1953 [2] and he published complete proofs in [3]. These relations were also determined by H. Cartan, using a different method; see [15].

8. The generalized Pontrjagin powers

Just as the Steenrod squares could be generalized to pth power operations for every odd prime p, so it seemed reasonable to hope that the Pontrjagin squaring operation could be generalized to some kind of pth power operation for primes $p > 2$. This hope was realized by Emery Thomas, a student of Steenrod, in his 1955 Ph.D. thesis. He published an announcement of his results in 1956 (see [65]) and the complete details in 1957, [66]. In their simplest form, these operations are functions defined in the cohomology of any CW-complex K for any odd prime p, as follows:

$$\mathcal{P}_p : H^{2n}(K; \mathbb{Z}_{pm}) \to H^{2pn}(K; \mathbb{Z}_{p^2m}).$$

This function is not a homomorphism; as the name implies, it behaves like raising to the pth power. In order to explain its algebraic properties more concisely, Emery Thomas found it convenient to recast the entire theory of these operations in terms of rings with divided powers.

Unlike the case of the Pontrjagin squares and the Steenrod operations, no applications of these newest cohomology operations were forthcoming. One can only speculate as to the reason for this. Was it something basic in the nature of the universe of mathematics and its applications? Or was it because very few topologists ever bothered to become familiar with Pontrjagin pth powers?

9. Are there any more cohomology operations?

At this stage in the history of algebraic topology (about 1957) topologists had at hand a rather extensive list of cohomology operations: cup products, Pontrjagin squares, Steenrod squares and reduced pth powers, and Pontrjagin pth powers. In addition, there were a couple of more simple minded operations which had been known for many years: namely, the operation determined by a homomorphism of coefficient groups, and the Bockstein coboundary operator determined by a short exact sequence of coefficient groups. From this list of rather basic operations, one can construct new ones by iterating the known ones, adding two or more operations, etc. The question naturally arises, does one obtain *all* natural, first order cohomology operations in this way? The answer is affirmative at least in the case of finitely generated coefficient groups. This was proved by John Moore (unpublished) and by Albrecht Dold [19]. The proof ultimately depends on the correspondence pointed out by J.-P. Serre between cohomology operations and cohomology classes in Eilenberg–Mac Lane spaces (see Section 7 above).

10. Conclusion

This result of J.C. Moore and A. Dold brought to a close a chapter in the history of algebraic topology. Between the years 1935 and 1959 cohomology theory was first defined, the main details of the various types of cohomology theory were worked out, cup products, Pontrjagin powers and Steenrod reduced powers were developed, and it was proved that there are no more cohomology operations left to be discovered. All this was accomplished in this rather short time span in spite of the world-wide depression of the 1930's and the extreme destruction and disruption wrought by World War II.

Of course there has been interesting research on cohomology theory since 1959. New and better methods of exposition of various aspects of the theory have been published. Numerous applications of cohomology to homotopy theory and other parts of mathematics have been described. Although there are no more primary cohomology operations to be discovered, the field of second order and higher order cohomology operations is still open (although it does not seem to be a fertile field for research).

One of the most interesting observations about the history of cohomology operations is that often the first definition of an operation was by *ad hoc*, unmotivated formulas; they often seemed like magic formulas, because they worked. This was true in the case of cup products, the Pontrjagin square, and the Steenrod squares. It was only later that a more conceptual, well-motivated method of definition for these operations was developed. Is this something which is common in the history of mathematics, or is the history of cohomology operations rather unusual in this respect?

Bibliography

[1] J. Adem, *The iteration of Steenrod squares in algebraic topology*, Proc. Nat. Acad. Sci. **38** (1952), 720–726.

[2] J. Adem, *Relations on iteration reduced powers*, Proc. Nat. Acad. Sci. **39** (1953), 636–638.

[3] J. Adem, *The relations on Steenrod powers of cohomology classes*, Algebraic Geometry and Topology: A Symposium in Honor of S. Lefschetz, R.H. Fox, D.C. Spencer and A.W. Tucker, eds, Princeton Univ. Press, 1957, 191–239.

[4] J.W. Alexander, *On the chains of a complex and their duals*, Proc. Nat. Acad. Sci. **21** (1935), 509–511.

[5] J.W. Alexander, *On the ring of a compact metric space*, Proc. Nat. Acad. Sci. **21** (1935), 511–512.

[6] J.W. Alexander, *On the connectivity ring of an abstract space*, Ann. Math. **37** (1936), 698–708.

[7] J.W. Alexander, *A theory of connectivity in terms of gratings*, Ann. Math. **39** (1938), 883–912.

[8] J.W. Alexander, *Gratings and homology theory*, Bull. Amer. Math. Soc. **53** (1947), 201–233.

[9] A. Borel and J.P. Serre, *Détermination des p-puissances réduites de Steenrod dans la cohomologie des groupes classiques. Applications*, C. R. Acad. Sci. Paris **233** (1951), 680–682.

[10] A. Borel and J.P. Serre, *Groupes de Lie et Puissances réduites de Steenrod*, Amer. J. Math. **75** (1953), 409–448.

[11] H. Cartan, *Methodes modernes en topologic algébrique*, Comm. Math. Helv. **18** (1945), 1–15.
[12] H. Cartan, *Algebraic Topology*, Mimeographed notes based on lectures delivered at Harvard University during the spring of 1948. Edited by George Springer and Henry Pollak.
[13] H. Cartan, *Seminaire H. Cartan*, E.N.S. 1st year (1948/1949).
[14] H. Cartan, *Une théorie axiomatique des carrés de Steenrod*, C. R. Acad. Sci. Paris **230** (1950), 425–427.
[15] H. Cartan, *Sur l'iteration des opérations de Steenrod*, Comm. Math. Helv. **28** (1955), 40–58.
[16] E. Čech, *Multiplications on a complex*, Ann. Math. **37** (1936), 681–697.
[17] G. De Rham, *Sur l'analysis situs des variétés à n dimensions*, Jour. de Math. **10** (1931), 115–200.
[18] G. De Rham, *Sur la théorie des intersections et les intégrals multiples*, Comm. Math. Helv. **4** (1932), 151–167.
[19] A. Dold, *Sur les operations de Steenrod*, Bull. Soc. Math. France **87** (1959), 331–339.
[20] C.H. Dowker, *Hopf's theorem for non-compact spaces*, Proc. Nat. Acad. Sci. **23** (1937), 293–294.
[21] C.H. Dowker, *Mapping theorems for non-compact spaces*, Amer. J. Math. **69** (1947), 200–242.
[22] S. Eilenberg, *Cohomology and continuous mappings*, Ann. Math. **41** (1940), 231–251.
[23] S. Eilenberg, *Singular homology theory*, Ann. Math. **45** (1944), 407–447.
[24] S. Eilenberg and N. Steenrod, *Foundations of Algebraic Topology*, Princeton Univ. Press (1952).
[25] S. Eilenberg and N. Steenrod, *Axiomatic approach to homology theory*, Proc. Nat. Acad. Sci. **31** (1945), 117–120.
[26] M. Glezerman and L. Pontrjagin, *Intersections in manifolds*, Uspekhi Mat. Nauk (N.S.) **2** (1947), 58–155 (in Russian); English translation in Amer. Math. Soc. Transl. **50** (1951).
[27] H. Hopf, *Zur Algebra der Abbildungen von Mannigfaltigkeiten*, J. Reine Angew. Math. **163** (1930), 71–88.
[28] H. Hopf, *Die Klassen der Abbildungen der n-dimensionalen Polyeder auf die n-dimensionale Sphäre*, Comm. Math. Helv. **5** (1932), 39–54.
[29] W. Hurewicz, *On duality theorems*, Bull. Amer. Math. Soc. **47** (1941), 562–563.
[30] W. Hurewicz, J. Dugundji and C.H. Dowker, *Continuous connectivity groups in terms of limit groups*, Ann. Math. **49** (1948), 391–406.
[31] A.N. Kolmogoroff, *Über die Dualität im Aufbau der kombinatorischen Topologie*, Math. Sbornik **43** (1936), 97–102.
[32] A.N. Kolmogoroff, *Homologiering des Komplexes und des lokal-bikompakten Raumes*, Math. Sbornik **43** (1936), 701–706.
[33] A.N. Kolmogoroff, *Les groupes de Betti des espaces localement bicompact*, C. R. Acad. Sci. Paris **202** (1936), 1144–1147.
[34] A.N. Kolmogoroff, *Propriétés des groupes de Betti des espaces localement bicompact*, C. R. Acad. Sci. Paris **202** (1936), 1325–1327.
[35] A.N. Kolmogoroff, *Les groupes de Betti des espaces métriques*, C. R. Acad. Sci. Paris **202** (1936), 1558–1560.
[36] A.N. Kolmogoroff, *Cycles relatifs. Theoréme de dualité de M. Alexander*, C. R. Acad. Sci. Paris **202** (1936), 1641–1643.
[37] S. Lefschetz, *Topology*, Amer. Math. Soc. Colloq. vol. 12, New York (1930).
[38] S. Lefschetz, *Algebraic Topology*, Amer. Math. Soc. Colloq. vol. 27, New York (1942).
[39] W.S. Massey, *Pontrjagin squares in the Thom space of a bundle*, Pacific J. Math. **31** (1969), 133–142.
[40] L. Pontrjagin, *The theory of topological commutative groups*, Ann. Math. **35** (1934), 361–388.
[41] L. Pontrjagin, *The general topological theorem of duality for closed sets*, Ann. Math. **35** (1934), 904–914.
[42] L. Pontrjagin, *Classification des transformations d'un complexe $(n+1)$-dimensionelle dans une sphère n-dimensionelle*, C. R. Acad. Sci. Paris **206** (1938), 1436–1438.
[43] L. Pontrjagin, *A classification of mappings of a 3-dimensional complex into the 2-dimensional sphere*, Math. Sbornik **9** (1941), 331–363.
[44] L. Pontrjagin, *Mappings of a 3-dimensional sphere into an n-dimensional complex*, Dokl. Akad. Nauk SSSR **34** (1942), 39–41.
[45] L. Pontrjagin, *Classification of mappings of the $(n+1)$-dimensional sphere into a polyhedron K_n whose fundamental group and Betti groups of dimensions $2, \dots, n-1$ are trivial*, Izv. Akad. Nauk Ser. Mat. **14** (1950), 7–44.
[46] H. Seifert and W. Threlfall, *Lehrbuch der Topologie*, B.G. Teubner, Leipzig (1934). English translation by Michael A. Goldman entitled *A Textbook of Topology*, Academic Press (1980).
[47] J.P. Serre, *Sur les groupes d'Eilenberg–Mac Lane*, C. R. Acad. Sci. Paris **234** (1952), 1243–1245.
[48] J.P. Serre, *Cohomologie des complexes d'Eilenberg–Mac Lane*, Comm. Math. Helv. **27** (1953), 198–232.

[49] E.H. Spanier, *Cohomology theory for general spaces*, Ann. Math. **49** (1948), 407–427.
[50] N.E. Steenrod, *On universal homology groups*, Proc. Nat. Acad. Sci. **21** (1935), 482–484.
[51] N.E. Steenrod, *Universal homology groups*, Amer. J. Math. **58** (1936), 661–701.
[52] N.E. Steenrod, *Products of cocycles and extensions of mappings*, Ann. Math. **48** (1947), 290–320.
[53] N.E. Steenrod, *Reduced powers of a cocycle*, Proc. of the International Congress of Mathematicians 1950 I, Amer. Math. Soc., Providence, RI, (1952), 530.
[54] N.E. Steenrod, *Reduced powers of cohomology classes*, Ann. Math. **56** (1952), 47–67.
[55] N.E. Steenrod, *Homology groups of symmetric groups and reduced power operations*, Proc. Nat. Acad. Sci. **39** (1953), 213–217.
[56] N.E. Steenrod, *Cyclic reduced powers of cohomology classes*, Proc. Nat. Acad. Sci. **39** (1953), 217–223.
[57] N.E. Steenrod, *Cohomology operations derived from the symmetric group*, Comm. Math. Helv. **31** (1957), 195–218.
[58] N.E. Steenrod, *The work and influence of professor S. Lefschetz in algebraic topology*, Algebraic Geometry and Topology: A Symposium in Honor of S. Lefschetz, Princeton Univ. Press (1957), 24–43.
[59] N.E. Steenrod, *Cohomology Operations*, Annals of Mathematics Studies vol. 50, Princeton University Press (1962). Written and revised by D.B.A. Epstein.
[60] N.E. Steenrod and E. Thomas, *Cohomology operations derived from cyclic groups*, Comm. Math. Helv. **32** (1957), 129–152.
[61] E. Stiefel, *Richtungsfelder und Fernparallelismus in Mannigfaltigkeiten*, Comm. Math. Helv. **8** (1936), 3–51.
[62] R. Thom, *Classes caractéristiques et i-carrés*, C. R. Acad. Sci. Paris **230** (1950), 427–429.
[63] R. Thom, *Varietés plongées et i-carrés*, C. R. Acad. Sci. Paris **230** (1950), 507-508.
[64] R. Thom, *Espaces fibrés en sphères et carrés de Steenrod*, Ann. Sci. École Norm. Sup. **69** (1952), 109–182.
[65] E. Thomas, *A generalization of the Pontrjagin square cohomology operation*, Proc. Nat. Acad. Sci. **42** (1956), 266–269.
[66] E. Thomas, *The generalized Pontrjagin cohomology operations and rings with divided powers*, Mem. Amer. Math. Soc. **27** (1957).
[67] A. Weil, *Sur les théorèmes de De Rham*, Comm. Math. Helv. **26** (1952), 119–145.
[68] J.H.C. Whitehead, *On simply connected 4-dimensional polyhedra*, Comm. Math. Helv. **22** (1949), 48–92.
[69] J.H.C. Whitehead, *A certain exact sequence*, Ann. Math. **52** (1950), 51–110.
[70] H. Whitney, *Sphere spaces*, Math. Sbornik **43** (1936), 787–791.
[71] H. Whitney, *On products in a complex*, Proc. Nat. Acad. Sci. **23** (1937), 285–291.
[72] H. Whitney, *On matrices of integers and combinatorial topology*, Duke Math. J. **3** (1937), 35–44.
[73] H. Whitney, *The maps of an n-complex into an n-sphere*, Duke Math. J. **3** (1937), 51–55.
[74] H. Whitney, *On products in a complex*, Ann. Math. **39** (1938), 397–432.
[75] H. Whitney, *Moscow 1935: Topology moving toward America*, A Century of Mathematics in America, Vol. 1, P. Duren, ed., Amer. Math. Soc., Providence, RI (1988), 97–117.
[76] W.T. Wu, *Classes caractéristiques et i-carrés d'une varieté*, C. R. Acad. Sci. Paris **230** (1950), 508–511.
[77] W.T. Wu, *Les i-carrés dans une varieté grassmanniene*, C. R. Acad. Sci. Paris **230** (1950), 918–920.
[78] W.T. Wu, *On squares in Grassmannian manifolds*, Acta Sci. Sinica **2** (1953), 91–115.
[79] W.T. Wu, *On Pontrjagin classes III*, Acta Acad. Sinica **4** (1954), 323–346 (in Chinese, English summary). English translation in Amer. Math. Soc. Transl. Ser. 2 **11** (1959), 155–172.

This Page Intentionally Left Blank

CHAPTER 22

Fibre Bundles, Fibre Maps

M. Zisman

Université Paris 7, UFR de Mathématiques, Tours 45–55, 2 Place Jussieu, 75251 Paris Cedex 05, France

E-mail: zisman@math.jussieu.fr

Fibrations today form one of the basic notions in topology and are considered as a rather simple structure, easy to define and to think about, even if one can read, in a paper by Norman E. Steenrod [55]: *The concept of fibre bundle is somewhat complicated.*

There are essentially two points of view in this topic, both starting from a continuous map $p : E \to B$ where p is called the *projection*, E the *total space*, B the *base* and, for any $x \in B$, $p^{-1}(x)$ the *fibre* over x.

In the first one, finding its origin in differential geometry one requires that p be a locally trivial map. More precisely, these are given an open covering $U = (U_i)_{i \in I}$ of a space B, a space F called the fibre, and homeomorphisms $\varphi_i : p^{-1}(U_i) \to U_i \times F$ over U_i. For any given $x \in U_i \cap U_j$, $i, j \in I$, the formula $g_{ij}(x)(y) = \varphi_j \circ \varphi_i^{-1}(x, y)$, $y \in F$, defines an automorphism $g_{ij}(x)$ of F satisfying the relations $g_{ik}(x) = g_{jk}(x) \circ g_{ij}(x)$ for $x \in U_i \cap U_j \cap U_k$, the relation of a 1-cocycle. It is generally required that the values of the $g_{ij}(x)$ all belong to the same topological transformation group G of F, and that the maps $g_{ij} : U_i \cap U_j \to G$ be continuous. In that case the collection (E, p, B, F, G) is called a fibre bundle with structure group G.

The second point of view results from homotopy considerations. Given a commutative diagram of topological spaces and continuous maps,

$$\begin{array}{ccc} X \times \{0\} & \xrightarrow{\ u\ } & E \\ \big\downarrow & & \big\downarrow{\scriptstyle p} \\ X \times [0,1] & \xrightarrow[\ v\ ]{} & B \end{array} \qquad (*)$$

does there exist a continuous map $V : X \times [0, 1] \to E$ such that $p \circ V = v$ and $V|X \times \{0\} = v$? If such a map V always exists for any space X of a given class $\mathcal{X}$ of topological spaces and any maps u and v, then p is said to satisfy the *covering homotopy property* (= CHP) for the class $\mathcal{X}$. With this perspective in mind one would like to define a fibration by a property general enough to get many "interesting" fibrations but also allowing an easy

HISTORY OF TOPOLOGY
Edited by I.M. James

proof of the CHP for a large class of spaces. The most natural idea, viz. to call a fibration a map such that the CHP holds for any space, came the last (1955).

It took many years to allow all these notions to emerge, to clarify among the works of many geometers and topologists. The common ancestor of both points of view consists in the theory of coverings, and almost all authors until 1950 or 1951 will set their hearts on verifying that their definitions include coverings. Already in 1922, Oswald Veblen in §36 of his *Analysis Situs* constructs the universal covering of a complex as a complex on which the fundamental group of the given complex operates, and probably it is possible to find, in particular cases, earlier constructions of this kind, for example in Hermann Weyl's *Die Idee der Riemannschen Fläche* (1913).

The first point of view appeared in the early twenties, but did not reach its present state before the late forties; the second one came later, in 1940. At that time the second world war had begun, making communications between mathematicians almost impossible. As a consequence, the same kind of theories, leading to the same theorems, appeared in different places almost simultaneously.

The Brussels colloquium (1950) devoted to fibre bundles, marks the end of this age during which this concept was being elaborated, and the beginning of the modern era, when it can be studied for itself and through its applications in many branches of mathematics.

1. First approaches

One of the first – if not the first – fibrations appeared in the mathematical world on May 19th, 1879, in a short note by the young Emile Picard, 23 years old at the time, and entitled *Sur une propriété des fonctions entières* [48]. In this brilliant paper, a model of conciseness and accuracy, Picard proved that a holomorphic function, defined on $\mathbb{C}$, and with values in $\mathbb{C}$ with two points removed, is a constant ("little Picard theorem"). Here is his argument, where only the notations and vocabulary have been modernized. Let $\mathcal{H}$ be the Poincaré half-plane. Let $A \subset \mathcal{H}$ be the set of all points of the form gi or $g \exp 2i\pi/3$, where g is running over the modular group $PSL(2, \mathbb{Z})$ acting on $\mathcal{H}$ in the usual way. Finally let $\lambda : \mathcal{H} \to \mathbb{C}$ be the modular function $\lambda = g_2^3/\Delta$, some combination of Eisenstein series. Then the restricted map induced by λ, $\mathcal{H} - A \to \mathbb{C} - \{0, 1\}$ is a covering, and so any continuous map $f : \mathbb{C} \to \mathbb{C} - \{0, 1\}$ has a lifting $\tilde{f} : \mathbb{C} \to \mathcal{H}$, and $\tilde{f}$ is holomorphic if f is. In that case, being holomorphic and bounded, $\exp \tilde{f}$ must be a constant (Liouville's theorem), and so must be f and $\tilde{f}$. Half of the paper is devoted to a proof of Liouville's theorem. Then for having identified a particular property of the function λ – the lifting property for some maps – and having used, in order to prove a great theorem, a typical argument of a theory that would only appear some 50 years later, I suggest we should consider Picard as one of the founder fathers of the fibration concept.

The other founder father is Elie Cartan. In a long series of papers published between 1922 and 1925, [2–7], in relation to differential geometry and connections, the author engaged himself in computations which may seem a little repetitive today but where one can however see how E. Cartan was endeavouring to associate, in a more and more precise way a vector, affine or projective space with each point of a manifold, providing a coherent system of relations in order to link all these spaces together. It is difficult in front of these texts not to see E. Cartan there describing for us, from this already distant past, the fibre bundles in the language of the time, since ours did not exist yet.

As an example, here is what he wrote in his paper *Sur les variétes à connexion affine et la theorie de la relativité généralisée* [4, §28]:

> *Faisons correspondre par la pensée à chaque point m un espace affine contenant ce point et soient* e_1, e_2, e_3 *trois vecteurs formant avec m un système de référence pour cet espace. La variété sera dite à connexion affine lorsqu'on aura défini, d'ailleurs d'une manière arbitraire, une loi permettant de repérer l'un par rapport à l'autre les espaces affines attachés à deux points infiniment voisins quelconques m et m′ de la variété.*

The same idea is found a little further (§29) in a shorter sentence:

> *Les lois de la connexion affine définissent en quelque sorte le raccord des espaces affines tangents en deux points infiniment voisins.*

In another paper [7] he introduced the ancestor of the structure group of a fibre bundle, giving the feeling that he was close to understanding its role in the existence of global structures: the future theorems about the reduction of structure groups. How to understand differently the following sentence?

> *A tout espace à connexion euclidienne on peut attacher un sous groupe du groupe des déplacements euclidiens qui joue, vis à vis de l'espace considéré, un rôle analogue à celui du groupe de Galois d'une equation algébrique. De même en effet ...*

It is true, however that Cartan was still far away from the modern concept of a fibration, and this for many reasons; the main of these being that he always kept a local point of view although the future builders of that new object would have from the very beginning the ambition to construct it globally, another one is that he never distinguished, or not clearly, the local product structure from the extra datum of a connection; both notions are nearly always ambiguously intertwined.

It fell to the next generation to isolate the precise concepts from Elie Cartan's intuitive ideas.

We must note, however, that at the same time, in October 1924, Harold Hotelling submitted a paper [38] where he constructed 3-dimensional manifolds as circle-bundles on a 2-dimensional manifold. His work came from studies on dynamical systems and results by Birkhoff where these manifolds appeared naturally. Using *Heegard's diagrams* the author determined their fundamental group. That approach without explicit topology, very geometric and elegant used beyond doubt some explicit bundles but Hotelling never abstracted any general idea, and the techniques involved are different.

2. The elaboration of a new concept

2.1.1. It is Herbert Seifert who created the term *gefaserter Raum* (fibre space) in his 1932 paper *Topologie drei-dimensionale gefaserter Raüme* [52]. That work, still motivated by the study of 3-dimensional manifolds, contains a definition of fibre spaces, sometimes called *Seifert fibrations*, through seven axioms, entirely self-contained, and perfectly stated. Axiom 1 is nothing but the definition of topological spaces (from neighbourhoods, and satisfying the Hausdorff axiom). The next three are devoted to the definition of 3-dimensional topological connected manifolds, such that any covering by neighbourhoods is reducible to a countable covering. Axioms 5 and 6 provide the manifold with a partition into subspaces, called *fasern* (fibres) homeomorphic to a circle. Axiom 7 is the most original, the newest. It gives a model for neighbourhoods of the fibres. Let $B_2 \times [0, 1]$ be a cylinder in $\mathbb{R}^3$ where B_2 stands for the unit disk of $\mathbb{R}^2 = \mathbb{C}$, and let two relatively prime integers, μ and ν, be given. Then, for any x and y belonging to B_2, one identifies $(x, 0)$ and $(y, 1)$ iff $y = x \exp(2i\pi\nu/\mu)$. Let $S_{\mu,\nu}$ be the quotient space. Axiom 7 states that any fibre has a neighbourhood homeomorphic to a $S_{\mu,\nu}$ for some (μ, ν), through a fibrewise homeomorphism sending the given fibre onto the image of the cylinder axis. Seifert next provides the *Zerlegungsfläche* – that is to say the *space of fibres* – with a topology for which it is a 2-manifold and the natural projection from the given space to that manifold is continuous. That projection is not in general what we now call a fibration: this occurs iff we have $\mu = \nu$ for every fibre. The typical example Seifert gave is the 3-sphere S^3 fibred by the circles defined by the following equations:

$$t \mapsto (z_1 \exp 2i\pi nt, z_2 \exp 2i\pi mt), \qquad t \in [0, 1],$$

where $(z_1, z_2) \in \mathbb{C}^2$ is a point in S^3, i.e. satisfies the relation $|z_1|^2 + |z_2|^2 = 1$, and where m and n are relatively prime integers. The projection onto the space of fibres is a fibration iff $m = n$ and in that case it is nothing but the map introduced in 1931 by Heinz Hopf in his famous paper *Über die Abbildungen der dreidimensionalen Sphäre auf die Kugelfläche*.

Seifert's paper is doubly interesting for the history of fibrations. To begin with, it introduces, explicitly for the first time, a new structure quite close to the one we are studying here, and secondly it provides us with definitions and proofs of great rigour and accuracy, where nothing is left in the shadow – that will not always be the case for some papers in the years to come!

2.1.2. Almost at the same time, in 1931, during a lecture at a colloquium in Leipzig, William Threlfall – he too was motivated by the study of 3-manifolds – constructed such manifolds as the bundle – in our language – of projective lines associated to the tangent bundle of a surface. In the corresponding paper [62] he observed in a short footnote that these manifolds belong to the framework of Seifert fibrations [52]; below (on p. 95) he explained in a long sentence using everyday words what this could mean.

Though the beginning of the paper reveals some kind of bitterness – here is its first sentence:

> *Die Topologie ist noch keine klassische Disziplin wie die Funktionentheorie*

it also measures – unintentionally justifying the preceding opinion – the degree of abstraction allowed at that time. Thus he proves carefully some results which are today considered as perfectly trivial: we shall meet again this phenomenon later. For example when the surface is a torus, Threlfall starts with a definition of the product of two topological spaces and refers to Steinitz for a proof that the torus is homeomorphic to $S^1 \times S^1$. He still needs a long proof to conclude that the associated 3-manifold is homeomorphic to $S^1 \times S^1 \times S^1$ and that its fundamental group is $\mathbb{Z}^3$.

Near the end of the paper, Threlfall indicates that he was told by Hopf that Hotelling had been at work on a matter very close to his own some eight years before [38]. Communications were still sporadic, and topologists formed a few isolated islets more than an international community. The second world war would soon make the communications still more difficult, a reason, as we shall see later, for some misunderstandings.

2.1.3. In 1935, Seifert brought another contribution to the theory of fibrations in a paper published in 1936 [53], and where he studied differentiable compact submanifolds M^n of class C^2 of $\mathbb{R}^m$. He begins by proving – and his proof is the one we still know today – that such a manifold has a tubular neighbourhood whose boundary – he called it *Umgebungs-Mannigfaltigkeit* – is a sphere-bundle with base M^n. In the case $m = n + 2$ the previous result allows him to prove the existence of a normal unit vector field along the manifold M^n since the sphere bundle in that case has a cross-section, introducing *en route* the present terminology (*Schnitt*). In order to prove that last statement, Seifert assumes the existence of a triangulation on M^n and then he embarks, perhaps for the first time, on an obstruction calculus where the principal argument comes from a duality property strongly related to the fact that M^n is embedded in an Euclidean space.

With Seifert and Threlfall, as with Picard, the total space, the base and the projection, subject to some properties, are given data. It will be Hassler Whitney's and Charles Ehresmann's task to adopt the opposite process – coming from differential geometry – which consists in the construction, from a given base and a given fibre, of a total space according to Elie Cartan's intuition.

2.2. Between 1935 and 1940, Whitney published three papers crucial for the nascent theory. In opposition to Seifert's flawless writing, Whitney's papers look like seminar lectures, or announcements of work in progress. And this is really what those papers are, in which we can see Whithey's ideas developing and acquiring accuracy all along those years.

From the very first lines he gives us, in *Sphere spaces* [63] published in 1935, the fibre bundle philosophy:

> *Locally, sphere spaces are product spaces, but in the large, this may no longer hold.*

In that very short paper, Whitney introduced many fundamental topics among which the characteristic classes soon called *Stiefel–Whitney classes*. But many important details were still remaining in the shadows, and would only be made precise in [64]. It is in [63] that the following classical terminology appeared for the first time: the *base*, the *total space*; *coordinate systems* are defined in [63] but they only got their name in [64]. There also appear the tangent space of a differentiable manifold, the normal space of an embedded manifold, and in particular the case of an n-manifold embedded in $\mathbb{R}^{n+2}$: in that case the second characteristic class of the normal bundle vanishes, so the bundle is trivial. Concerning that last result, as he was proof-reading his paper [53], Seifert, who had made himself acquainted with Whitney's work, observed in a footnote that it was nothing but his own theorem 1, emphasizing however that Whitney did not give any proof! It would appear in [64].

In order to define his characteristic classes, Whitney introduced the space – he denoted it by Q_s^l – of all sequences of s pairwise orthogonal vectors of unit length in $\mathbb{R}^{l+1}$, and stated, without proof, the following lemma (with today's notations for the homology groups and the group of integers):

> *If* $s = 1$ *or* $l + 1 - s$ *is even, then* $H_{l+1-s}(Q_s^l) = \mathbb{Z}$;
> *if* $s > 1$ *and* $l + 1 - s$ *is odd, then* $H_{l+1-s}(Q_s^l) = \mathbb{Z}/2\mathbb{Z}$.

Meanwhile ("Sphere spaces" was submitted to the editors on June 12th and [57] on August 20th) Stiefel finished writing his account of the characteristic classes (for differentiable manifolds). The first third of the paper studies, in "European style" with all the details to be wished for, some spaces precisely called *Stiefel manifolds* soon after its publication. The lemma above can be found in [57, p. 320] (Stiefel named, as we shall do, $V_{n,m}$ the space of m-frames in $\mathbb{R}^n$). Thus in 1935, Seifert, Stiefel and Whitney were really approaching the same kind of problems: Seifert's and Stiefel's approaches were more specific, and Whitney's more universal.

In his next paper [64], published in 1937, Whitney first makes all the topics he introduced in the previous one more precise. No doubt, Elie Cartan's philosophy is at work:

> *To each point $p \in K$ there corresponds a ν-sphere $S(p)$; if $p \neq q$ we assume that $S(p)$ and $S(q)$ have no common points.*

(here K is a complex). But the underlying space in which all these spheres live was left hazy, and it would be the same in [65]: it is only in 1941 that Ch. Ehresmann and J. Feldbau [28] will really make Cartan's idea clear. In the same way the topology of the total space $\mathcal{S}(K)$ – the union of all the spheres $S(p)$ for p running over K – is not explicitly given although everywhere present: we have to wait until [65] to settle that detail. To each closed cell σ in K and each point $p \in \sigma$ is associated a homeomorphism $\xi_\sigma(p): S_0^\nu \to S(p)$ called *σ-coordinate system* where S_0^ν stands for the unit sphere in the Euclidean space $\mathbb{R}^{\nu+1}$. It is implicitly required that the map $\xi_\sigma : \sigma \times S_0^\nu \to \mathcal{S}(K)$ defined by the equality $\xi_\sigma(p,q) = \xi_\sigma(p)(q)$ should be continuous. When p belongs to an intersection of two cells σ and σ', Whitney requires the composite $\xi_{\sigma'}^{-1}(p) \circ \xi_\sigma(p): S_0^\nu \to S_0^\nu$ to be an orthogonal transformation (for $p \in \sigma \cap \sigma'$) and the map $\xi_{\sigma'}^{-1} \circ \xi_\sigma : \sigma \cap \sigma' \to O(\nu+1)$ thus defined, to be continuous (in fact he writes $G^{\nu+1}$ for the orthogonal group $O(\nu+1)$). Owing to the ξ_σ he is able to define orthogonality in the spheres $S(p)$. What he calls a *sphere space* $S(K)$ in [63, 64], and a *sphere bundle* in [65] is the data consisting of K, and for each point p in the cell σ, of the $S(p)$ and ξ_σ subject to the previous conditions: the concepts introduced by Hotelling, Seifert and Threlfall are all sphere bundles.

A few pages in [64, 65] are enough for Whitney to outline many concepts and the main theorems of the nascent theory:

- The *equivalence* between two sphere space [64]: $S(K)$ and $S'(K)$ are equivalent if there exists a homeomorphism $f: \mathcal{S}(K) \to \mathcal{S}'(K)$ such that, for any $p \in K$, the restricted map $f|S(p)$ induces an orthogonal transformation $S(p) \to S'(p)$.
- The *Whitney sum* of two sphere spaces (he called it their product) is defined in [64] and the *duality formula* which gives the characteristic classes of the Whitney sum in terms of the characteristic classes of the given sphere bundles, appears in [65] with the following comment:

  > *The proof is very difficult if $r \geqslant 4$.*

- The sphere space *induced* by a map $f: K \to K'$ is defined in [64] in connection with the next point.
- The *classification* of sphere spaces. After giving the definition of a *universal* sphere space [64], Whitney writes that any sphere space (with base K, let us say) can be obtained as the sphere space induced by a suitable map from X to the base of the universal one. A proof is sketched for 1-sphere spaces.
- The theory of coverings fits into the theory of sphere bundles [65].
- As an example of application of the duality formula, Whitney proves the following theorem: the complex projective plane cannot be embedded in $\mathbb{R}^6$.

Finally, in order to show the richness of these papers, unfortunately difficult to read today, let us point out that between two other brilliant ideas, Whitney introduced the formal series $\sum_{r \geqslant 0} w^r t^r$ (the w^r are the characteristic classes) and computed its inverse: a topic of great promise some 15 years later.

2.3. Advances are now coming from Strasbourg (and Clermont-Ferrand) with Charles Ehresmann and Jacques Feldbau. Already in the early 30's Ehresmann had obtained various results which appeared later as being relevant to fibration theory [20, p. 477]. But it was in May 1939 on the brink of Word War II that Feldbau published his note entitled *Sur la classification des espaces fibrés* [30] explicitly devoted to this topic, and communicated by Elie Cartan, like most of Ehresmann's and Feldbau's papers.

2.3.1. Feldbau who intended to extend Seifert and Whitney's works [52, 64], to the case of a "general bundle" is assuming however as they did, that the total space and the base are manifolds, and that the fibres are compact manifolds, thus revealing his vocation to be a geometer. His definition is very close to Seifert's (a fibration is a space with a partition for which it is locally a product) and the generalisation comes, in addition to the choice of a more general manifold than a circle or a sphere for the fibres, from the introduction of a family of homeomorphisms $H(x): F_x \to F$ from the fibre over a point x in the base into the generic fibre F. This is the first step in the direction of the concept of a structure group for a bundle, the next step will be achieved two years later in [28], where the same notations are kept.

Beyond the previous definitions, [30] contains two important results:

- Theorem A, which states that a bundle whose base is a simplex must be trivial, is cleverly proved, after a suitable subdivision of the base via the following gluing lemma: if A and B are two simplices with a common face, and if a fibre bundle $E \to A \cup B$ is trivial when restricted to A and B, then it is trivial.
- Theorem B which is nothing but the *classification theorem* for bundles with base the n-sphere and fibre F: they are classified by $\pi_{n-1}(\operatorname{Aut} F)$ modulo the action of $\pi_0(\operatorname{Aut} F)$. Feldbau infers at once from theorem B the non vanishing of $\pi_{2n-1}(SO(2n))$ – with today's notations – since the sphere-bundle with fibre S^{2n-1} and structure group $SO(2n)$ associated to the tangent bundle of the sphere S^{2n} is non-trivial.

In fact Feldbau did not use the action of $\pi_0(\operatorname{Aut} F)$ on $\pi_{n-1}(\operatorname{Aut} F)$, nor did he justify the substitution for Aut F of a subgroup G of Aut F, for example $SO(2n)$ when $F = S^{2n-1}$. Both corrections will appear, with partial proofs, in [31], a paper published in 1942, under the name of J. Laboureur, where one can also find interesting results on the parallelisability of spheres, making use of the art of killing the first non-trivial homotopy group of a space. The first complete proof of the "corrected" theorem B is Steenrod's: [56, theorem 18-5, p. 99].

2.3.2. 1941 saw first the beginning of the homotopy theory of fibrations, after the simultaneous discovery of the CHP by five people, and secondly the first general definition of a fibre bundle with structure group.

This last definition appears in [21, 28]. Feldbau was Ehresmann's first research student (there were 76 of them). He co-signed "*Sur les propriétés d'homotopie des espaces fibrés*" [28], he should have co-signed "*Espace fibrés associés*" [21] too, as we learn from his adviser who wrote at the beginning of the note:

> *Les résultats qui vont être exposés sont dus à la collaboration de l'auteur et de l'un de ses élèves.*

After the Occupation by the Nazis of the northern half of France in 1940, the University of Strasbourg had withdrawn to Clermont-Ferrand in non-occupied France. Although in principle not directly under Hitler's power, Maréchal Pétain's government soon passed antisemitic laws, and especially forbade Jewish professors to teach in French universities. Feldbau was probably afraid to see his name at the top of a paper to be published in Paris. In 1942 and 1943 he published in fact under the name of Jacques Laboureur (in German, feldbau means *ploughing*, in French *labourage*, however not a French patronymic; *laboureur* instead (= ploughman) is one). He was arrested in June 1943, and kept prisoner in Drancy before his internment in Auschwitz concentration camp, then in Canaker where he died on April 22, 1945. The interested reader can consult [20], the unnumbered pages following page XXIV, in order to be better acquainted with his short mathematical life.

War still had other consequences, fortunately less dramatic: to suppress or only delay the communications between mathematicians mainly as far as we are concerned, between the United States, France and Switzerland, as we shall soon notice.

2.3.3. But let us go back to mathematics. In the two notes [21, 28], Ehresmann and Feldbau gave a definition of bundles with structure groups which would remain almost unchanged afterwards. They also resolved, finally, the problem sketched in [30, 64]: to construct a bundle with a given base and fibre; it is indeed the objective of the first paragraph of [21] entitled:

I. Méthode de construction d'un espace fibré.

Given two topological spaces B and F, a group G of automorphisms of F, and an open covering Φ of B, they indeed defined an equivalence relation on the disjoint union space $\coprod_{U\in\Phi} U \times F$ by means of maps $t_{U_1U_2} : U_1 \cap U_2 \to G$. These maps, already met in a less satisfactory presentation in [30], took their inspiration from Whitney's coordinate systems ξ_σ and would be called *coordinate transformations* by Steenrod [56]. The maps $(U_1 \cap U_2) \times F \to (U_1 \cap U_2) \times F$ defined by $(x, y) \mapsto (x, t_{U_1U_2}(x)(y))$ are homeomorphisms (by definition), but since G does not carry any topology, there are no questions about continuity for the $t_{U_1U_2}$ at this stage. That is the reason why Steenrod in *The Topology of Fibre Bundles* [56] distinguished his definition, where G is always a topological group, from the one given by Ehresmann–Feldbau, and insisted the two definitions were distinct from each other. However, a few lines down in [21], one assumes the group G to be a topological group, so Steenrod's objection seems aimless. It is true, however, that the passage from discrete to topological group remains rather unclear – as remained unclear in Whitney's papers some aspects in his construction of sphere bundles – but a few years later (and Steenrod admitted this in [56, p. 20]) Ehresmann entirely cleared his definition of bundles in his talk at the 1947 Algebraic Topology colloquium in Paris where he introduced for the first time a *transformation pseudogroup* [25]. It is worth reporting André Haefliger's opinion concerning that concept [34]:

L'élégance et la concision de la présentation contraste avec la lourdeur des premières pages du livre de Steenrod publié en 1951.

2.3.4. Let us come now to the CHP almost simultaneously discovered, in chronological order, by Hurewicz–Steenrod, Ehresmann–Feldbau and B. Eckmann.

The deformation lemma in [28] establishes the CHP for the class of finite complexes and its proof, although only sketched, is quite convincing (a complete proof would appear in [24] for the class of all complexes, finite or not, in 1944). The authors immediately deduced *the exact homotopy sequence of a bundle* from the lemma, or more precisely, they deduced that which took the place of an exact sequence at a time when exact sequences did not yet exist: some subgroups are isomorphic to some quotients. Only one of the three isomorphisms is explicitly given, and the proof that it is really an isomorphism is only sketched, but correctly. Some applications, considered today as being very elementary, are given: homotopy groups of a covering in terms of those of the base, homotopy groups of $P_n(\mathbb{C})$ in terms of those of S^{2n-1}.

Their next paper [21] is mainly concerned with the definitions of associated and principal bundles. After introducing the notion of isomorphisms between two bundles (without expressing the corresponding conditions in terms of cocycles), they prove that two bundles with the same group, base and fibre are isomorphic iff so are their associated principal bundles: in particular a bundle is isomorphic to a trivial bundle iff the associated principal bundle has a section (sections are called *système continu de représentants* in this paper. Ehresmann only introduced the word *section* in 1944).

2.3.5. The important problem of reducing the structure group of a bundle (Ehresmann spoke about *recherche d'une structure plus précise*) was discussed a year later in [22]. Here we find all the classical results of the theory including the reduction of that problem to the search for a section of a suitable bundle, a problem Ehresmann knows how to solve in the case when the homotopy groups of the fibre vanish up to degree $n-1$, if the dimension of the base is n (after Stiefel [57], he said). In May 1943 he applied all these results to the study of differentiable manifolds; we find in [23] a very elegant definition of these manifolds making use of *local charts* and *atlases*, and the definition of the *tangent bundle*; former results [22] allow him to prove that a differentiable manifold always has a Riemannian structure. He also shows that neither S^4 nor $P_4(\mathbb{R})$ could be a universe for Einstein's general relativity: a result which must have impressed the academic world at the time.

2.3.6. Such are Ehresmann's contributions to the theory of fibrations during the years 1941–1944. Later, in 1947, he would prove [26] that a submersion from a compact manifold onto another is a fibration (Hurewicz and Steenrod [43] published a very similar theorem in 1941) and the same year in [25] he would pose the problem of knowing when there exists an almost-complex structure on an even dimensional differentiable manifold. Making use of his knowledge concerning the reduction of the structure group he proves that S^4 does not carry such a structure, but that S^6 does. A. Kirchoff would soon give an explicit almost complex structure on S^6, using Cayley numbers.

Ehresmann's and Feldbau's work clarified many fundamental notions concerning bundles, and brought to our knowledge many basic and classical theorems. However, they suffered – in addition to the French isolation during the war – from a lack of topological tools: the compact open topology for example. There was also a lack of rigour in some arguments, as when Ehresmann applies the CHP to more general spaces than those for which he proved it to hold. Things are still not completely settled in the theory of fibrations.

2.4. On June 17th 1941, a few days after the day [28] was published (June 4th), Beno Eckmann communicated his dissertation entitled *Zur Homotopietheorie gefaserter Räume*. We are now leaving the underlying differential geometry present until now in all works on fibrations, to join homotopy theory; more precisely we are leaving Elie Cartan's school (for, perhaps, Heinz Hopf's school).

2.4.1. Eckmann was endeavouring to compute as many homotopy groups of spaces provided by geometry as possible: spheres, Stiefel manifolds, orthogonal groups, and he was going much further than the Strasbourg mathematicians in this direction, the latter giving the feeling, by comparison, that they only wanted to show that their theory was able to work. Eckmann was also trying to solve some problems, related to these computations, like, for example, under which conditions on an odd integer n does (or does not) the n-sphere admit a continuous tangent 2-field?, and he proved that the 5-sphere does not admit such a 2-field.

Since the spaces he was working with were all compact metric spaces, Eckmann restricted himself naturally to a theory involving only such spaces. He was also looking for a theory able to give easily a theorem on the CHP which is at the origin of almost all his results. In consideration of which, he propounded a nice definition of a kind of fiberings – he called them *retrahierbare Zerlegungen* – fitting his purpose perfectly (Ehresmann translated in [24] this German name into partition contractible: *partition retractable* would be better). For him a fibering is the data consisting of two compact metric spaces R and Z (R for *Raum* = space and Z for *Zerlegung* = partition) and of a surjective continuous map $P : R \to Z$, the projection. Let ρ stand for the metric in Z and for the Hausdorff metric on the space of compact subspaces of R. Then one assumes that for any pair (A, B) of points in Z the equality $\rho(A, B) = \rho(P^{-1}(A), P^{-1}(B))$ holds (it is good to think of the points $A \in Z$ as the elements of a partition of R given by the closed subsets $P^{-1}(A)$). It is assumed moreover that for any point $A \in Z$ there is given a retraction $Q(A) : U(P^{-1}(A), r) \to P^{-1}(A)$, ($r$ is a fixed real positive number and $U(K, s)$ is the open neighbourhood of all points of R at a distance from K less than r). Let $E \subset R \times Z$ be the set of pairs (b, A) satisfying the relation $\rho(P(b), A) < r$. For such a pair one gets $b \in U(P^{-1}(A), r)$ and consequently $Q(A)(b)$ is well defined. One requires that not only the map $Q(A)$ be continuous but also the map $Q : E \to R$ defined by $Q(b, A) = Q(A)(b)$. If E' is the subspace of E such that $(b, A) \in E'$ iff $\rho(P(b), A) \leqslant r' < r$, then $Q|E'$ is now uniformly continuous.

2.4.2. With these definitions, Eckmann proves very carefully the CHP for all compact spaces [19, Lemma 3.d, p. 155] and deduces from it a perfect flawless proof of what stood for the homotopy exact sequence of a fibration, as in [28]. It is nothing else, he said, than the natural generalisation (*Weitgehende Verallgemeinerung*) of Hurewicz's theorem connecting the homotopy groups of a Lie group G, of a closed subgroup $H \subset G$, and of the quotient G/H. Indeed we can find, in the first paper Hurewicz wrote in 1935 about homotopy groups, such a theorem in [41, pp. 118–119], stated without proof and in a way which Ehresmann and Feldbau would hardly change in [28].

About the previous projection $G \to G/H$ Eckmann pointed out, without any real proof, that it is a "retrahierbare Zerlegung", referring to [41] where, he said, Hurewicz stated a property analogous to the "retrahierbare Zerlegung" – and that is true (conditions α and β [41, p. 116]) – but in [41] there are no proofs! In [43], published before [19], there is a "minimum" proof of an analogous result. Ehresmann in [22] made use of the same fibration without references. In the very interesting notice about his works which he wrote as he was applying for a professorship at the *Faculté des Sciences de Paris* in 1955, Ehresmann said [20, p. 477] that he proved that result in his thesis (*Topologie de certains espaces homogènes*, Ann. of Math. 35 (1934), 396–443): as a matter of fact there is a short proof in p. 398. As for Steenrod [56] he refers to Chevalley's *Topology of Lie groups* (Princeton University Press, 1948) for the same topic.

Beno Eckmann of course produced many other "retrahierbare Zerlegungen". At first he generalised the previous example, considering fibre bundles whose total space and fibre are sufficiently differentiable manifolds. The proof is only sketched in the general case, but the retraction Q is explicitly given in all the examples he would consider later: the Hopf maps between spheres, the classical fibrations involving Stiefel manifolds.

2.4.3. In a footnote (p. 141) Eckmann pointed out that when he was proof reading his paper it had come to his knowledge that "*Sur les propriétes d'homotopie des espaces fibrés*" [28] contained – he said *ohne Beweis* (= without proof) – many of his own results. He was too harsh since, as we have seen, the proofs in [28] are quite satisfactory, at least for the modern reader who already knows the results, but it is true that there is a huge difference between an announcement draft and a perfectly written paper. On the other hand, Ehresmann never mentioned [19] until 1944 although he went to Zurich for a lecture in 1942 (but Feldbau mentioned it in [31]!).

2.4.4. Before using homotopy groups, Eckmann introduced some extra conditions on the space R, assuming it to be connected and locally contractible (and then Z satisfies the same properties). These hypotheses shed light on a mysterious footnote in [28]:

> *On suppose que tous ces groupes d'homotopie existent.*

As a matter of fact, the only reference in 1941 for these homotopy groups consists in Hurewicz's original paper [41] where it is always assumed that the spaces are metric compact, connected and locally contractible.

2.5. The most innovative paper from that time about homotopy theory and fibrations comes a little earlier than [19, 28]: *Homotopy relations in fibre spaces* [43] by Hurewicz and Steenrod dates from November 1940. For them, a *fibre space* (and it is the first occurrence of these words) is given by a topological space E, a metric space B with distance ρ, and a continuous map $\pi : E \to B$ such that there exists a continuous function ϕ, called a *slicing function*, with values in E, and defined on the following subspace $A \subset E \times B$: $A = \{(x, b) \mid \rho(\pi(x), b) < \varepsilon_0\}$, where ε_0 is a given positive real number. Finally one requires ϕ to satisfy the following two conditions:

(i) $\pi \circ \phi(x, b) = b$ for any $(x, b) \in A$,

(ii) $\phi(x, \pi(x)) = x$ for any $x \in E$.

The examples of fibre spaces given in the paper are, roughly speaking, the same as those we have already seen: product spaces, coverings, submersions from one differentiable manifold onto another (with a proof making use of Riemannian structures on the manifolds, as would be done by Eckmann in a similar situation [19]), the projection of a compact Lie group onto its quotient by a closed subgroup. It is also said (without proof) that Whitney's sphere bundles are fibre spaces. Let us also remark that that concept of fibre space is more general than Eckmann's "retrahierbare Zerlegungen" which suppose both E and B to be compact metric spaces (take $\phi = Q$).

Theorem 1 is nothing but the CHP for *any* space, but where it is assumed that the homotopy $v: X \times [0, 1] \to B$ (notations of diagram *) is *uniformly continuous*, which is not a restriction if X is a compact metric space as in [19]. Its proof is quite easy since the slicing function provides an explicit covering homotopy. Hurewicz and Steenrod first deduced from theorem 1 that all fibres have the same homotopy type if the base is arcwise connected (a result Eckmann also found in [19], but he relegated it to a footnote as if he did not realize its theoretical importance), then, like Eckmann they proved that the Hopf maps $S^3 \to S^2$, $S^7 \to S^4$ and $S^{15} \to S^8$ are essential maps. What a simplification, if one compare this homotopical proof with Hopf's difficult original homological argument!

2.5.1. The homotopy exact sequence of a fibre space is not explicitly given in their paper, but on the other hand, Hurewicz and Steenrod introduce a fundamental ingredient for all future proofs of this exactness: as a matter of fact, theorem 2 asserts that the projection $\pi : E \to B$ induces an *isomorphism* from $\pi_i(E, \pi^{-1}(b_0))$ onto $\pi_i(B, b_0)$ for any $b_0 \in B$. By the way it is in that paper that the relative homotopy groups are defined for the first time and that the restrictive conditions imposed on spaces to define homotopy groups are released (but the authors did not point it out!). Owing to a nice purely homotopic theorem (if F is a proper arcwise connected closed subset of a sphere S^n then there is an isomorphism $\pi_i(S^n, F) \approx \pi_i(S^n) \oplus \pi_{i-1}(F)$), they prove the following direct sum decomposition: $\pi_i(S^n) \approx \pi_i(S^{2n-1}) \oplus \pi_{i-1}(S^{n-1})$ for $n = 2, 4, 8$, and deduce from it some results on the homotopy groups of spheres, including some special cases of Freudenthal's theorem and results also found by Eckmann [19] with a completely different argument. After some computations of homotopy groups concerning the fibration $SO(n-1) \to SO(n) \to S^{n-1}$, they prove that only S^{2n-1} can be a proper sphere space over the sphere S^n (with fibre S^{n-1}). Feldbau would later, 1942, recover the same theorem [31] with an argument based on the parallelisability of spheres (of course [43] was unknown to him).

Finally the paper comes to an end with an example of a fibre space which is not a fibre bundle (a full right-angled triangle projected on one of its sides). For Henri Poincaré[1] a nice function should be regular and be useful in the proof of nice theorems, like the function λ in Picard's theorem. The others were banished from the universe of nice mathematics (see, for example, the entry "*Peano*" by G. Glaeser in *Encyclopaedia Universalis*). What would he have thought of such an example whose interest consists only in showing that two theories are not quite the same?

2.5.2. Recently, at the colloquium "Materiaux pour l'histoire des mathématiques au XX$^{\text{eme}}$ siècle en l'honneur de Jean Dieudonné" held at Nice (January 6–8, 1996) Eckmann gave a lecture where he returns to these old times. In the corresponding paper, entitled *Naissance des Fibrés et Homotopie*, to be published with the proceedings of the colloquium,[2] he recalls how amazing it was, for him and the four other discoverers of the CHP, to work with the new homotopy objects we described above, in order to recover Hopf's theorems and to prove much other interesting new results:

> *le lemme et la suite exacte d'homotopie établis, les resultats tombaient du ciel!*

he writes.

---

[1] "Autrefois, quand on inventait une fonction nouvelle, c'était en vue de quelque but pratique; aujourd'hui on les invente tout exprès pour mettre en défaut les raisonnements de nos pères, et on n'en tirera jamais que cela".

[2] Added in proof: published in the series *Séminaires et Congrès* 3, Soc. Math. France (1997).

2.5.3. *The object of the fundamental Hurewicz–Steenrod definition* wrote Ralph H. Fox at the beginning of *On fibre spaces I* in 1943 *is to state a minimum set of readily verifiable conditions under which the covering homotopy theorem holds.*

Since this definition depends on the metric of the base B, it is not topologically invariant. That is the reason why Fox proposes a purely topological definition as follows. Let $\pi : E \to B$ be the projection, and let us denote – as he does – by $\bar{\pi} : E \times B \to B \times B$ the map $\bar{\pi}(x, b) = (\pi(x), b)$. Finally, let U be a neighbourhood of the diagonal of B. Then he calls a continuous map $\phi : \bar{\pi}^{-1}(U) \to E$ a *slicing function* if it satisfies the same two conditions as above, and π is called a *fibre mapping* relative to U if there exists a slicing function. In the case when the base B is a compact metric space, it is not hard to see that this definition is equivalent to the previous one. Fox's theorem is concerned with the case of a metrizable base and asserts that if $\pi : E \to B$ is a fibre mapping, then, there exists a metric on B for which π is a fibre space. This nice result, important at that time for the development of the fibration concept lost of course a part of its interest when in the late 50's the notion of fibre space advanced once again.

In fact, Fox fell into a paralogism (as one says in Aristotelean logic) that is worth pointing out in order to show how topics very well known today, if not considered as being completely trivial, could at their beginning give rise to hidden traps. Let B be a metric space, and let $\mathcal{C}(B, [0, 1])$ be the space of all continuous functions from B into $[0, 1]$, with the *uniform convergence* topology. In order to construct a continuous function $\Phi : B \to \mathcal{C}(B, [0, 1])$ such that $\Phi(x)(x) = 0$ for $x \in B$ and $\Phi(x)(y) = 1$ if $(x, y) \notin U$ where U is an open neighbourhood of the diagonal of B, Fox applies Urysohn's lemma to get a continuous function $f : B \times B \to [0, 1]$ satisfying the relations $f(x, x) = 0$ and $f(x, y) = 1$ for $x \in B$ and $(x, y) \notin U$, and then he defines Φ by $\Phi(x)(y) = f(x, y)$. But if f is not uniformly continuous, Φ may not be continuous! In fact such a function Φ always exists, but a correct construction is more elaborate than the one above, and uses partitions of unity which did not yet exist in 1943.

2.6. *Whether every fibre bundle is a fibre space is not yet determined* wrote Steenrod the same year in [55]. To answer this question or similar questions with other definitions of fibre spaces would be the task of many topologists after 1950.

2.6.1. In 1943 Steenrod was "only" concerned in [55] with an easier problem. He proved there that a fibre bundle whose base is a normal space satisfies the CHP for the class of compact spaces: it was the first general and precise result in a subject where the CHP was often proved under restrictive hypotheses but applied without worrying about them. The real aim of the paper is however the classification of sphere bundles. Its principle had already been introduced by Whitney [64] as early as 1937, leaving a tremendous task to be accomplished by those who undertook to make the forerunner's ideas precise and generalise them. That is what Steenrod did in this almost self-contained paper which begins with a nice definition of a fibre bundle. But what a surprise to read that Steenrod, like Whitney, Ehresmann or Feldbau does not give the structure group a topology until he specifies it to be a classical group coming from geometry! Another surprise comes from the references at the end which show that communications in 1943 between Europe and the United States were not as bad as one could expect.

2.6.2. Shiing-Shen Chern and Yi-Fone Sun [8] extended in 1949 Steenrod's work by classifying the fibre bundles with one of the following classical groups $SO(n)$, $Gl^+(n)$, $U(n)$, $Sp(n)$ as structure group and whose base is a compact metric ANR space (instead of a complex as in [55]). Like Steenrod they went back to the beginning of the theory and wrote in a very polished style. Curiously they were the first to consider morphisms between bundles which they called *admissible mappings* and which would later be called more simply *bundle maps* by Steenrod [56]. Thus they were able to state and prove a variant of the CHP for bundle maps (Steenrod [56] would call it the first covering homotopy theorem, the usual one being the second) which on its own provides part of the classification: two bundles induced by homotopic maps are equivalent. But when Steenrod imposed a hypothesis on the base of the *target* bundle (it has to be a normal space), here instead there are no hypotheses on that base, but the base of the *source* bundle must be compact. Curiously the proofs are similar, as will be the proof of the generalisation to C_σ-spaces by Steenrod in [56] where he also extends the existence of universal bundles to any compact Lie group.

It is worth remarking that that paper, forgotten by Jean Dieudonné in his *History of Algebraic and Differentiable Topology* 1900–1960, Birkhäuser 1989, contains an interesting observation, which passed unnoticed at the time when it was published. Let $f : B \to B_G$ be the classifying map of a given G-bundle with base B. Then the image of the cohomology ring $H^*(B_G)$ under the homomorphism $H^*(f)$ is a subring of $H^*(B)$. Chern and Sun called it the *characteristic ring* of the bundle. They seemed aware of its importance although unable to prove it is independent of the choice of a universal G-bundle.

2.6.3. A way to prove at little cost that a bundle is always a fibre space consists in localizing the notion of a fibre space. That is what Sze Tsen Hu did [39] in 1949. For him a fibre space over B relatively to a map π is a space X with a continuous map $\pi : X \to B$ such that for any point $b \in B$ there exist an open neighbourhood U of b and a continuous function $\Phi_U : \pi^{-1}(U) \times U \to X$ satisfying the two conditions of slicing functions (see Section 2.5). Then it is clear, and Hu proved it very well after giving an impeccable and simple definition of fibre bundles, that a fibre bundle is a fibre space. Unfortunately what he has just gained, Hu loses immediately since, in order to prove the CHP (for the class of compact spaces and normal base B) he can only resort to Steenrod's proof in [55]. Globally this whole work sounds a little hollow today, although it had a positive effect in its day. William Huebsch [40] used it to prove the CHP for the class of paracompact spaces by means of a difficult transfinite induction, a considerable improvement on Hu's proof (or Steenrod's in [55]). Later E. Fadell [29] was probably inspired by Hu's paper when he proposed an interesting generalisation of fibre spaces.

2.7. During the topology colloquium held in Brussels from the 5th to the 10th of June 1950, and especially concerned with fibre bundles (and in which only European mathematicians took part), when a lecturer recalled the definition of a bundle, it was only to fix the notations. Two of them, Henri Cartan and Charles Ehresmann each with a different aim, defined and used connections in differentiable fibre bundles. Thus, the loop Elie Cartan opened in 1924 was finally closed 26 years later by his own son and one of his students. What a delight for symbol lovers! No doubt a period had come to an end. It was time to bring the new structure out of the specialized reviews; indeed a little later the first book on bundles was published: the famous *Theory of Fibre Bundles* [56] by Steenrod. Despite a few criticisms addressed to it, this book had a considerable influence on the students of the 50's: there they learned, beyond the definition of a bundle, some topology, homotopy and homology, and also a few nice theorems.

3. The modern era

After the notion of fibre bundle and the leading part of the CHP have been well understood, around 1950, the theory of fibrations would progress along two axes: to know the relations between fibre bundles and fibre spaces defined by a CHP-condition thoroughly, and to extend the classification theorems. The notion of *fibre homotopy equivalence*, introduced by René Thom [61] would enrich the research in both directions.

3.1. But before, let us note an interesting presentation of a fibre bundle with structure group by Friedrich Hirzebruch [35] in 1956. If B stands for a topological space (respectively a differentiable manifold, respectively a complex analytic manifold) and if G is a topological group (respectively a Lie group, a complex Lie group), then Hirzebruch denotes by G_c (respectively G_d, G_ω) the sheaf of germs of continuous (respectively differentiable, holomorphic) functions from B into G. With these notations, the homotopy classes of principal bundles with group G and base B are in bijection with the cohomology set $H^1(B, G_c)$, as one easily sees from the definitions and the interpretation of the coordinate transformations g_{ij} as a 1-cocycle. Taking G_d or G_ω in place of G_c, one gets differentiable or holomorphic bundles, notions already considered before, but which find their natural definition here, allowing us by analogy to define, after the same model, for example algebraic bundles. Hirzebruch presents in a few pages the essential part of the theory up to and including Chern and Pontryagin characteristic classes of vector bundles with structure group $U(n)$ or $O(n)$, about which he gives a pleasant axiomatic definition and proves the formulas giving the characteristic classes of a Whitney-sum or a tensor product of two vector bundles using a purely algebraic method, based on Armand Borel's algebraic theorems on spectral sequences [1]. Further, in 1960 Bernard Morin [47] would give in the *Seminaire Cartan* a similar presentation, but self-contained, avoiding having to resort to the difficult theorems on the cohomology of classifying spaces in [1] (however, easier to prove at that time after J.C. Moore).

3.2. Jean-Pierre Serre in his thesis [54], resolutely got out of the notion of a locally trivial map: a fibration for him is a continuous map satisfying the CHP for all finite polyhedra (cubes are sufficient). That condition is enough to imply that all fibres have the same weak homotopy type provided the base is arcwise connected, and that the usual homotopy exact sequence holds. The existence of the homology and cohomology spectral sequences is more difficult to prove: Serre must introduce cubical singular homology and cohomology groups in order to obtain the result; a "readable" proof using the usual singular simplexes would appear much later in 1967, by A. Dress [18]. Serre fibrations considerably enlarge the range of applications of fibrations. To begin with, it is clear that all previously defined fibrations are Serre fibrations: that comes almost from the very definition for those of Eckmann or Hurewicz–Steenrod and Fox and from [28] for fibre bundles. On the other hand, the fibre space of paths starting at a given point and sending such a path to its end point – for which Serre constructed his theory – is the source of countless works in homotopy theory.

3.3. The restriction contained in the definition of a Serre fibration is not strongly constraining since it is easy to prove that these fibrations satisfy the CHP for the class of all CW-complexes (and even a stronger condition as we shall see in Section 3.3.3), already a large class of topological spaces. On the other hand, what can be said about maps $p: E \to B$ satisfying the CHP for all topological spaces, as is the case for the fibre space of paths, the definition of which was recalled above?

3.3.1. At the end of the year 1955, M.L. Curtis [19] gave a criterion for a continuous map $p: E \to B$ to satisfy the CHP for all spaces. Let C be the fibred product of the diagram

$$E \xrightarrow{p} B \xleftarrow{q} B^{[0,1]},$$

where q sends any path to its starting point. Curtis proved that if the map p satisfies the CHP for C, then it satisfies the CHP for all spaces. Using this result and the work of Huebsch [40], he proved then that if p is a locally trivial map and E and B are metric spaces – hence C also – then p verifies the CHP for all spaces.

3.3.2. Curtis does not seem to know Hurewicz's lecture at the Institute for Advanced Study in January 1954, nor his paper [42] published in August 1955. The subject of both was precisely the study of those maps for which the CHP for all spaces holds, and soon called *Hurewicz fibrations* or, even shorter, *fibrations*. In fact, Hurewicz's definition is slightly different, although equivalent to the former. He is using the same fibred product as Curtis's – probably one of the first fibred products in topology but this name does not appear yet – that he called Ω_p. Let $\tilde{p}: E^{[0,1]} \to \Omega_p$ be the map which assigns to each path τ of E the pair $(\tau(0), p \circ \tau)$; then Hurewicz called the triple (E, B, p) a *fibre space* if the map $\tilde{p}$ has a section (called a *lifting function*). It is necessary of course to show that the class of these fibre spaces is rich enough, in order to give some value to the previous definition. And that is the case, since the property to be a Hurewicz fibration is a local property, provided a suitable condition on the base holds. To be more precise, let us call, as Hurewicz does, an open covering $U = \{U_r\}$ of B *normal* if, for any r there exists a continuous function $f_r : B \to [0, \infty[$ such that $U_r = f_r^{-1}(]0, \infty[)$. Then he proves the following fundamental theorem: suppose that B admits a locally finite normal open covering $U = \{U_r\}$, such that for any r, the map $p^{-1}(U_r) \to U_r$ induced by p, is a Hurewicz fibration; then p is a Hurewicz fibration. As a corollary he got his *uniformization theorem*: if p is locally a fibre space over a paracompact base, then it is globally a fibre space. Thus every locally trivial map over a paracompact space is a Hurewicz fibration, and every fibre bundle (over any base) satisfies the CHP for the class of paracompact spaces: we recover Huebsch's theorem [40], with an easier and more conceptual proof.

The proof of the theorem is beautiful and would serve as a pattern for all those on similar questions which would come later: let us only note the use of a partition of unity associated to an open covering (even if this name does not appear in the paper).

Hurewicz also brought an elegant solution to the comparison between all previously defined kinds of fibrations: when the base is a compact metric ANR, then the fibrations of Hurewicz–Steenrod [43], Fox [32], Hu [39] and Hurewicz [42] are all the same.

3.3.3. A homotopy $v : X \times [0, 1] \to B$ is said to be stationary (for example, in [55]) on a subspace $A \subset X$ if $v(x, t)$ is independent of $t \in [0, 1]$ for any $x \in A$. It suffices to read a little carefully the proofs of the CHP for fibre bundles to allow oneself to be convinced that if the homotopy v, stationary on A, can be lifted, then there exists a lifting stationary on the same subspace A. This property is not true anymore for Hurewicz fibrations, so that Hurewicz introduced the concept of a *regular fibration* which is a fibration for which this property is always true. When the base B is a metric space, he proved that every Hurewicz fibration is regular. The proof of the uniformization theorem shows that the property of being a regular fibration is also a local property.

More generally let us consider a commutative diagram (in continuous lines) where $\mathcal{X}$ is a subspace of $\mathcal{Y}$:

$$\begin{array}{ccc} \mathcal{Y} & \xrightarrow{u} & E \\ \big\downarrow & \overset{V}{\nearrow} & \big\downarrow p \\ \mathcal{X} & \xrightarrow{v} & B \end{array}$$

and let us call (if there exists) *diagonal map* a map V such that $V|\mathcal{Y} = u$ and $p \circ V = v$. In particular, if A is a closed subset of X, $\mathcal{X} = X \times [0, 1]$ and $\mathcal{Y} = X \times \{0\} \cup A \times [0, 1]$ then if v is stationary on A, a diagonal $V : X \times [0, 1] \to E$ would be stationary on A. In 1952 I.M. James and J.H.C. Whitehead [44] showed that such a diagonal always exists if p is a Serre fibration, X a CW-complex and A a sub complex of X. Arne Strøm [58] in 1966 would extend this result to the case where the inclusion $A \hookrightarrow X$ is any *closed cofibration* and p a Hurewicz fibration; two years later in [59] he proved more generally that the diagonal V always exists, provided that p be a Hurewicz fibration and the inclusion $\mathcal{Y} \hookrightarrow \mathcal{X}$ a closed cofibration and a homotopy equivalence at once. Meanwhile, in 1967, Daniel G. Quillen [51] defined a homotopy theory in general categories axiomatically, by means of what he called a *model category*: a category together with three classes of maps called fibrations, cofibrations and homotopy equivalences, satisfying a list of axioms: one of these is exactly the existence of a diagonal for the diagram above under the hypotheses given by Strøm (one half of axiom M1 in [51]). It was proved later that Quillen's axioms are all satisfied in the category of topological spaces, so that it is a model category [60].

3.4. Let us recall that given two maps $p : E \to B$, $p' : E' \to B$ (= *spaces over* B), a map $f : E \to E'$ is called a *map over* B (and written $f : p \to p'$) if $p' \circ f = p$. A homotopy in the category of maps over B is called a fibre homotopy. Fibre homotopy equivalences are defined in the same way.

3.4.1. It is easy to see that if p and p' are two fibre homotopy equivalent spaces over B, then p might be a Serre or Hurewicz fibration but not p'. That remark motivated Edward Fadell [29] to modify the definition of a *fibre space* in order to obtain an object invariant under fibre homotopy equivalences.

In his interesting work of 1957 he began by proving now classical properties of Hurewicz fibrations: for example all fibres have the same homotopy type if the base is arcwise connected; all arcwise connected components of a fibre have the same homotopy type when the total space is arcwise connected; for a regular fibration, if the fibre $p^{-1}(b)$ is contractible to a point in the total space, then the fibre is naturally an H-space (a similar result was also proved by Sugawara); if the open subspace U of B is contractible in B onto a point $b \in U$, then the induced fibration $p^{-1}(U) \to U$ is fibre homotopy equivalent to the trivial fibration $U \times p^{-1}(U) \to U$. It follows from that last result that a Hurewicz fibration is fibre homotopy equivalent to a locally trivial map provided that, for any point $b \in B$, there exists an open neighbourhood U of b, contractible into b in B (Fadell said then that B was *weak locally contractible*).

That last property led Fadell to call a *fibre space* a triple $p : E \to B$ such that there exists a space F and an open covering $\{U_\alpha\}$ of B such that for any α, the restriction map $p^{-1}(U_\alpha) \to U_\alpha$ is fibre homotopy equivalent to the product bundle $U_\alpha \times F \to U_\alpha$. Then for example a Hurewicz fibration with a weak locally contractible base is a fibre space in the preceding sense. However it is clear from the definition that Fadell fibre spaces are invariant under fibre homotopy equivalences.

In order to prove that these fibre spaces satisfy the homotopy exact sequence of a fibration and the usual spectral sequences, it suffices to prove that they satisfy a suitable modified CHP: indeed it is the case. With the notations of the diagram * one can lift the homotopy $v : X \times [0, 1] \to B$ to a homotopy $V : X \times [0, 1] \to E$ such that $V | X \times \{0\} = u$; but instead of an equality between the two maps $p \circ V$ and v, one has only a homotopy, stationary on $X \times \{0\}$ and $X \times \{1\}$. Fadell proves that theorem only when the space X is compact; a delicate proof, inspired by Steenrod's in [56] but much more difficult. He said that

> *by fitting homotopies together*

one can extend the above result to C_σ spaces. But if, as he also said, for any fibre space p there would exist a Hurewicz fibration q fibre homotopy equivalent to p, then the theorem would be true for every space. It would result from later theorems by Albrecht Dold [13] that it is indeed the case subject to some hypotheses on the base. These theorems show also that Fadell fibrations satisfy his modified CHP for all paracompact spaces without any hypothesis on the base (in Dold's terminology a Fadell fibration satisfies the WCHP for paracompact spaces). Fadell's paper contains also many results on section extension properties under technical hypotheses.

With this work a step was got over, new prospects for future research were opened but it would fall to Dold [13] to define the good concepts and prove the main theorems.

3.4.2. Dold introduced two new topological ingredients in his paper entitled *Partitions of unity in the theory of fibrations* [13], both shedding a new light on the theory of fibrations: the notions of *numerable covering* and of *halo*.

A (not necessarily open) covering $\{V_\lambda\}_{\lambda \in \Lambda}$ of B is called *numerable* if there exists a locally finite partition of unity $\{\pi_\gamma : B \to [0, 1]\}_{\gamma \in \Gamma}$ (called a *numeration* of $\{V_\lambda\}_{\lambda \in \Lambda}$ such that for any $\gamma \in \Gamma$ there exists a $\lambda \in \Lambda$ with $\pi_\gamma^{-1}(0, 1] \subset V_\lambda$. That notion is then very close to the concept of a *normal covering* introduced by Hurewicz [42].

A halo around $A \subset B$ (called *Hof* in German, for example, in [11]) is a subspace V of B such that there exists a continuous function $\tau : B \to [0, 1]$ with $A \subset \tau^{-1}(1)$ and $B - V \subset \tau^{-1}(0)$.

Dold introduced also two new properties for a map $p : E \to B$:

- The *section extension property* (= SEP): p has the SEP if every partial section over $A \subset B$ which can be extended to an halo around A can be extended over B.

- The *weak covering homotopy property* (= WCHP) also called *h-Faserung* in [11] or *weak-fibration*: p has the WCHP (for X) if it has the usual CHP for every homotopy $v : X \times [0, 1] \to B$ such that $v(x, t) = v(x, 0)$ for any $t \in [0, 1/2]$. That property is equivalent [13, Proposition 5.13] to the following: every homotopy $v : X \times [0, 1] \to B$ can be lifted to a homotopy $V : X \times [0, 1] \to E$ such that the equality $p \circ V = v$ holds and the two maps $V | X \times \{0\}$ and u are homotopic over B. That property involves the CHP in the sense of Fadell and is stable under fibre homotopy equivalences.

3.4.3. In a first series of theorems, Dold shows that the SEP, the CHP and the WCHP are local properties: if they are satisfied for all maps $p^{-1}(V_\lambda) \to V_\lambda$ induced by p, where $\{V_\lambda\}_{\lambda \in \Lambda}$ is a *numerable covering* of B, then they are satisfied by p also. If the maps $p^{-1}(V_\lambda) \to V_\lambda$ satisfy the CHP or the WCHP for all V_λ of an *open* covering of B, then p satisfies the same condition for the class of paracompact spaces. Finally if $p : E \to B$ and $p' : E' \to B$ are two spaces over B, then a continuous map $f : p \to p'$ over B is a fibre homotopy equivalence, provided the restricted maps $p^{-1}(V_\lambda) \to p'^{-1}(V_\lambda)$ are fibre homotopy equivalences over V_λ for every element V_λ of a numerable covering of B.

All these results, proved by starting from the SEP – the simplest case – and reducing the other to the first, recover all previous results on similar topics, unifying them and giving them the right degree of generalization.

3.4.4. In a second series of theorems, in a completely new spirit, A. Dold gives conditions for a morphism $f : p \to p'$ between two maps satisfying the WCHP over the same base B to be a fibre homotopy equivalence, justifying, if there would be need, the choice of the WCHP to weaken when necessary the notion of fibration.

Already in 1955, Dold had studied a similar problem [12], and proved the following theorem: let $p : E \to B$ and $p' : E' \to B$ be two *fibre bundles* with locally compact fibres, over the same *polyhedron* B; let $f : p \to p'$ be a continuous map over B (and not a bundle map!); then if for every point $x \in B$ the map $p^{-1}(x) \to p'^{-1}(x)$ induced by f is a homotopy equivalence, f itself is a fibre homotopy equivalence. Hypotheses on the fibres came from the proof making use of the functional spaces $F^{F'}$ and F'^{F} (with the compact open topology) and of the adjunction of the two functors exponentiation and product. A few years later in [12] he states the two following theorems where, with the same notations as above, p and p' are now satisfying the WCHP. The first is called *fundamental theorem of homotopy theory* in [11], the second is a natural generalisation of the above theorem in [12]:

- The map f over B is a fibre homotopy equivalence iff, considered as a map $f : E \to E'$ it is an ordinary homotopy equivalence.
- If B has a numerable covering $\{V_\lambda\}_{\lambda \in \Lambda}$ with each V_λ contractible to a point in B, then f is a fibre homotopy equivalence iff for any $x \in B$, the map $p^{-1}(x) \to p'^{-1}(x)$ induced by p is a homotopy equivalence.

3.4.5. Between 1955 and 1958 B. Eckmann and P. Hilton had popularized through many colloquia and lectures at seminars the idea of a duality in the category of topological spaces which allows, starting from a given notion to get another one by reversing the arrows in diagrams, by changing the functor "product by a space" to the functor "exponentiation by this space", and vice versa: so that the reduced suspension corresponds to the loop space, the reduced cone to the space of paths with a fixed origin and fibrations to cofibrations.

That duality is at the origin of many researches just by trying to prove propositions known to be true in the dual situation.

One of the first haloes met in topology, long before this word appeared, comes from the following context. Let (Z, z_0) be a pointed space and let $[0, 1]$ be pointed by 0. Dieter Puppe [49] in 1958 denoted by $\check{Z}$ the space $Z \vee [0, 1]$ pointed by the image of 1, and he called the point z_0 *nicht ausgeartet Grundpunkt* (= non-degenerate base point) when Z and $\check{Z}$ have the same pointed homotopy type. He introduced that concept since he was able to prove that if (X, x_0) and (Y, y_0) are pointed spaces, both with non degenerate base points, then the cone of the inclusion $X \vee Y \hookrightarrow X \times Y$ and the smash product $X \wedge Y$ have the same pointed homotopy type. It is equivalent to say that the point $z_0 \in Z$ is non degenerate, and that z_0 has a contractible halo [49, Hilfsatz 14, p. 332], or to say that the inclusion $\{z_0\} \hookrightarrow Z$ is *eine h-cofaserung* = h-cofibration [11, Satz 3.13]. However h-cofibrations and h-fibrations – another name for the WCHP – correspond under Eckmann–Hilton duality.

The same mathematicians have often worked on both sides of the duality. So, for example, in a series of lectures on homotopy theory and half exact functors he gave in October 1963 at the University of Amsterdam, Dold decided from the beginning to study both theories simultaneously, after he had given a precise and clear definition of the duality. Later T. tom Dieck, K.H. Kamps and D. Puppe would do almost the same when writing their *Homotopietheorie* [11]. Let us note also that in Quillen's *Homotopical Algebra* [51], fibrations and cofibrations are not only simultaneously defined, but also each through the other. In order to show the influence a theory may have on its dual let us indicate the following result [15] where the HEE = *Homotopieerweiterungseigenschaft* stands for the dual of the CHP: let $\{V_\lambda\}_{\lambda \in \Lambda}$ be a numerable covering of the space B, and let $A \subset B$ be a subspace such that for any $\lambda \in \Lambda$ the inclusion $A \cap V_\lambda \hookrightarrow V_\lambda$ has the HEE; then so does $A \hookrightarrow B$.

On the other hand, contrasting with the past, there are more and more exchanges between topologists for the best benefit of topology, exchanges which are made possible at once since it becomes easy to travel from one University to another – at least in the Western World – and the same time, as a consequence of the growing of the number of professorships in Universities after the war, since it occurs that more than only one mathematician is working in the same field in the same University. A good example, as far as topology is concerned, is Heidelberg. For example, during a lecture at Amsterdam, Dold gave a sufficient condition for an inclusion to be a closed cofibration. Talking about that result once back in Heidelberg, he was told by Puppe that in fact the condition is necessary also: that is the way Satz 3.14 in [14] was established.

3.5. The existence of n-universal bundles – that is to say [56] principal bundles with an $(n-1)$-connected total space, was known for all compact Lie groups after the works of Chern–Sun and Steenrod (cf. Section 2.6.2).

3.5.1. In January 1955 John Milnor upset that landscape completely through two short papers [45, 46] where he got results, valid for all topological groups and all n. Serre [54] had constructed a fibration with a given base, a contractible total space and an H-space as fibre. After modifying slightly that construction John C. Moore got for every space X a loop space $\Omega(X)$ provided with the structure of a strictly associative H-space; its singular complex $S\Omega(X)$ then is a simplicial monoid, to which it is possible to apply the $\overline{W}$-construction. He proved in that way the existence of a kind of ∞-universal principal bundle in the *simplicial category* with base $\overline{W}S\Omega(X)$ having the same homotopy type as the singular complex $S(X)$ of X (Seminaire Cartan 1954–1955 exposés 18 et 19). In [45] Milnor was going much further since he proved here the existence of a true ∞-universal fibre bundle with a given base provided that base was a connected countable simplicial complex in the weak topology, or a space of the same homotopy type. In [46] he went back to a more classical approach which consists in the search of an ∞-universal fibre bundle with a given structure group (instead of a given base) and he proved that such a bundle always exists.

What is noteworthy about both constructions of Milnor, is that both are very simple and explicit at once. The group $G(X)$ in [45] having the homotopy type of the loop space of the complex X with base point x_0 consists of the set of all stationary sequences $\{x_0, x_1, \dots, x_n, \dots\}$ such that x_i and x_{i+1} belong to the same simplex for any $i \geqslant 0$, and satisfying the equality $x_n = x_0$ for n big enough, modulo a suitable equivalence relation which ensures that the quotient is a group (for the concatenation of sequences as operation). The total space of the ∞-universal fibre G-bundle is the infinite join [46] $G \circ G \circ \cdots \circ G \circ \cdots$ with a suitable modified topology so that right translations by G is continuous on that space.

Since the constructions are explicitly given it is (relatively) easy to make use of them. So Milnor shows that every principal G-bundle with base a countable complex X, is induced by a continuous homomorphism $G(X) \to G$; he shows also that – if $p_G : E_G \to B_G$ denotes the ∞-universal principal G-bundle – then the E^1 term of the homology spectral sequence with coefficients in a field is given by

$$E^1_{n,q} = \bigoplus_{i_1+\cdots+i_n=q} \widetilde{H}_{i_1}(G) \otimes \cdots \otimes \widetilde{H}_{i_n}(G)$$

a formula where the *bar-construction* appears, and which would be crucial later for the new proofs of Borel's theorems on the cohomology of the classifying spaces by S. Eilenberg and J.C. Moore and for their generalisations to H-spaces (R.J. Milgram, J. Shasheff).

To any H-space H is associated *via* the *Hopf construction* its projective plane $P_2(H)$. If the H-space is homotopy associative it is possible to go further and to construct an extension $P_2(H) \subset P_3(H)$. The more the product of H is homotopically nice, the more is it possible to push further the sequence of these projective spaces $P_2(H) \subset P_3(H) \subset \cdots \subset P_k(H) \subset \cdots$. In fact, the Milnor construction B_G is modelled on that for an infinite projective space $P_\infty(G)$, and so this construction will serve as a pattern for the later generalisations.

3.5.2. We have already seen that Steenrod proved in addition to the existence of an n-universal principal G-bundle (for G a compact Lie group), a classification theorem for principal G-bundles with base a polyhedron of dimension $\leqslant n$, and that Chern–Sun, using a delicate result by Hu extended that theorem to bundles with base any compact ANR space of dimension $\leqslant n$. In Milnor's papers [45, 46] there are no indications that he worked in that direction too, unlike what Jean Dieudonné (op. cit.) asserts who credited him a result close to a theorem which would be stated and proved only eight years later by Dold. As a matter of fact, the end of his paper [13] is devoted to that problem. Adopting Alexander Grothendieck's point of view of representable functors, he defined, for each topological group G, a functor k_G from the category of topological spaces up to homotopy into the category of sets which assigns to any space X the set of isomorphism classes of *numerable* principal G-bundles with base X (a fibre bundle is called *numerable* if there exists a numerable covering $\{V_\lambda\}_{\lambda\in\Lambda}$ of the base such that the restriction of the bundle over V_λ is a trivial bundle for any $\lambda \in \Lambda$). Dold proves that the functor k_G is representable, i.e. that there exists a space B_G, unique up to homotopy, and a functorial bijection (in X) of sets $[X, B_G] \to k_G(X)$. The proof consists of showing that Milnor's ∞-universal G bundle is in fact a numerable bundle and that its total space is contractible (and not only weakly contractible as Milnor proved). The theorems on the SEP allow him to complete the proof. Later he would extend these results in [14].

3.6. In 1957 A. Dold and R. Thom [17] working at the time on infinite symmetric products encounter the following situation. For some continuous map $\varphi: X \to Y$ and a covering of Y by subspaces U, they know that for some U the restricted map $\varphi^{-1}(U) \to U$ is a Serre fibration, but for others all that is known is that the usual exact homotopy sequence of a bundle holds. What can be said about φ? That leads them to define, after a first attempt which H. Cartan thought was too complicated, a *quasi-fibration* as follows: it is a continuous surjective map $p: E \to B$ such that for each point $b \in B$ the map $\pi_i(E, p^{-1}(b), y) \to \pi_i(B, b)$ is an isomorphism, for any $i \geqslant 1$ and any $y \in p^{-1}(b)$; for $i = 0$ one requires to the sequence $\pi_0(p^{-1}(b)) \to \pi_0(E) \to \pi_0(B) \to 0$ to be exact as a sequence of pointed sets. So quasi-fibrations retain only the minimal property one could ask for a fibration in homotopy theory.

The success of that concept comes from three fundamental properties which allow us to construct non-trivial quasi-fibrations: a gluing property, and good behaviour towards a particular kind of deformations and towards direct limits. To prove the last two properties is an easy exercise, but the first constitutes a difficult theorem. These properties allowed Dold and Thom to show that the maps they studied were quasifibrations so that they would apply the exact homotopy sequence to them.

3.6.1. A little after (in 1958) quasifibrations were once more able to prove their efficiency via the *Dold–Lashof construction* [16] which is a generalisation of the classical *Hopf construction* for H-spaces. Dold and Lashof are starting from an H-space H with a two sided unit such that left multiplications are weak homotopy equivalences (which occurs for example when H is arcwise connected), and a quasi-fibration $p: E \to B$ on which H is acting on the right, i.e. there exists a continuous map $m: E \times H \to E$ over B such that the unit of H acts as the identity of E, and such that for any $y \in E$ the map $\{y\} \times H \to p^{-1}(p(y))$ induced by m, is a homotopy equivalence. They construct in that situation, inspired by the Milnor construction [46], a quasi-fibration $DL(p)$ (a notation due to J. Stasheff) and a commutative square

$$\begin{array}{ccc} E & \xrightarrow{f} & E' \\ {\scriptstyle p}\downarrow & & \downarrow{\scriptstyle DL(p)} \\ B & \xrightarrow[\bar{f}]{} & B' \end{array}$$

where f and $\bar{f}$ are inclusions and where f is null homotopic. When, in addition, the H-space H is associative and is acting "associatively" on the quasi-fibration $p: E \to B$, i.e. one has, in addition to the previous conditions, the equality $m(m(y, h), h') = m(y, hh')$

for any $y \in E, h, h' \in H$, then p is called a *principal* H-quasi-fibration. In that case, it turns out that $DL(p)$ itself is a principal H-quasi-fibration, so by iterating the DL-construction infinitely many times and taking the direct limit one gets a principal H-quasi-fibration $DL_\infty(p) : E_\infty \to B_\infty$, and the space E_∞ is weakly contractible. (In fact, the topology of E_∞ is not the direct limit topology: as in [46] that topology must be modified in order to ensure that H acts continuously on E_∞.) In the particular case where we start with the quasi-fibration $H \to pt$, for an associative H-space H such that left translations are weak homotopy equivalences, we obtain *in fine* a principal H-quasi-fibration denoted by $E_H \to B_H$ with a weakly contractible total space E_H. James Stasheff is very enthusiastic with regard to this result, and he writes in *H-Spaces from a Homotopy Point of View* (Lecture Notes in Mathematics vol. 161, Springer, 1970):

> *This theorem is the result of considerable evolution. A restricted version appears in Steenrod's Topology on Fibre Bundles. Such a space B_H was constructed by Milnor for arbitrary topological groups and then by Dold and Lashof in the present generality.*

3.6.2. Let F be a locally compact space, G be a group of automorphisms of F and H be the H-space of all homotopy equivalences of H into itself. The inclusion $G \hookrightarrow H$ is continuous provided the two functional spaces are equipped with the compact open topology. By functoriality one gets a commutative square

$$\begin{array}{ccc} E_G & \xrightarrow{\varphi} & E_H \\ \downarrow & & \downarrow \\ B_G & \xrightarrow[\bar{\varphi}]{} & B_H \end{array}$$

Since it is proved in [16] that for groups the DL-construction and Milnor construction are the same, it then follows that fibre bundles with fibre F structure group G and base X are classified by the set $[X, B_G]$. Considering now two fibre bundles (E, p, X, F, G) and (E', p', X, F, G) with the same base space, fibre and group, classified respectively by f and $f' : X \to B_G$, then, they are fibre homotopy equivalent iff the two maps $\bar{\varphi} \circ f$ and $\bar{\varphi} \circ f'$ are homotopic. In fact, the theorem was proved in [16] only for X a polyhedron since in 1958 the results of [13] were still unknown; in addition the definition of a fibre homotopy equivalence given in [16] is stronger than the usual one, however equivalent if X is a paracompact space (still a consequence of [13]).

As early as 1955 Dold got a similar result in [12] for spheres as base of the bundles. In that case the fibrations are classified via the theorem of Feldbau [30, 31]. Using recent – at the time – results of Serre on homotopy groups of spheres, he produced families of fibrations with base S^m fibre S^n, structure group O_{n+1} fibre homotopy equivalent but not equivalent. The lowest case appears for $m = 4$ and $n = 3$.

3.6.3. Stasheff (op. cit.) states the existence theorem of the principal H-quasi-fibration $E_H \to B_H$ (notations of Section 3.6.1) as the following: under the above hypotheses on H, there exists a space B_H and a weak homotopy equivalence $H \to \Omega B_H$, compatible with the multiplications up to homotopy. Then the statement appears as a particular case of a theorem by Graeme Segal (*Categories and cohomology theories*, Topology **13** (1974), 293–312) about *simplicial spaces*. The connexion between the remark and our study comes from a preprint of that paper, entitled *Homotopy everything H-spaces* where Segal said (without proof) that the proof proceeds from quasi-fibrations and their properties. Soon after, Volker Puppe who wrote at the beginning of his paper [50]:

> *This note reflects part of the attempt to understand the main theorem in G. Segal famous preprint,*

shows, roughly speaking, that one can prove Segal's theorem using maps satisfying the WCHP (called Dold fibration in this paper), which allows him to obtain statements more pleasant and more precise than those using quasi-fibrations. By the way he had to use, following Martin Fuchs [33], a modified Dold–Lashof construction with *Dold fibrations* too, instead of quasi-fibrations.

Bibliography

[1] A. Borel, *Sur la cohomologie des espaces fibrés principaux et des espaces homogènes de groupes de Lie compacts*, Ann. of Math. **59** (1953) 115–207.

[2] E. Cartan, *Sur une généralisation de le notion de courbure de Riemann et les espaces à torsion*, C. R. Acad. Sci. Paris **174** (1922), 593; or *Œuvres Complètes* tome III.

[3] E. Cartan, *Sur un théorème fondamental de Monsieur H. Weyl*, J. Math. Pures Appl. **2** (1923), 167–192; or *Œuvres Complètes* tome III.

[4] E. Cartan, *Sur les variétés à connexion affine et la théorie de la relativité généralisée*, Ann. Ecole Norm. Sup. **40** (1923); or *Œuvres Complètes* tome III.

[5] E. Cartan, *Les espaces à connexion conforme*, Ann. Soc. Math. Polon. **2** (1923), 171–221; or *Œuvres Complètes* tome III.

[6] E. Cartan, *Sur les variétés à connexion projective*, Bull. Soc. Math. France **52** (1924), 205–241; or *Œuvres Complètes* tome III.

[7] E. Cartan, *Les récentes généralisations de la notion d'espace*, Bull. Soc. Math. **48** (1924), 294–320; or *Œuvres Complètes* tome III.

[8] S.S. Chern and Y.F. Sun, *The imbedding theorem for fibre bundles*, Trans. Amer. Math. Soc. **67** (1949), 286–303.

[9] *Colloque de Topologie*, B.C. R.M., Bruxelles (1950).

[10] M.L. Curtis, *The covering homotopy theorem*, Proc. Amer. Math. Soc. **7** (1956), 682–694.

[11] T. tom Dieck, K.H. Kamps and D. Puppe, *Homotopietheorie*, Lecture Notes in Math. vol. 152, Springer (1970).

[12] A. Dold, *Über fasernweise Homotopieäquivalenz von Faserraumen*, Math. Zeit. **62** (1955), 111–136.

[13] A. Dold, *Partitions of unity in the theory of fibrations*, Ann. of Math. **78** (1963), 223–255.

[14] A. Dold, *Halbexacte Homotopie Funktoren*, Lecture Notes in Math. vol. 12, Springer (1966).

[15] A. Dold, *Die Homotopieerweiterungseigenschaft ist eine locale Eigenschaft*, Invent. Math. **6** (1968), 185–189.

[16] A. Dold and R. Lashof, *Principal quasi-fibrations and fibre homotopy equivalences of bundles*, Illinois J. Math. **3** (1959), 285–305.

[17] A. Dold and R. Thom, *Quasifaserungen und unendliche symmetrische Produkte*, Ann. of Math. **67** (1958), 239–281.

[18] A. Dress, *Zur Spectralsequenz von Faserungen*, Invent. Math. **3** (1967), 172–178.

[19] B. Eckmann, *Zur Homotopietheorie gefaserter Räume*, Commentarii Math. Helv. **14** (1942), 141–192.

[20] Ch. Ehresmann, *Œuvres Complètes et Commentédes Amiens* (1984).*

[21] Ch. Ehresmann, *Espaces fibrés associés*, C. R. Acad. Sci. Paris **213** (1941), 762–764 /15/.

[22] Ch. Ehresmann, *Espaces fibré de structures comparables*, C. R. Acad. Sci. Paris **214** (1942), 144–147 /16/.

[23] Ch. Ehresmann, *Sur les espaces fibrés associés à une variété differentiable*, C. R. Acad. Sci. Paris **216** (1943), 628–630 /17/.

* In the following bibliography from [21] until [28] the number between bars /–/ indicates the corresponding reference in the *Œuvres Complètes*.

[24] Ch. Ehresmann, *Sur les applications continues d'une espace dans un espace fibré ou dans un revêtement*, Bull. Soc. Math. France **72** (1944), 27–54 /18/.
[25] Ch. Ehresmann, *Sur la théorie des espaces fibrés*, Colloque de Topologie Algébrique, Paris (1947), 3–15 /20/.
[26] Ch. Ehresmann, *Sur les espaces fibrés differentiables*, C. R. Acad. Sci. Paris **224** (1947), 1611–1612 /22/.
[27] Ch. Ehresmann, *Sur les variétés plongées dans une variété différentiable*, C. R. Acad. Sci. Paris **226** (1948), 1879–1880 /23/.
[28] Ch. Ehresmann et J. Feldbau, *Sur les propriétes d'homotopie des espaces fibrés*, C. R. Acad. Sci. Paris **212** (1941), 945–948 /14/.
[29] E. Fadell, *On fibre spaces*, Trans. Amer. Math. Soc. **90** (1959), 1–14.
[30] J. Feldbau, *Sur la classification des espaces fibrés*, C. R. Acad. Sci. Paris **208** (1939), 1621–1623.
[31] J. Feldbau (alias J. Laboureur), *Les structures fibrée sur le sphère et le problème du parallélisme*, Bull. Soc. Math. France **70** (1942), 181–185.
[32] R.H. Fox, *On fibre spaces*, Bull. Amer. Math. Soc. **49** (1943), 555–557.
[33] M. Fuchs, *A modified Dold–Lashof construction that does classify H-principal fibrations*, Math. Ann. **192** (1971), 328–340.
[34] A. Haefliger, *Ehresmann: un géomètre*, Gaz. Math. **13** (1980), 27–35, and [20] 555–561.
[35] F. Hirzebruch, *Neue Topologische Methoden in der Algebraischen Geometrie*, Ergebnisse der Math., Springer (1956).
[36] H. Hopf, *Über die Abbildungen der dreidimensionalen Sphäre auf die Kugelfläche*, Math. Ann. **104** (1931), 637–665.
[37] H. Hopf, *Über die Abbildungen von Sphären auf Sphären niedriger Dimension*, Fund. Math. **XXV** (1935), 427–440.
[38] H. Hotelling, *Three dimensional manifolds of states of motion*, Trans. Amer. Math. Soc. **27** (1925), 329–344.
[39] S.T. Hu, *On generalizing the notion of fibre spaces to include the fiber bundles*, Proc. Amer. Math. Soc. **1** (1950), 756–762.
[40] W. Huebsch, *On the covering homotopy theorem*, Ann. of Math. **61** (1955), 555–563.
[41] W. Hurewicz, *Beitrage zur Topologie der Deformationen I*, Proc. Akad. Wetensch. Amsterdam **38** (1935), 112–119.
[42] W. Hurewicz, *On the concept of fibre space*, Proc. Nat. Acad. Sci. USA **41** (1955), 956–961.
[43] W. Hurewicz and N.E. Steenrod, *Homotopy relations in fibre spaces*, Proc. Nat. Acad. Sci. USA **27** (1941), 60–64.
[44] I.M. James and J.H.C. Whitehead, *Note on fibre spaces*, Proc. London Math. Soc. **4** (1954), 129–137.
[45] J. Milnor, *Construction of universal bundles I*, Ann. of Math. **63** (1956), 272–284.
[46] J. Milnor, *Construction of universal bundles II*, Ann. of Math. **63** (1956), 430–436.
[47] B. Morin, *Les classes caractéristiques d'un espace fibré à fibres vectorielles*, Seminaire H. Cartan–J.C. Moore 12^{e} année exposé 9 (1959–1960).
[48] E. Picard, *Sur une propriété des fonctions entières*, C. R. Acad. Sci. Paris **88** (1879), 1024–1027.
[49] D. Puppe, *Homotopiemengen und ihre induzierte Abbildungen I*, Math. Zeit. **69** (1958), 299–344.
[50] V. Puppe, *A remark on "homotopy fibrations"*, Manuscripta Math. **12** (1974), 113–120.
[51] D.G. Quillen, *Homotopical Algebra*, Lecture Notes in Math. vol. 43, Springer (1967).
[52] H. Seifert, *Topologie dreidimensionaler gefaserter Räume*, Acta Math. **60** (1932), 147–238.
[53] H. Seifert, *Algebraische Approximation von Mannigfaltigkeiten*, Math. Zeit. **41** (1936), 1–17.
[54] J.P. Serre, *Homologie singulière des espaces fibrés*, Ann. of Math. **54** (1951), 425–505.
[55] N.E. Steenrod, *The classification of sphere bundles*, Ann. of Math. **45** (1944), 294–311.
[56] N.E. Steenrod, *The Topology of Fibre Bundles*, Princeton University Press (1951).
[57] E. Stiefel, *Richtungsfelder und Fernparallelismus in n-dimensionalen Mannigfaltigkeiten*, Commentarii Math. Helv. **8** (1936), 305–353.
[58] A. Strøm, *Note on cofibration*, Math. Scand. **19** (1966), 11–14.
[59] A. Strøm, *Note on cofibration II*, Math. Scand. **22** (1968), 130–142.
[60] A. Strøm, *The homotopy category is a homotopy category*, Arch. Math. **23** (1972), 435–441.
[61] R. Thom, *Espaces fibrés en sphères et i-carrés de Steenrod*, Ann. Ecole Norm. Sup. **69** (1952), 109–182.
[62] W. Threlfall, *Raüme aus Linienelementen*, Jahresber. Deutsch. Math.-Verein. **42** (1933), 87–110.
[63] H. Whitney, *Sphere spaces*, Proc. Nat. Acad. Sci. USA **21** (1935), 464–468.
[64] H. Whitney, *Topological properties of differentiable manifolds*, Bull. Amer. Math. Soc. **43** (1937), 785–805.
[65] H. Whitney, *On the theory of sphere bundles*, Proc. Nat. Acad. Sci. USA **26** (1940), 148–153.

This Page Intentionally Left Blank

CHAPTER 23

A History of Spectral Sequences: Origins to 1953

John McCleary

Department of Mathematics, Vassar College, Poughkeepsie, NY, USA
E-mail: mccleary@vassar.edu

To the memory of Samuel Eilenberg

In December 1946 Princeton University held a conference to celebrate its bicentennial. The sessions on mathematics were titled *"The Problems of Mathematics"*. Princeton enjoyed a leadership role in topology and a problem list [57], prepared by S. Eilenberg (1913–1998), came out of the conference, received at the *Annals of Mathematics* on July 1, 1947 and published in volume 50 (1949). The problems were chosen to give a "birds-eye view of some of the trends of present day Topology". This paper is concerned with a particular entry in this list:

PROBLEM 17. What are the relations connecting the homology structure of the bundle, base space, fiber and group?

Eilenberg mentions briefly recent results of Jean Leray (1906–) announced in the 1946 *Comptes Rendus* [123–126][1] without proof and "indicating interesting new methods".

Less than seven years later, May 3–7, 1953, Cornell University hosted an international conference titled *Fiber bundles and differential geometry*. This conference led to two problem sets in algebraic topology; one [82] was prepared by F. Hirzebruch (1927–) and treats questions of differential topology, especially characteristic classes; the other [141][2] was prepared by W.S. Massey (1920–) and treats questions of homotopy theory. There Massey writes

> It is now abundantly clear that the spectral sequence is one of the fundamental algebraic structures needed for dealing with topological problems.

This paper describes the development of the spectral sequence and its impact on algebraic topology during these years. Because we are discussing an algebraic technique, a

[1] Eilenberg reviewed these notes for the Mathematical Reviews – MR #8,49d; 8,49e; 8,166b; 8,166e.

[2] Received at the *Annals of Mathematics* on February 12, 1955.

HISTORY OF TOPOLOGY
Edited by I.M. James

key role is played by the problems against which these ideas are developed rather than the parallel development of theories where such a technique might apply (contrast [138]). We begin by establishing the background for the development of spectral sequences by tracing the roots of Eilenberg's *Problem 17* and the other lines of research that were affected by the work of Leray, Koszul, Cartan, Serre, and Borel. This discussion splits into three parts, divided conveniently if not naturally, by events before, during, and after the Second World War.

The sketch of algebraic topology before and during the war sets the scene. The problems, prevailing frameworks, and dominant figures determined the audience for the arrival of spectral sequences.

Leray introduced spectral sequences (then called *l'anneau d'homologie*) in 1946. The middle third of this history treats Leray's work[3] leading up to and describing his development of spectral sequences. The reception, transformation, and elaboration of Leray's work by Henri Cartan (1904–) and Jean-Louis Koszul (1921–) is also described in this section.

The last part of the paper, beginning in 1950, concerns the Paris theses of Jean-Pierre Serre (1926–) and Armand Borel (1923–). The changes in algebraic topology brought about by their work demonstrate the considerable depth of Massey's remark. It marks a remarkable moment in the history of topology. The paper closes with a discussion of the spread of these developments.

1. Algebraic topology before the Second World War

Looking back from 1940, the then recent developments in algebraic topology display a remarkable vitality. New homology theories were being developed for larger and larger classes of spaces. The simplicial theory for polyhedra had produced a set of powerful results (e.g., fixed point theorems, Poincaré duality) that set benchmarks for the newer approaches. The theories of J.W. Alexander (1888–1971) [1], E. Čech (1893–1960) [38], and S. Lefschetz (1884–1972) [115] offered successful extensions of the simplicial theory that functioned as tools for new applications of topological ideas.

In 1935–1936 the higher homotopy groups were introduced by W. Hurewicz (1904–1956) [100].[4] A central problem for algebraic topology was (and still is) the computation of these groups for well-understood spaces. Beginning with the first nontrivial example of an essential map between spheres of differing dimension [88], Heinz Hopf (1894–1971) extended his study of linking invariants to obtain nontrivial homotopy classes of mappings $S^{4n-1} \to S^{2n}$ [90]. Further progress was obtained by Hans Freudenthal (1905–) [71] who proved his landmark suspension theorem. However, precious little more was known or even conjectured about the homotopy groups of spaces, even up to 1950 when Hopf addressed the Cambridge, Massachusetts International Congress of Mathematicians [96], asking "Wie kann man einen Überblick über sie gewinnen ...?"

Hurewicz's higher homotopy groups could be seen to be denumerable for a polyhedron by an application of the simplicial approximation theorem. In some cases, such as the space

[3] For more detail, see also C. Houzel's history of sheaf theory [98], and A. Borel's introduction to the forthcoming selected papers of Leray [17].

[4] E. Čech had already defined the higher homotopy groups at the 1932 International Congress of Mathematicians in Zürich. However, these groups were treated then as a mere curiosity.

$S^1 \vee S^2$, the higher homotopy groups are not finitely generated. How the fundamental group influenced such cases and the general question of the higher homotopy groups of a simply-connected space being finitely generated was still open in 1950. Hurewicz had also shown many connections between the homotopy groups and other invariants – the Hurewicz Theorem for homology, the fact that the Hopf fibration $\eta : S^3 \to S^2$ leads to an isomorphism for $i > 2$, $\pi_i(S^2) \cong \pi_i(S^3)$, and a study of aspherical spaces (spaces for which the higher homotopy groups vanish) showing that their homotopy type and homology groups are determined solely by the fundamental group.

In the mid 1930's the newly defined cohomology[5] ring was also developed. Furthermore, Hassler Whitney (1907–1989) had introduced the notion of sphere spaces [201] and had made progress toward a classification of them using characteristic classes, introduced independently by Whitney and E. Stiefel (1909–1978), a student of Hopf [182]. The Stiefel–Whitney classes of manifolds are obstructions to the extension of families of vector fields. The combination of cohomology and homotopy groups (the lower-dimensional homotopy groups of Stiefel manifolds in this case) lies at the heart of seminal work of Eilenberg [53] giving a general obstruction theory for the extension of mappings. The ideas of simple spaces (where the fundamental group acts trivially on the higher homotopy groups), local coefficients (introduced by K. Reidemeister (1893–1971) [152]), Hopf invariants, and the various classification theorems of Hopf [89] and Whitney [203] are all encompassed by Eilenberg's general method.

Outside the centers of Princeton and Vienna, another approach to topological questions was being developed in France and Switzerland during the 1920's and 1930's. Recalling the differential methods of Poincaré, Élie Cartan (1869–1951) published a series of papers [22–24] in which the topology of a Lie group is used to deduce its global analytical properties. In his study of the linear independence of differential forms on a Lie group up to coboundary [22], Cartan conjectured that the resulting numbers ought to be the Betti numbers, that is, combinatorial invariants of a manifold could be deduced from the differentiable structure. In 1931, Georges de Rham (1903–1990) proved Cartan's conjecture [153] establishing differential forms as a subtle tool for the study of algebraic topology. For Cartan the topology of Lie groups could be studied beginning with differential forms on the underlying manifold then restricting to left invariant forms, arriving finally at the exterior powers of the dual of the Lie algebra which represent such forms [23]. Applying these ideas, R. Brauer (1901–1977) carried out the computation of the Betti numbers of the classical groups [21]. The Cartan program of deducing the topology of Lie groups algebraically via the Lie algebra was completed in the work of Weil, Chevalley, Koszul, and Henri Cartan to be discussed below.[6]

2. Some algebraic topology during the war

The interruption caused by the Second World War damped the vitality of research in algebraic topology but in no way did it stop it. Though communication became difficult during the war, considerable advances in topology appeared in these years. In countries at war,

[5] See Massey's article in this volume.

[6] See also the third volume of the series *Connections, Curvature, and Cohomology* by Greub, Halperin and van Stone, Academic Press, New York (1976).

mathematicians were by and large involved in the war effort. Without graduate students and communication with other mathematicians, progress slowed.

Isolated from the war, Switzerland represents a notable exception. Four papers came out of the Swiss school of Hopf and his students during these years that are central to this account.

The first and most important appeared in the 1941 *Annals of Mathematics* when the war had stopped publication of *Compositio Mathematica.* In [92] Hopf introduced a generalization of compact Lie groups, his notion of a *Γ-Mannigfaltigkeit*, which today is called an *H-space.*[7] An H-space (or H-manifold in Hopf's case) is a space endowed with a continuous multiplication and a unit with respect to this multiplication. The main application of this idea is to prove a generalization of a result of L. Pontryagin (1908–1988) [146] who had computed the rational homology of the classical Lie groups. Pontryagin proved that the homology was isomorphic to the rational homology of a product of odd-dimensional spheres by the direct construction of representing cycles for the homology classes together with a study of the mapping induced by the multiplication on homology. Hopf showed that Pontryagin's result followed from the structure of an H-space and not the special case of having a Lie group. To establish his generalization of Pontryagin's theorem, Hopf analyzes the Pontryagin product on homology together with his *Umkehrshomomorphismus* which expresses the dual of the cup product by using Poincaré duality. Hopf was mindful of Cartan's work [24]. The methods he introduced are global and obtain that the rational cohomology of the exceptional groups is that of a product of spheres, a result left open by the case-by-case analysis of Pontryagin and Brauer.[8] Hopf acknowledged that he could not obtain the closed form of the Poincaré polynomials for the exceptional groups in this way, but the corollaries (for example, fixed point arguments) followed directly. This paper is a landmark in algebraic topology. Hopf had shown how to reverse the flow of ideas to go from the topological to the analytic, thus demonstrating the potential of certain fundamental topological and algebraic structures.

The algebra generators of the rational homology ring of an H-space satisfy certain dual generating hypotheses set out by Hopf (his *minimale, maximale Elemente*) who conjectured that the *minimale* elements (primitives) span a generating set. Hans Samelson (1916–) [155] proved Hopf's conjecture by studying the duality between the Pontryagin ring and the intersection product ring. In modern terms, the duality is between indecomposables in cohomology and primitives in homology. Samelson applied these results to study homogeneous spaces G/U for U a closed subgroup of G, a compact Lie group. Viewing the subgroup as a cycle in G, one can ask if it bounds or not. A subgroup U of G is said to be *not homologous to zero in* G (*nicht homolog 0*) if a nonbounding cycle of U does not bound in G. Equivalently, the inclusion homomorphism $U \subset G$ induces an injection on homology $H_*(U) \to H_*(G)$. Samelson proved that if U is not homologous to zero in G, then $H_*(G;\mathbb{Q})$ is isomorphic to $H_*(U;\mathbb{Q}) \otimes H_*(G/U;\mathbb{Q})$.

Though Samelson's theorem treats the homological properties of the fibre space $G \to G/U$, it was not expressed in this generality. Whitney's work on sphere spaces had unified many of the applications of topology to geometric questions and so their structure came to be taken as fundamental. A natural question is to express the relation between the various topological invariants of the base, fibre and total space of a sphere space. A major step in

[7] The terminology *H-space* (for *Hopf space*) is due to Serre [165, p. 476].

[8] Chih-Tah Yen computed the Poincaré polynomials for the exceptional groups using techniques similar to Brauer in 1949 [213].

solving this problem is taken in the thesis of Werner Gysin (1915–) [76], again a student of Hopf in Zürich. Gysin studied the homology structure of a sphere space composed of manifolds via a construction associated to a simplicial mapping, say $f : M \to B$. If ζ is a cycle on B, suppose that $(f^*(\zeta^{pd}))^{pd}$, the cycle Poincaré dual to the image in $C^*(M)$ of the cocycle Poincaré dual to ζ (this is the *Umkehrhomomorphismus* of Hopf) bounds in M. Then $(f^*(\zeta^{pd}))^{pd}$ is in $C_{p+d}(M; \mathbb{Q})$. Let $\partial C = (f^*(\zeta^{pd}))^{pd}$ and define $h(\zeta) = f(C) \in C_{p+d+1}(B; \mathbb{Q})$. If f is a sphere space with fibre S^d, then $h(\zeta)$ is a cycle. The construction of Gysin takes the subgroup of $H_p(B; \mathbb{Q})$ of classes whose Poincaré duals map to zero under f^* to the quotient of $H_{p+d+1}(B; \mathbb{Q})$ by the image of $H_{p+d+1}(M; \mathbb{Q})$ under f_*. Gysin showed that this construction is well-defined and homotopy invariant. It follows that if the correspondence $[\zeta] \mapsto [h(\zeta)]$ is nontrivial, the mapping f is essential. This fact generalizes one of Hopf's constructions of the Hopf invariant [91].

In modern parlance, Gysin had identified a form of the transgression homomorphism which, in the case of sphere spaces, is realized as a long exact sequence called the *Gysin sequence*:

$$\cdots \to H_p(M; \mathbb{Q}) \to H_p(B; \mathbb{Q}) \to H_{p-d-1}(B; \mathbb{Q}) \to H_{p-1}(M; \mathbb{Q}) \to \cdots.$$

Because Whitney's theory of sphere spaces was central to the application of topology to geometric questions, the results of Gysin mark a major advance. Like the fixed point theorems of simplicial homology, Gysin's theorem set a benchmark for any new approach to the computation of the homology of fibre spaces and his paper [76] is among the most cited papers of the post-war research on fibre spaces. The Gysin sequence was taken up after the war by Norman Steenrod (1910–1971) in [176] who gave a cohomological proof. André Lichnerowicz (1915–) gave a de Rham interpretation of the sequence [135] where the Gysin construction is realized as integration along a fibre. In another landmark paper [41], S.S. Chern (1911–) and E.H. Spanier (1921–1996) removed the manifold hypotheses of Gysin and constructed the Gysin sequence for a fibre space of CW-complexes.

The fourth important paper to appear during the war is again by Heinz Hopf [94].[9] Whitney reviewed it for *Mathematical Reviews* stating,

> This paper is, in the reviewer's mind, one of the most important contributions to combinatorial topology in recent years. It gives far reaching results concerning the relations between the fundamental group, the first and second homology and cohomology groups, and the products between these groups, with beautiful and simple methods.

Hopf proved a remarkable relation: For a polyhedron X, let $h : \pi_i(X) \to H_i(X)$ denote the Hurewicz homomorphism; then

$$H_2(X)/h(\pi_2(X)) \cong R \cap [F, F]/[F, R],$$

where $\pi_1(X) \cong F/R$ with F, a free group and R, the relator subgroup for $\pi_1(X)$. This theorem advances to higher dimensions Poincaré's relation, $H_1(X) \cong \pi_1(X)/[\pi_1(X), \pi_1(X)]$. Hopf goes on to extend this relation for spaces with $\pi_i(X) = \{0\}$ for $1 < i < n$, for which

9 In the Proceedings of the University of Michigan Conference of 1940 [208], a two page announcement of the results of this paper appears, dated May 14, 1940, and "arriving from Zürich too late to have been presented at the conference".

the fundamental group determines the quotient $H_n(X)/h(\pi_n(X))$. The manner in which the quotient is determined was left unsaid except in the case of $n = 2$.

These results mark the beginning of the study of the homological study of groups, a subject founded simultaneously during the war by Hopf and Beno Eckmann (1917–) [47] in Switzerland, by Freudenthal [73] in the Netherlands, by D.K. Faddeev (1907–) in the former Soviet Union [68], and by Eilenberg and Saunders Mac Lane (1909–) in the United States. Hopf's work was the basis of the famous wartime collaboration of Eilenberg and Mac Lane whose resulting stream of papers introduced the cohomology of groups, Eilenberg–Mac Lane spaces, the bar construction, and other important notions in algebraic topology. In Section 17 of their paper [60, I], the spaces $K(\Pi, m)$ are introduced to prove that if $\pi_i(X) = \{0\}$ for $1 \leqslant i < m$ and $m < i < r$, then $H_i(X; G) \cong H_i(K(\Pi, m); G)$, where $\Pi = \pi_m(X)$ and $i < r$, and $H_r(X)/h(\pi_r(X)) \cong H_r(K(\Pi, m))$. This generalizes the scheme observed by Hopf to higher dimensions. Thus the homology groups of the $K(\Pi, m)$ measure the failure of the Hurewicz homomorphism to be an isomorphism in these dimensions. The computation of the homology groups of the $K(\Pi, m)$ later becomes an important ingredient in the analysis of homotopy groups.

Outside Switzerland, other developments were published during the war. Early on Hurewicz [100] had appreciated the relations between homotopy groups implied by a mapping with the homotopy lifting property. In an effort to clarify this relation, Hurewicz and Steenrod [102] gave a broad definition of fibre space[10] from which the homotopy lifting property follows. As corollaries of this definition they prove the homotopy equivalence of fibres over different points and the long exact sequence of homotopy groups.[11] The definition is based on the idea of a *slicing function* that satisfies certain properties stated in terms of a metric space structure on the base. The examples given in [102] include covering spaces, the Hopf maps, and homogeneous spaces.

Others sought to isolate the foundations of the bundle structure. In a more geometric context Ch. Ehresmann (1905–1979) and Jacques Feldbau (1914–1945) took as fundamental the notion of a connection on a manifold. They studied Whitney's sphere spaces and extracted the role of the structure group. This led to classification theorems, in particular, the theorem of Feldbau [70] which reduces the problem of classification for bundles over S^n to the computation of the $(n-1)$-st homotopy group of the structural group. Eckmann, another student of Hopf, studied the homotopy lifting property in the case of $G \to G/U$, where U is a closed subgroup of a compact Lie group G [44].

During the war Steenrod carried out a program of research that culminated in his celebrated book, *The Topology of Fibre Bundles* [179]. In [173] he considered tensor bundles over a manifold M. The obstructions to the existence of tensor fields on a manifold are found in an extension of Whitney's theory of characteristic classes to non-Abelian fundamental groups using local coefficients. Steenrod returns to local coefficients in [174] to explore these more sophisticated tools as an extension of ordinary homology.

In [175] Steenrod investigated Whitney's sphere bundles, admitting that the "concept of fibre bundle is somewhat complicated". His definition for this paper follows Ehresmann and Feldbau [51]. He included an interesting section in his opening discussion on the analogy between extensions of groups and fibre bundles and likened the classification of bundles to the classification of group extensions. The main result of the paper is that if B is a

[10] See the paper of M. Zisman in this volume.

[11] Exact sequences were not part of the mathematical parlance of the time, however.

complex and $l \geqslant \dim B$, then the sphere bundles over B are in one-to-one correspondence with the homotopy classes of maps of B to a space M_l^k, the space of great k-spheres on the $(k+l+1)$-sphere. The homotopy theoretic classification of bundles via *classifying spaces* emerges as a major theme in topology after these developments.

We next take up our main story, the work of Leray during and after the Second World War.

3. Leray

Before the war, Jean Leray had made substantial contributions to the mathematical study of fluid dynamics. He had participated in the mathematical circle around Élie Cartan.[12] Among those contributions is a paper [134] written with Julius Schauder (1899–1943) in which they extend the fixed point methods of Brouwer to Banach spaces in order to establish the existence of solutions of certain classes of differential equations.

At the outbreak of the war, Leray was an officer in the French Army. After France was occupied by the Germans he was arrested by the Germans and was taken to an officers' prison camp in Austria, OFLAG XVIIA, where he spent the remaining war years. *Une université en captivité* was organized with Leray as director for which the captors provided library access from the University of Vienna. Leray feared that his specialty of applied mathematics might lead to forced support of the German war effort, and so he admitted only his experience in topology as his focus of research and teaching.

Leray's fixed point work with Schauder [134] utilized an approximation procedure based on the classical Brouwer theorem that, in a limit, proved the desired result for a Banach space. Leray sought to avoid having to go through the simplicial theory and so chose to work solely at the level of the topological: From his course [122]:

> Mon dessein initial fut d'imaginer une théorie des équations et des transformations s'appliquant directement aux espaces topologiques. J'ai dû recourir à des procédés nouveaux, renoncer à des procédés classiques, et il m'est impossible d'exposer cette théorie des équations et des transformations, sans, d'une part, donner une nouvelle définition de l'anneau d'homologie et d'autre part, adapter les raissonnements cités de M. Hopf à des hypothèses plus générales que les siennes.

The results of his work appeared in four *Comptes Rendus* notes in 1942 [118–121] and the complete *Cours de Topologie Algébrique professé en captivité* was published in Liouville's Journal in 1945 [122].

Among the sources Leray acknowledges in his introduction are the papers of de Rham [153], Alexander [2], Kolmogoroff [106], Alexandroff [4], and Čech [38] on cohomology, as well as Hopf's seminal paper on H-spaces [92]. Leray takes the cohomology ring as the fundamental topological invariant of study and names his theory *l'anneau d'homologie*, reserving the term *groupes de Betti* for the homology groups (following Alexandroff and Hopf [5]).

Leray's basic object is the *couverture*.[13] To define a couverture Leray begins with an abstract *complex*, a graded module K over a ring R, required to be finitely generated and

---

[12] Leray had written up Cartan's lectures leading to the book, *La Méthode du Répère Mobile, la Théorie des Groupes Continus et les Espaces Généralisés*, Hermann (1935).

[13] Borel writes in [17]: "I do not know of any translation of *couverture* in the mathematical literature". In his reviews of these papers, Eilenberg uses *cover* for couverture, but this is obviously inadequate. I follow the notation of [17] in this survey of Leray's work.

equipped with a differential d of degree 1. An abstract complex K is made *concrete over a space* X if there is an assignment to each nonzero $k \in K$ of a nonempty subset of X, called the *support of* k, written $|k| \subset X$, and required to satisfy $|d(k)| \subset |k|$. A basic operation on concrete complexes is that of taking a *section* over a subspace of X. Let $F \subset X$ and K a concrete complex over X, then one obtains a new concrete complex FK with supports $|Fk| = F \cap |k|$. It is necessary to modify K to ignore elements of K whose supports do not meet F. Let $K_{X-F} = \{k \in K\colon |k| \cap F = \emptyset\}$. Then K_{X-F} is a submodule of K, closed under the differential. One defines the complex $FK = K/K_{X-F}$. When $F = \{x\}$, we write xK for $\{x\}K$. Given two concrete complexes K and K' over X and a point $x \in X$, consider

$$r_x : K \otimes_R K' \to xK \otimes_R xK'$$

given by the tensor product of the quotients. If h is in $K \otimes K'$, let $|h| = \{x \in X\colon r_x(h) \neq 0\}$. This defines a concrete complex $K \circ K'$, the *intersection* of K and K'.

A *couverture* is a concrete complex K for which all supports are closed and xK is acyclic for all $x \in X$, that is $H^p(xK) = \{0\}$ for all $p > 0$ and $H^0(xK) \cong R$ with generator the *unit cocycle*, $K^0 = \sum_\alpha x^{0,\alpha}$, the sum of the generators of degree zero.

If X is normal, then the collection of all couvertures on X is a differential graded R-algebra with product given by $\circ$. Its cohomology is Leray's *anneau d'homologie* of X, $H(X, R)$. If a couverture has acyclic supports, then $H(X, R)$ may be computed from that of the underlying abstract complex.

Having set up his cohomology theory, Leray turns to applications. Most are the classical theorems of fixed point sort (his *théorie des équations*), thus giving a new way to obtain his results with Schauder [134]. Leray draws particular attention to his ability to prove the main theorems of Hopf on manifolds with multiplication [92] in this context.

As pointed out later by H. Cartan [28] and by Borel [13], for locally compact spaces, Leray's homology ring agrees with Alexander–Spanier cohomology, introduced by Alexander in [2] and extended in the Ph.D. thesis of Spanier [172]. In his 1950 basis-free exposition of couvertures [130], Leray points out that Čech, Alexander–Spanier, singular and de Rham cohomology theories all admit a description as *un anneau d'homologie* (see also [17]).

At the heart of Leray's development of his cohomology theory there is an argument, his *Lemme 2* of No. 4 in [122], in which he proves that the product of a given complex with an acyclic complex has the same homology as the given complex. The lemma shows that the product of point sections of two couvertures $xK \circ xK'$ is again acyclic. This result is key to Leray's proof of the Künneth theorem [122, Theorem 9]. The argument is an induction on the *weight* of a cocycle, which, in this case, is the maximal degree of an element of the acyclic complex. The same argument occurs in four places in [122] and it is the precursor of what will become the underlying structure of a spectral sequence.

Leray's work drew the attention of Henri Cartan who returned to Strasbourg at the end of the war. Cartan, like André Weil[14] (1906–1998), saw the need for the consolidation of the main results of algebraic topology. His first paper on algebraic topology [26] is based on Lefschetz's extension of Čech theory for locally compact spaces. He makes the observation that, after setting up the basic properties of the homology theory, "subsequent results

[14] Weil reports in [190, Vol. 2, p. 527] that Bourbaki was seriously considering a treatise on "topologie combinatoire" in July 1945.

are deduced from these properties alone, thus avoiding any additional use of complexes or assumptions about triangulability".[15] Around this time Eilenberg and Steenrod published their axioms for a homology theory [65] in which the panoply of homology theories developed since Poincaré were seen to compute uniquely the homology of reasonable spaces. This result that ought to have put an end to the development of homology theories in the face of the generality of Lefschetz–Eilenberg singular theory was met with a cool reception by Weil. He [190, Vol. 2, p. 526] reports meeting with Eilenberg in New York in 1944 and discussing the work on axiomatic homology of Eilenberg and Steenrod:

> ma première question fut pour demander si celle-ci rendait compte des théorèmes de de Rham; ma déception fut grande d'apprendre qu'il n'en était rien.

Weil goes on further to record having met Leray in July 1945 and hearing about Leray's "cohomologie à coefficients variables", the next step in Leray's research.

Already in [122, p. 114] Leray had studied the effect of a mapping (*représentation*) $\phi : E' \to E$ at the level of the couvertures. Let ϕ^{-1} be the inverse transformation of ϕ, generally multivalued, and define $\phi^{-1}(k)$ when k is a *forme de* E, that is, a class of an element in a couverture on E. The mapping ϕ effects a change of supports for a concrete complex with $|\phi^{-1}(k)| = \phi^{-1}(|k|)$, and hence, defines a new concrete complex on E'. Though one obtains a complex in this way, a couverture need not go over to a couverture. The generalization of Steenrod [174] of homology to local coefficients offered a model for Leray that could be extended and incorporated into his cohomology theory. The result was a series of remarkably original notes appearing in *Comptes Rendus* in 1946, where Leray introduced sheaves (*faisceaux*[16]) and spectral sequences.

These developments are aimed at the study of fibre spaces, though the methods apply more generally to any continuous mapping of locally compact spaces. A *sheaf* (*faisceau*) $\mathcal{B}$ of modules over a ring R over a space X associates to each closed subset $F \subset X$ a module $\mathcal{B}(F)$, and to each inclusion of closed subsets $F_1 \subset F_2 \subset X$, a homomorphism of modules $\mathcal{B}(F_2) \to \mathcal{B}(F_1)$, subject to certain axioms. The pairing of a sheaf with a couverture gives rise to a complex $K \bigcirc \mathcal{B}$ leading to cohomology with coefficients in the sheaf $H(X, \mathcal{B})$. The forms are expressions $\sum_\alpha b_\alpha k_\alpha$ where $b_\alpha \in \mathcal{B}(|k_\alpha|)$. The main example of Leray associates to a mapping $\phi : E' \to E$, the sheaf of R-modules $F \mapsto H^p(\phi^{-1}(F), R)$ for each closed subspace $F \subset E$. Leray sought to analyze the cohomology of E' as approximated by the cohomology $H(E, \{F \mapsto H(\phi^{-1}(F), R)\})$.

The relation between the approximation and the target is expressed by a sequence of subquotients of the approximation: In [124] Leray described a family of submodules

$$\begin{aligned} 0 = Q_{-1}^{p,q} \subset Q_0^{p,q} \subset \cdots \subset Q_{p-1}^{p,q} \subset P_{r+1}^{p,q} \subset \cdots \subset P_2^{p,q} \\ \subset H^p\big(E, \{F \mapsto H^q(\phi^{-1}(F), R)\}\big). \end{aligned}$$

There also exist submodules

$$0 = E^{-1,p+q+1} \subset E^{0,p+q} \subset \cdots \subset E^{p+q-1,1} \subset E^{p+q,0} = H^{p+q}(E', R)$$

[15] From Steenrod's review, MR#7,138a.

[16] Weyl, in his 1954 ICM address on the work of the Fields Medal winners, Kodaira and Serre, remarked that "Princeton has decreed that 'sheaf' should be the English equivalent of the French 'faisceau'."

along with isomorphisms

$$\Delta_r : P^{p,q}_{p+1}/Q^{p,q}_{q-1} \to E^{p,q}/E^{p-1,q-1}.$$

Leray refers to all of this structure as *l'anneau d'une représentation*. The definition of the modules bears a resemblance to the weight argument of [122]: Classes of total degree $p+q$ may come from classes in $H^m(E, \{F \mapsto H^n(\phi^{-1}(F), R)\})$ for any $m + n = p + q$ and so may be assigned a weight m. Relating the weights to classes of $H^{p+q}(E', R)$ as well as determining the subquotients of the cohomology of E in the sheaf of cohomology rings of the inverse images is the technical triumph that appeared, rather obscurely, in the sketchy *Comptes Rendus* notes.

The immediate applications of this structure that Leray announced in this note are to obtain the Gysin sequence [75] and the results of Samelson [155] on the action of a compact group acting on a sphere. The third and fourth notes treat further applications: the Poincaré polynomial of a mapping, Poincaré duality among the modules $P_r^{p,q}$, product formulas, and the identification of the Gysin homomorphism with Δ_d, and finally, an analysis of the cohomology of a homogeneous space given by a Lie group modulo a subgroup of maximal rank.

Eilenberg[17] reviewed Leray's 1946 *Comptes Rendus* notes very briefly and somewhat cryptically. The next year, however, Leray's work was transformed at the hands of Jean-Louis Koszul and Henri Cartan. In an elegant note in the *Comptes Rendus* of 1947 [107], Koszul extracted from Leray's description of the homology ring of a mapping an algebraic construction that gives rise to all of the structure. He begins with a $\mathbb{Z}/2\mathbb{Z}$-graded differential ring (A, d), together with a decreasing multiplicative filtration

$$\cdots \subset B^p \subset B^{p-1} \subset \cdots \subset A,$$

that is, the collection $\{B^p\}$

(1) satisfies $B^p \cdot B^q \subset B^{p+q}$,
(2) has trivial intersection,
(3) has union all of A, and
(4) the B^p are preserved under the differential.

To such a structure Koszul associates subrings

$$C_r^p = \{x \in B^p \mid d(x) \in B^{p+r}\}$$

and $D_r^p = d(C_r^{p-r})$. Let $C^p = B^p \cap \ker d$ and $D^p = B^p \cap d(A)$. This gives a sequence of inclusions

$$\cdots \supset C_r^p \supset C_{r+1}^p \supset \cdots \supset C^p \supset D^p \supset \cdots \supset D_{r+1}^p \supset D_r^p \supset \cdots.$$

Koszul sets $\mathcal{E}_r^p = C_r^p/(D_{r-1}^p + C_{r-1}^{p+1})$ and $\mathcal{E}^p = C^p/(D^p + C^{p-1})$. The differential induces a differential $d_r : \mathcal{E}_r^p \to \mathcal{E}_r^{p+r}$ on each $\mathcal{E}_r$ and the main result of the note is that $\mathcal{E}_{r+1} \cong H(\mathcal{E}_r, d_r)$. The sequence $(\mathcal{E}_r)$ is called *la suite d'homologies* of A. Furthermore, Koszul identifies a decreasing sequence of submodules of $H(A, d)$, H^p, defined as the

[17] MR#8,49d; #8,49e; #8,166b; #8,166c.

collection of homology classes containing a cycle in B^p and $\mathcal{E}^p \cong H^p/H^{p+1}$. Koszul ends the note with remarks on the morphisms between such filtered differential rings.

In a subsequent note [108], Koszul considered an additional grading on A by degree giving everything in sight a bigrading. The main example is the exterior algebra generated by the dual of the Lie algebra associated to a compact, connected Lie group. When $U \subset G$ is a closed connected subgroup of G, then a filtration may be defined from the left invariant forms on G that come from U. Koszul notes that

> le calcul de la suite d'homologies ($\mathcal{E}_r$) de A donne l'anneau d'homologie à coefficients réels de l'application canonique p de G sur G/U ainsi que sa structure.

He identifies in a footnote $\mathcal{E}_r^{p,q} \cong P_{r-1}^{p,q}/Q_{r-2}^{p,q}$ in Leray's notation. Koszul's main applications of this structure are to the determination of relations among Poincaré polynomials and to the case of U, a one-dimensional subgroup. The explicit decomposition of the cochain complex in this case allowed Koszul to classify via the filtration the possible cases in terms of Lie theory. This theme was explored thoroughly in his thesis [109].

June 26 to July 2, 1947, the *Colloque International de Topologie Algébrique* took place in Paris. The participants included H. Cartan, Leray, Ehresmann, Freudenthal, Hirsch, Hodge, Hopf, Koszul, de Rham, Stiefel, J.H.C. Whitehead, and Whitney. The published papers of Cartan [28] and Leray [127] differ considerably from their delivered talks. By this time Cartan had received a letter from Weil, then in Brazil, in which he sketched his celebrated proof of the de Rham theorem [190, Vol. 2]. Cartan, in his published short report to the conference, offered his criticism of the work of Eilenberg and Steenrod axiomatizing homology:

> Je voudrais néamoins tenter de caractériser ici en quelques lignes la tentative qui avait fait l'objet de mon exposé du 27 juin 1947. ... Cette théorie diffère de celle d'Eilenberg et Steenrod ... sur plusieurs points importants. Tout d'abord, elle ne vise à axiomatiser que la cohomologie au sens de Čech, ou au sens d'Alexander ... au contraire, la théorie d'Eilenberg–Steenrod se donnait pour but d'englober toutes les théories de l'homologie ou de la cohomologie. Précisément à cause de sa généralité, la théorie d'Eilenberg–Steenrod ne pouvait comporter de théorème d'unicité que pour des espaces de nature très particulière; tandis que l'intérêt principal de notre théorie réside dans ses théorèmes d'unicité.

The potential for application of the methods of algebraic topology to questions in analysis rested on having analytic means apparent and accessible at the level of cohomology, as in the de Rham theory. Furthermore, the unique properties of spaces such as manifolds, or polyhedra, leading to duality theorems, are glossed over in the global approach of Eilenberg and Steenrod. Cartan had recognized the potential of Leray's theory of *couvertures* and *faisceaux* and he began a program[18] to clarify the foundations. Between the lecture in Paris and the publication of the proceedings, Cartan gave a course on algebraic topology at Harvard (Spring 1948) in which he presented a version of Leray's complexes with supports. In the proceedings paper Cartan sketched his program, naming the relevant topological structure *carapace*.

Koszul[19] recalls Leray's lecture treating the action of a discrete group on a topological space in which the Cartan–Leray spectral sequence is introduced. He writes: "Cela

[18] In a lengthy footnote in [17], Borel discusses Cartan's various versions of these foundations.

[19] Letter of April 30, 1997.

était tout inattendu et a fait sensation". A short note coauthored by Cartan and Leray [34] sketching this application appeared in the proceedings. Leray's proceedings paper is based on his courses at the Collège de France 1947/48. Already in this paper, he has adopted the algebraic refinements of Koszul [107] and the "perfectionnements" of Cartan on differential graded and filtered algebras. Leray here uses the terminology *l'anneau spectral d'homologie de l'anneau filtré* for the present-day spectral sequence.

Cartan published two important notes in the 1947 *Comptes Rendus* [27]. In them he introduced formally the definition of a filtered differential graded algebra and the associated spectral sequence, here called *une suite de Leray–Koszul.* In the second note he introduced the spectral sequence of a double complex in the case of a group acting on a space to obtain the Cartan–Leray spectral sequence. This spectral sequence has as E_2-term $H^*(G, H^*(A))$, where (A, d) is a differential graded algebra on which G acts, and it converges to $H^*(G, A)$. In the case of a group G acting properly discontinuously on a space E, then the E_2-term is $H^*(G, H^*(E; R))$ and the spectral sequence has *anneau terminal* the associated graded ring for a filtration of $H^*(E/G; R)$. Cartan claims results of Hopf [94], Eilenberg–MacLane [60], Eckmann [47], and Freudenthal [73] as consequences, especially in the case of the fundamental group acting on the universal cover. Of interest in these notes is the emphasis on the algebraic structures, coming from the known cochain algebras, but here presented in considerable generality. Furthermore, the spectral sequence of a double complex is introduced as well as the passage from the homology of a differential graded object for coefficients to coefficients in that differential graded object via a spectral sequence.

Around this time some new tools emerged in the study of algebraic topology, and, in particular, for fibre spaces. The terminology of the exact sequence, found first in Hurewicz [101] and Eilenberg and Steenrod [65], was formalized in [104]. In [189] H.C. Wang (1918–1978) solved another special case of the Eilenberg's *Problem 17*: Let X be a fibre space over S^n with fibre F. Then there are isomorphisms $H_r(X) \cong H_r(S^n \times F)$ when F has dimension less than n, and if F has dimension $n-1$ with $H_{n-1}(F) \cong \mathbb{Z}$, then either $H_r(X) \cong H_r(S^n \times F)$ for all r, or $H_n(X) = \{0\}$, $H_{n-1}(X) \cong \mathbb{Z}/m\mathbb{Z}$, and $H_r(X) \cong H_r(S^n \times X)$ for $r \neq n, n-1$. The integer m is interpreted by Wang at the level of homotopy groups. The result may be summarized in an exact sequence (the Wang sequence)

$$\cdots \to H_{n+k}(X) \to H_k(F) \to H_{n+k-1}(F) \to H_{n+k-1}(X) \to \cdots.$$

The analysis is based on the properties of the homology of pairs.

In a 1948 *Comptes Rendus* note [78], Guy Hirsch (1915–1993) introduced another piece of structure into the study of fibre spaces – the transgression. Hirsch considered the mappings induced by the projection $P : E \to B$ and the inclusion of a fibre $I : F \to E$ on homology with rational coefficients. If a class z in $H_p(F; \mathbb{Q})$ lies in the kernel of I_*, consider a cycle representative for z, say Z, and a chain $\tilde{c}$ on E with $\partial(\tilde{c}) = Z$. The chain on B given by $P_*(\tilde{c})$ is a cycle and the choices made in this process can be rendered irrelevant by considering the class of $P_*(\tilde{c})$ in $H_{p+1}(B; \mathbb{Q})/P_*H_{p+1}(E : \mathbb{Q})$. The homomorphism determined by $(\ker I_* \subset H_p(F; \mathbb{Q})) \to H_{p+1}(B; \mathbb{Q})/P_*H_{p+1}(E : \mathbb{Q})$ is the (homology) *transgression.*[20] Hirsch also relates this homomorphism to dual classes in the cohomology

[20] The term is introduced in [109].

of B and proves that the characteristic cocycle of Whitney [202] and Steenrod [175] is obtained by this construction; this is defined as a cohomology class representing a mapping $H_{h+1}(B;\mathbb{Q}) \to H_h(F;\mathbb{Q})$. A different formulation of characteristic cocycles appears earlier in the paper [39] of Chern in the context of differentiable forms. Chern gives three definitions of *basic characteristic classes*, the third of which is a dual to Hirsch's definition. Both Cartan and Koszul take up the transgression in later work.

In 1948 Claude Chevalley (1909–1984) and S. Eilenberg published [43] an algebraic treatment of the cohomology of Lie groups, taking the Lie algebra as basic object and extracting a cohomology theory for Lie algebras based on Élie Cartan's algebraic characterization [23] of the de Rham complex of left invariant differential forms. Chevalley and Eilenberg sought to place this cohomology theory of Lie algebras in the same framework as Hochschild's cohomology theory of algebras. The underlying complex for this investigation is an exterior algebra with basis the dual of the given Lie algebra, together with a differential. The doctoral thesis of Koszul [109], submitted in June, 1949, takes off from this point to explore the potential of this algebraic rendering of the cohomology of Lie groups for the study of homogeneous spaces. The main sources of Koszul's study, recommended to Koszul by Henri Cartan, are [23] of É. Cartan, [92] of Hopf, and [155] of Samelson.

After giving algebraic derivations of the theorems of Hopf and Samelson, Koszul studied the relative cohomology algebra for pairs of Lie algebras, $\mathfrak{b} \subset \mathfrak{a}$. This was defined by É. Cartan in [23] using the cochains on $\mathfrak{a}$ that are said to be *orthogonal to* $\mathfrak{b}$, that is, in the algebra of linear functions on $\mathfrak{a}$ that vanish on $\mathfrak{b}$. A cochain is called $\mathfrak{b}$-invariant, if, for all $b \in \mathfrak{b}$, the algebra extension of the homomorphism $a \mapsto [b, a]$ vanishes on the cochain. The subalgebra of cochains orthogonal to $\mathfrak{b}$ and $\mathfrak{b}$-invariant make up the relative cochains, and they form an exterior algebra, closed under the differential. This gives the relative cohomology $H^*(\mathfrak{a};\mathfrak{b})$. Koszul observed that there is a filtration of the cochains on $\mathfrak{a}$ by ideals B^p of cochains of degree at least p, decomposable as a product with at least p cochains of degree 1 orthogonal to $\mathfrak{b}$. The subsequent spectral sequence is analyzed from the filtered differential graded algebra (chez Cartan [30]) through E_0 and E_1 to identify, for $\mathfrak{b}$ a reductive subalgebra of $\mathfrak{a}$, $E_2 \cong H^*(\mathfrak{b}) \otimes H^*(\mathfrak{a};\mathfrak{b})$ with *l'algèbre terminale* the associated graded for the filtration of $H^*(\mathfrak{a})$. The immediate application of the spectral sequence is to the case when the inclusion $\mathfrak{b} \subset \mathfrak{a}$ is nonhomologous to zero, that is, when the homomorphism induced by the inclusion, $H^*(\mathfrak{a}) \to H^*(\mathfrak{b})$, is surjective. Koszul proves the analogue of Samelson's theorem [155] for Lie algebras. The next step is the analysis of the general case, when $I^*: H^*(\mathfrak{a}) \to H^*(\mathfrak{b})$ is not onto. In the cadre of spectral sequences, this is the case of nontrivial differentials. Koszul introduces the term *transgressive* for cocycles z of $\mathfrak{b}$ whose image among the cocycles of $\mathfrak{a}$ is a coboundary lying in the relative cocycles for the pair $(\mathfrak{a}, \mathfrak{b})$. The linear mapping $\Psi : (T^p \subset H^p(\mathfrak{b})) \to H^{p+1}(\mathfrak{a};\mathfrak{b})/M_1$, where T^p is the ideal of transgressive elements of degree p and M_1 is the ideal of cohomology classes which contain a cocycle in the invariant coboundaries of the first filtration. Koszul shows that $\ker \Psi = \operatorname{im} I^*$ and that, for a reductive subalgebra $\mathfrak{b}$ in $\mathfrak{a}$, all primitive elements of $H^*(\mathfrak{b})$ are transgressive. In the last few pages of the thesis, Koszul thoroughly analyzes the edge homomorphism thus identifying the image of $H^*(\mathfrak{a})$ in $H^*(\mathfrak{b})$ via the spectral sequence. In a footnote Koszul recalls Hirsch's analogous results [78] on characteristic cocycles.

The powerful use of algebra in Koszul's thesis demonstrated that purely algebraic objects, together with their apparent additional structure and associated objects like the spectral sequence, gave a detailed picture of the workings of geometric phenomena, in particular, the topology of homogeneous spaces. In the framework of Lie groups, where Lie algebras are available, the first nontrivial cases of fibre spaces, homogeneous spaces, had been analyzed. The next step in a more general direction was taken by Cartan.

It is important to recall that the Lie algebra methods work over the field of coefficients $\mathbb{R}$. In the late 1940's, Steenrod introduced the Steenrod squares and reduced powers for singular cohomology with coefficients in $\mathbb{F}_p$, p a prime [176, 180]. By 1951, Wen-Tsün Wu (1919–) and René Thom (1923–) had shown the relevance of the squaring operations for characteristic classes and Steenrod had demonstrated their use in detecting essential mappings. The computation of the mod p cohomology of spaces together with their squaring or reduced power operations for fibre spaces and Lie groups held considerable promise.

Cartan had spent the spring of 1948 at Harvard where he lectured on topology. Once back in Paris, Cartan established his famous seminar,[21] the *Séminaire Cartan* which met from 1948 to 1964. The first three years of topics dealt with algebraic topology. In the first year, the topics were foundational, dealing with simplicial, singular, and Čech theories. The last few lectures of Cartan, numbered 12–17, were not published in the 1955 reissue of the notes by the *Secrétariat mathématique*. These dealt with the theory of sheaves and carapaces, drawing from his lectures at Harvard, but still in a preliminary form for Cartan.[22] Among the first participants of the 1948/1949 seminar were J.-P. Serre, J. Cerf, P. Samuel, J. Dixmier, and J. Frenkel. The second year of the seminar treated fibre spaces and homotopy groups and contains lectures by Serre, A. Borel, A. Blanchard, and Wen-Tsün Wu. The final lectures of this year dealt with Cartan's work on principal bundles and homogeneous spaces, extending the results of Leray, Hirsch, Koszul, Chevalley, and Weil to be discussed below. The lectures of the Cartan Seminars remain among the clearest expositions of certain topics in algebraic topology. The role of this level of exposition is crucial in this account. The atmosphere of consolidation of a growing subject and the wealth of challenging problems open to the initiated made the proceedings of the *Séminaire Cartan* a window of opportunity to a maturing field.

4. The Summer of 1950

The summer of 1950 began with a major conference event in the history of topology, the *Colloque de Topologie (espaces fibrés)*, which took place in Brussels, 5–8 June, organized by Guy Hirsch. This conference provides a glimpse of the state of progress on the problem of the homology of fibre spaces. The proceedings[23] opens with a note of appreciation to Élie Cartan "whose works had opened the way for much of the research presented in the course of the meeting". The published speakers were Hopf and Eckmann from Switzerland, H. Cartan, Leray, Ehresmann, Koszul from France, and Hirsch from Belgium.

In two landmark papers given at this conference Cartan [30] gave his penetrating analysis of the transgression for homogeneous spaces and principal fibre spaces. A *principal fibre space* $E \to B$ with structure group G as fibre, is a G-space E with B as its orbit space. Based in part on Koszul's thesis and work of Chevalley [42] and Weil,[24] Cartan ex-

[21] See [160], volume 3 for a review of the seminar.

[22] See [69], the Ph.D. thesis of Florence Fasanelli for an analysis of the contents of these lectures.

[23] *Colloque de Topologie (espaces fibrés)*, Bruxelles (1950), CBRM, Liège (1951).

[24] See [190] and Weil's endnotes for a version of his unpublished work on this subject.

posed an algebraic formalism on which the structure of the cohomology of a differentiable principal bundle may be founded. This approach takes as fundamental differential graded algebras along with operators on such. The basic construction is the so-called *Weil algebra*[25] which extends the Chevalley–Eilenberg complex $(\Lambda(\mathfrak{g}^*), d)$, the exterior algebra on the dual of the Lie algebra $\mathfrak{g}$, to a tensor product $(\Lambda(\mathfrak{g}^*) \otimes S[\mathfrak{g}^*], \delta)$ with $S[\mathfrak{g}^*]$ the symmetric algebra on $\mathfrak{g}^*$. The differential δ is defined explicitly from the bracket on the Lie algebra. Weil had introduced the Weil algebra to explicate how a connection together with the transgression could relate certain cohomology classes to the Chern classes in the particular case of unitary bundles. Chevalley had shown, using Koszul's work, that primitive elements in the Weil algebra were transgressive in the reductive case. Cartan elegantly set out this framework in his papers. If $\Omega^*(E)$ denotes the differential forms on E, then the invariant forms in $\Omega^*(E)$ determine its cohomology and furthermore, the basic elements of $\Omega^*(E)$ determine the cohomology of B. (Cochains are basic if they are invariant and also vanish under interior products with all elements of $\mathfrak{g}$.)

Chevalley writes in his review[26]

> At this stage, the topology may be thrown out, leaving only the algebraic facts in evidence.

The principal result is that the Weil algebra is acyclic, and so

> the Weil algebra may be considered as the algebra of cochains of a (nonexistent) fiber space which would be classifying in all dimensions.

Cartan analyzed the case of a homogeneous space and identified the transgression in this context. The invariant elements I_S of the symmetric algebra on $\mathfrak{g}^*$ play a prominent role. One of the main results, referred to as being part of Hirsch–Koszul theory, is a cochain equivalence between $H^*(E) \otimes I_S$ endowed with a differential and the cochains on B. This identifies I_S with the cohomology of a classifying space when $H^*(E)$ is trivial in a range.

Though spectral sequences do not figure in Cartan's Bruxelles papers, the success of his methods put the underlying algebra of differential forms at the forefront of topology and it was this work that was planned as the basis of Bourbaki's version of algebraic topology.[27] However, all this would soon be overtaken.

The talks of Hopf and Eckmann emphasized the homotopy-theoretic approach to fibre spaces. In his talk Eckmann[28] reported some progress on a question of Deane Montgomery (1909–1992) and Samelson [142] as to whether Euclidean n-space $\mathbb{R}^n$ can be the total space of a fibre space with compact fibre. In the first issue of the *Proceedings of the American Mathematical Society*, April 1950, Gail S. Young (1915–) [214] published partial results on this problem. Borel and Serre decided to apply Leray's ideas to this problem and quickly came up with a complete solution – there is no fibration of $\mathbb{R}^n$ with a compact fibre that does not reduce to a point. The proof is *"une application simple"* [18] of the ideas of Leray: Since the E_2-term of the Leray spectral sequence (*suite de Leray–Koszul* in this paper) is $H^*(B) \otimes H^*(F)$ (for real coefficients in Leray's cohomology theory), $H^n(\mathbb{R}^n) = \mathbb{R}$, and B has dimension less than or equal to $n-1$, $H^p(B) \neq 0$ for some minimal $p \leqslant n-1$. When F is connected, $H^*(F)$ has a unit 1. Let $y \in H^p(B)$ be nonzero, then $1 \otimes y \in E^2$

[25] Weil preferred the nomenclature "universal algebra associated to a Lie algebra". See [190, I, p. 568].

[26] MR# 13, 107e,f.

[27] See [190], endnotes to Vol. II.

[28] Letter of A. Borel, May 5, 1997.

is of minimal total degree and cannot be a coboundary. Since it is already a cycle for all differentials in the spectral sequence, $H^p(\mathbb{R}^n) \neq 0$, a contradiction. When F is not connected, each connected component of the fibering determines a fibering of $\mathbb{R}^n$ and so we are reduced to the previous argument, unless F is a finite set. Then the fibering is a covering space and an argument using group cohomology yields the result.

In the fall of 1950, Serre introduced a new spectral sequence (*un anneau spectral* in this paper) for the cohomology of groups. Based on Koszul's example for Lie algebras and Cartan's notion of a filtered differential graded ring, Serre announced [163] a filtered differential graded ring of cochains on which the quotient G/g by a normal subgroup g of G acts. The resulting spectral sequence has E_2-term $H^*(G/g, H^*(g, A))$ and terminal ring $H^*(G, A)$ for coefficients in an Abelian group A. This gave a new orientation of the results of Roger C. Lyndon (1917–1988) [131] whose use of subquotients to compute cohomology of groups may be viewed as a precursor to the use of spectral sequences (without differentials). Serre applied the spectral sequence to acyclic spaces on which a group acts, to Galois cohomology, and particular cases known to G.P. Hochschild (1915–) [83] and Eilenberg and Mac Lane [62]. Shortly after the publication of [163], he received a letter from Hochschild who described an explicit filtration of the cochains. They agreed to publish together and the complete exposition of these results appeared in 1953 *Transactions of the AMS* [84].

In the late summer of 1950 (August 30–September 6) another major conference, the International Congress of Mathematicians in Cambridge, Massachusetts, took place. Leray spoke [133] twice, on fixed point theory,[29] and on *L'emploi, en topologie algébrique, du formalisme du calcul différentiel extérieur*, which dealt with his cohomology theory. Some of the major papers on homotopy theory treated fibre spaces and homotopy groups, with a major address by George Whitehead (1918–) who gave a survey of known results on the homotopy groups of spheres, including his result that $\pi_{n+2}(S^n) \cong \mathbb{Z}/2\mathbb{Z}$ for $n > 1$, correcting a computation of Pontryagin [150]. Global questions about homotopy groups are noticeably absent in his talk. J.H.C. Whitehead (1904–1960) described his program to devise an algebraic homotopy theory solving classification problems with the appropriate algebraic input [200]. Whitehead had recently proved the *Whitehead Theorem*, that a map of simply-connected CW-complexes inducing an isomorphism of homology groups induces an isomorphism of homotopy groups [197]. Other reports of interest to this history were given by Hirsch (on the homology of fibre spaces), by Hurewicz (on relations between homotopy and homology), by Chern (on the transgression in differential geometry), by Spanier (on the Gysin sequence for CW-complexes), by Massey (on the homotopy groups of triads), and by Morse (on variational problems and topology).

Post-war topology was thriving. The directions of research in place before the war were being played out successfully with the addition of new questions, new methods and new researchers. The task of consolidation was undertaken by the leaders of the field (Cartan in France, Eilenberg and Steenrod in the US). This remarkable time presented a rich field of opportunity for new and able researchers. The next two years realized that opportunity and changed the field of algebraic topology dramatically.

[29] The dedication of this note reads, "À la mémoire du profond mathématicien polonais Jules Schauder, victime des massacres de 1940".

5. Serre's thesis

After the successful application of Leray's spectral sequence to the question of Montgomery and Samelson [18], Serre turned to other possible applications of these ideas. An example of a fibre space that was well known in 1950 is given by the limit of the finite-dimensional complex projective spaces, $S^1 \hookrightarrow S^\infty \to \mathbb{CP}(\infty)$. Since S^1 is an Eilenberg–Mac Lane space, $K(\mathbb{Z}, 1)$, and S^∞ has trivial homotopy groups, this fibre space identifies $\mathbb{CP}(\infty)$ as a $K(\mathbb{Z}, 2)$. The *algèbre spectrale* of Leray for this case was clear and suggested the possibility that an induction might lead to the computation of the cohomology of the higher Eilenberg–Mac Lane spaces from this initial case. Serre writes [160, Vol. 1, p. 585]:

> J'avais remarqué que la théorie de Leray permet d'aborder ce calcul en procédant par récurrence sur n, pourvu que l'on dispose d'un espace fibré E ayant les propriétés suivantes: (a) E est contractile, (b) la base de E est un $K(\pi, n)$, ce qui entraîne (c) les fibres de E sont des $K(\pi, n-1)$.

The identification $\Omega K(\pi, n) \simeq K(\pi, n-1)$ soon led Serre to consider the sequence of spaces $\Omega X \hookrightarrow PX \to X$ where $PX = \{\text{continuous maps } \lambda : (I, 0) \to (X, x_0)\}$ and the mapping $PX \to X$ is evaluation at 1, $\lambda \mapsto \lambda(1)$.

If you assume that such a space is a fibre space, and that Leray's theory applies (even though his hypotheses do not), this gives a route to the computations of $H^*(K(\mathbb{Z}, n); \mathbb{Q})$ by induction. The first hurdle to clear was the right notion of fibre space for the path space evaluation map. This came from the homotopy lifting property, emphasized by Eckmann [48] in his Bruxelles lecture. Serre, in [164, I], defined a fibre space to be a mapping that verifies the homotopy lifting property with respect to mappings of polyhedra to the total space. This notion generalized the fibre spaces of Hurewicz and Steenrod [102] and hence included the geometric examples of fibre bundles. Furthermore, such a fibre space, now called a *Serre fibration*, gives rise to a homotopy equivalence of fibres at various points, and the long exact sequence of homotopy groups

$$\cdots \to \pi_i(F) \to \pi_i(E) \to \pi_i(B) \to \pi_{i-1}(F) \to \cdots.$$

Finally, and perhaps most importantly, the path space evaluation map satisfied this property.

Though this admitted more general spaces as fibre spaces, it did not settle the question of applicability of Leray's methods. The singular theory of Lefschetz and Eilenberg [54] offered the best properties for the study of homotopy-theoretic constructions, especially via the Hurewicz homomorphism. However, there lacked an analogue of Leray's theory for singular homology. At a Bourbaki meeting in the fall of 1950, Serre discussed this problem with Cartan and Koszul. He writes [160, Vol. 1, p. 585]:

> ... fort heureusement, J.-L. Koszul et H. Cartan m'ont suggéré une certaine filtration du complexe singulier ... qui s'est révélée avoir toutes les vertus nécessaires.

The suggestion led to the technical part of Serre's thesis [165, Chapter II] which is based on cubical singular theory for which ([165]) the cubes

> se prêtent mieux que les simplexes à l'étude des produits directs, et, a fortiori, des espaces fibrés qui en sont la généralisation.

Having overcome the technical difficulties, the consequences were announced in a series of three *Comptes Rendus* notes. In the first note [164, I] the term *spectral sequence* (*suite* *spectrale*) makes its appearance. It applies to the homology spectral sequence for which the terms *spectral ring* or *spectral algebra* of Leray and Koszul did not apply. The identification of the initial term of this spectral sequence with $H_*(B; \mathcal{H}_*(F))$, here homology with local coefficients, is made and the role of the simple system acknowledged. This note also shows how the transgression and suspension may be identified with differentials in the spectral sequence. Serre states that Leray's results may be transferred to the singular setting and that the Wang sequence may be proved from the spectral sequence.

In the second note [164, II] the path evaluation map is identified as a fibre space. If X is a space for which there is field k and an $n > 0$ with $H_n(X; k) \neq \{0\}$ and $H_i(X; k) = \{0\}$ for $i > n$, then the nontriviality of $H_i(\Omega X; k)$ for infinitely many i follows. This settled a question of Morse [145], to establish topological conditions for a Riemannian manifold to satisfy in order to show, using Morse theory, that any pair of distinct points in the manifold are joined by infinitely many geodesic segments. From the finite generation of $H_i(X)$ for all i, the finite generation of $H_i(\Omega X)$ for all i follows. Serre also recognizes ΩX as an H-space to which Leray's extension [122] of Hopf's theorem applies giving the rational cohomology of ΩX as a tensor product of polynomial and exterior algebras.

In the third note [164, III] the consequences for the computation of homotopy groups are presented. The main tool is an iterative construction: If X is path-connected, let $X_0 = X$ and $T_1 = \widetilde{X}_0$, the universal cover of X_0. Let $X_1 = \Omega T_1$, and finally, let $T_2 = \widetilde{X}_1$, etc. If such a sequence is possible to construct, then it has the property that $\pi_1(X_n) \cong \pi_{n+1}(X)$. Furthermore, $H_1(X_n) \cong \pi_1(X_n)$. Using the Cartan–Leray spectral sequence, one could make the passage of homology data from X_{n-1} to T_n, and using the new homology spectral sequence, one could pass data from T_n to X_n. From this construction, Serre proved that a space, for which the fundamental group acts trivially on the higher homotopy groups, and for which $H_i(X)$ is finitely generated for all i, has $\pi_i(X)$ finitely generated for all i. This result opened up new territory for homotopy theory – a global result about homotopy groups improving the sparse evidence of 1950 and admitting algebraic refinements that were unknown at the time. Using homology with field coefficients, Serre showed that $H_i(X; k) = \{0\}$ for $0 < i < n$ implied $\pi_i(X) \otimes k = \{0\}$ for $i < n$ and that $\pi_n(X) \otimes k \cong H_n(X; k)$. Since a finitely generated group G is finite if and only if $G \otimes \mathbb{Q} = \{0\}$, this leads to a global result for the homotopy groups of spheres: $\pi_{n+k}(S^n)$ is finite for $k > 0$ except in the case $\pi_{4l-1}(S^{2l})$ which is isomorphic to $\mathbb{Z}$ direct sum a finite group. For p an odd prime, Serre showed that $\pi_i(S^n) \otimes \mathbb{F}_p = \{0\}$ for $n < i < n + 2p - 3$ and $\pi_{n+2p-3}(S^n) \otimes \mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$ when p is an odd prime. Finally, results on the homology of Eilenberg–Mac Lane spaces were announced following the heuristic inductive argument that started the research.

The thesis was finished in the spring of 1951 and sent to Steenrod for the *Annals of Mathematics* at the advice of Eilenberg. Steenrod gave it priority and it appeared at the end of 1951. The thesis is a remarkable mix of technical detail and simple direct argument. The plan follows the *Comptes Rendus* notes. Among the details one might view as technical are proofs of the simplicity of H-spaces, the development of cubical singular theory, and the identification of a condition (ULC, *uniformement localement contractile*) which implies the existence of universal covering spaces and which is passed on to the based loop space.

The ability to compute the homology of Eilenberg–Mac Lane spaces led to new computations in homotopy theory. In particular, Cartan and Serre introduced a new homotopy-theoretic construction giving fibre spaces susceptible to analysis with the Serre spectral sequence. In [35], they introduced the method of *killing homotopy groups*: To each path connected space X there is a sequence of spaces (X, n) with $(X, 1) = X$, together with a sequence of mappings $f_n : (X, n+1) \to (X, n)$, so that (X, n) is n-connected, the mapping f_n induces an isomorphism $\pi_i((X, n+1)) \to \pi_i((X, n))$ for $i > n$, and further, f_n is a fibration with fibre a $K(\pi_n(X), n-1)$, base space (X, n) and total space homotopy equivalent to $(X, n+1)$. The construction answered in the affirmative Problem 32 of Eilenberg's Princeton list [57].

> Dans la mesure où l'on connait les groupes d'Eilenberg–Mac Lane d'un groupe π donné, on obtient une méthode de calcul (partiel) ... des groupes d'homotopie de X.

In degrees out to twice the connectivity, the spectral sequence gives a long exact sequence mixing homology groups of the nth homotopy group of X with the homology of the spaces (X, n). Analysis of this sequence with computations of certain homology groups of Eilenberg–Mac Lane spaces allowed them to compute that the p-component of $\pi_{2p}(S^3)$ is $\mathbb{Z}/p\mathbb{Z}$, that $\pi_7(S^3)$ and $\pi_8(S^3)$ are 2-primary and that $\pi_9(S^3)$ is the direct sum of $\mathbb{Z}/3\mathbb{Z}$ with a 2-primary group.

George Whitehead [193] independently considered such a tower of spaces as the (X, n). The common source of background was the so-called Eilenberg complex, $\mathcal{S}(X, x, q)$, obtained by considering that part of the singular complex with simplices whose $(q-1)$-faces are at a point x. Eilenberg [54] had shown that $H_q(X, x, q) = H_q(\mathcal{S}(X, x, q)) \cong \pi_q(X, x)$. The killing homotopy groups construction of Cartan and Serre obtains a space (X, n) with $H_j((X, n)) \cong H_j(\mathcal{S}(X, x, n))$ for all j. Whitehead discovered the same long exact sequences as Cartan and Serre from the relative homology of consecutive Eilenberg complexes, which is isomorphic to the homology of $K(\pi_n(X), n)$. At around the same time M.M. Postnikov (1927–) introduced a dual construction, the *Postnikov tower* [151], using the simplicial methods of Eilenberg in an effort to understand the degree to which the homology of a space is determined by its homotopy groups.

In the spring of 1952, Serre announced his complete computation of the mod 2 cohomology of the Eilenberg–Mac Lane spaces [166]. The expected inductive argument was successful with the introduction of the idea of a simple system of generators, due to Borel [10]. The transgression, which commutes with Steenrod operations, allows the identification of polynomial generators between consecutive $K(\mathbb{Z}/2\mathbb{Z}, n)$'s. By studying the limit with respect to suspension, Serre computed $A^*(\mathbb{Z}/2\mathbb{Z})$ whose q-th degree elements may be identified with $H^{n+q}(K(\mathbb{Z}/2\mathbb{Z}, n); \mathbb{F}_2)$ for n large. Since stable mod 2 cohomology operations are identified with $A^*(\mathbb{Z}/2\mathbb{Z})$, Serre had computed the complete set of primary mod 2 cohomology operations. When the complete proofs of this note appeared [169], it included another global result about the homotopy groups of finite complexes. By a subtle argument involving the convergence of Poincaré series for the Eilenberg–Mac Lane spaces and the killing homotopy tower of spaces, Serre proved that a path-connected, simply-connected space whose integral homology is of finite type in all degrees and whose mod 2 homology is nontrivial in some positive degree and vanishes after a certain degree, has infinitely many homotopy groups $\pi_i(X)$ containing $\mathbb{Z}$ or $\mathbb{Z}/2\mathbb{Z}$. Such a result would have been unthinkable only three years earlier when even the finite generation of homotopy groups for finite CW-complexes was open. It showed decisively how the behavior of the homotopy groups is not at all like the homology groups. A new era of computation and structure opened up.

The final major paper of this period takes off from the proof of the finitude of the homotopy groups of spheres. By using homology with coefficients in a field k, Serre proved a Hurewicz theorem for the Hurewicz map tensored with k, $\pi_n(X) \otimes k \to H_n(X; k)$. Since tensoring with the field of p elements destroys q torsion for q prime and $q \neq p$, a kind of fracturing of homotopy problems is possible, one version for each prime p. A comprehensive approach to this fracturing is the subject of [168]. Serre introduced the general notion of a *class of Abelian groups*: $\mathcal{C}$ is a class of Abelian groups if

(a) every trivial group belongs to $\mathcal{C}$;

(b) every group isomorphic to a subgroup or a quotient group of a group in $\mathcal{C}$ is in $\mathcal{C}$, and

(c) every extension of two groups in $\mathcal{C}$ is in $\mathcal{C}$.

Examples include the class of finitely generated Abelian groups, all finite Abelian groups, all finite groups whose order is divisible only by primes belonging to a given set. By neglecting groups appearing in this class, one could prove "mod $\mathcal{C}$" versions of the classical results, including the Hurewicz theorem and the Whitehead theorem. More subtle applications include expressions of homotopy results mod $\mathcal{C}$. For example, if $\mathcal{C}$ is the class of finite groups of order a power of 2, then $\pi_i(S^{2m})$ is isomorphic mod $\mathcal{C}$ to $\pi_{i-1}(S^{2m-1}) \oplus \pi_i(S^{4m-1})$. For compact Lie groups, Hopf's theorem can be considered as a homotopy equivalence mod $\mathcal{C}$ between G and a product of odd spheres, when $\mathcal{C}$ is the class of all torsion groups. A prime p is called *regular* for the Lie group G if the same result holds mod the class of finite groups of order prime to p. Serre determined conditions for a prime to be regular for the classical groups.

Throughout this period of productive homotopy theory, Serre made many other contributions, particularly to algebraic geometry [160, Vol. 1]. We discuss the impact of Serre's work in later sections. We next turn to the work of Armand Borel.

6. Borel's thesis

Borel spent the academic year 1949/50 in Paris at the CNRS and attended Leray's course [131] at the Collège de France, in fact, he helped to complete the exposition in a note appearing at the end of Leray's paper [8]. Borel's thesis begins with the substance of the course of Leray in 1949/50. Since Lie groups and homogeneous spaces are locally compact, and Leray's methods are general enough to include both de Rham cohomology and singular cohomology with coefficients in any ring, they became his tool of choice. In particular, the basic methods were topological and the analytic structure of Lie groups played a minor role.

All of the strands of research up to 1950 concerning the topology of Lie groups find a place in Borel's thesis. In particular, Hopf's theorem, Samelson's thesis, the transgression of Hirsch, Koszul, and Cartan, and the characteristic classes of Stiefel, Whitney, Chern, and Steenrod all find a topological context and algebraic methods for their generalization and development.

Borel published a series of *Comptes Rendus* notes in 1951 outlining his results and methods. In the first [9] he investigated an analogous situation to the question of Montgomery and Samelson – which spheres are the total space of a fibre space with products of spheres as fibre? In the associated spectral sequence the differentials must be nontrivial: Suppose (F, E, B) is a fibre space with E a sphere, F a product of spheres, and B a finite complex. Then the spheres making up the fibre must be of odd dimensions $\{m_j\}$ and the algebra generators of $H^*(F; k)$ must transgress to the base making the cohomology of the base $H^i(B; k) \cong (k[u_1, u_2, \ldots, u_s])^i$ with the $\deg u_j = m_j + 1$. The main theorem of the note is a new proof of a theorem of Eckmann, Samelson and G. Whitehead [49] that, in fact, only a single sphere can appear in the fibre.

In [10] Borel introduced the notion of a *simple system of generators* over the field $\mathbb{F}_2$: An algebra A over $\mathbb{F}_2$ has a simple system of generators $h_1, \dots, h_m$ if the products, $h_{i_1} h_{i_2} \cdots h_{i_s}$ with $i_1 < i_2 < \cdots < i_s$, form a basis for A over $\mathbb{F}_2$. If $\widetilde{H}^*(X; \mathbb{F}_2)$ has a simple system of generators and X is connected, then the mod 2 Poincaré polynomial of X is given by

$$(1 + t^{r_1})(1 + t^{r_2}) \cdots (1 + t^{r_s}),$$

where $\deg h_i = r_i$. With this definition, the mod 2 cohomology of the various Stiefel manifolds (real, $V_{n,p}$; complex, $W_{n,p}$; quaternionic, $U_{n,p}$) may be expressed easily. The torsion in the integral cohomology of $\mathrm{Spin}(n)$ and the exceptional groups G_2 and F_4 is also described in this note.

In [11] Borel described the cohomology of a classifying space by considering the transgression for $E(n, G) \to B(n, G)$, the universal bundle for each dimension n which, as defined by Steenrod in [179], is a principal G-bundle for which $E(n, G)$ is at least n-connected. The algebra that emerges in his analysis of fiberings of a sphere by products of spheres goes over more generally as a compact Lie group G has cohomology with coefficients in a field k like that of a product of spheres when G has no torsion divisible by the characteristic of k. Furthermore, when $k = \mathbb{F}_2$, a simple system of universally transgressive generators go by the transgression to a basis for a polynomial algebra. Borel remarks at the end of the note that the action of the mod 2 Steenrod squares together with the formula of Wu [210] determine the product structure for $H^*(\mathrm{SO}(n); \mathbb{F}_2)$ completely because the Sq^i commute with the transgression.

In the final note mentioned in the thesis [12], the full power of the previous methods come to bear on the problem of computing $H^*(G/U; k)$. A new construction is introduced, now called the *Borel construction*: If (U, X, Y) is a locally compact principal U-bundle, then one forms the space $E_U \times_U X$ which is the quotient of the product of $E(n, U)$ (for n large enough) and X by the relation $(s, t) \sim (u \cdot s, u \cdot t)$, where $u \in U$. This space comes equipped with two fibrations; one with base Y gotten by projection from the second factor, and the other with base B_U, the classifying space $B(n, U)$ and fibre X. It follows that there is a spectral sequence (*un anneau spectral* in [12]) with E_2-term $H^*(B_U; H^*(X; k))$ and converging to $H^*(Y; k)$. When U is discrete, this is the Cartan–Leray spectral sequence and when $X = G$, and U is a closed subgroup of G, then the spectral sequence goes from $H^*(B_G; H^*(G/U; k))$ to $H^*(B_U; k)$. For the case of $U = T$, a maximal torus of G, under the assumption that G and G/T are without p-torsion, $H^*(B_G; \mathbb{F}_p)$ can be identified via the mapping induced by $BT \to BG$ with the invariant ring in $H^*(BT; \mathbb{F}_p) \cong \mathbb{F}_p[x_1, \dots, x_l]$ under the action of the Weyl group, where l is the rank of G. When S is the maximal torus of U, then there is a fibration $(U/S, G/S, G/U)$ and these methods lead to conditions under which U/S is totally nonhomologous to zero in G/S with respect to $\mathbb{F}_p$ coefficients. These results generalize considerably the formula of Hirsch which is for rational coefficients. The note ends with a remark on Hopf's theorem – when G is a connected, compact, Lie group, then $H^*(G; \mathbb{F}_p)$ has a p-simple system of generators, that is, a set of classes $x_1, x_2, \dots, x_s$ such that products of the form $x_1^{r_1} x_2^{r_2} \cdots x_s^{r_s}$, where $0 \leqslant r_i < p^{t_i}$ and $x_i^{p^{t_i}} = 0$, but $x_i^{p^{t_i} - 1} \neq 0$.

On March 25, 1952, Borel submitted his thesis before a committee of Leray (Président), Cartan and Lichnerowicz to obtain his doctorate from the Université de Paris. The paper appeared in the 1953 *Annals of Mathematics*. It explicates the *Comptes Rendus* notes in all details. The heart of the thesis is algebraic and turns on the extension to mod p coefficients of Hopf's theorem on the structure of a finite dimensional Hopf algebra and on the algebraic necessity of a nontrivial transgression in a spectral sequence of algebras with trivial E_∞-term. The topological inputs are principally the Borel construction and the use of the maximal torus.

The topological properties of compact Lie groups at a prime p were inaccessible by the methods of Cartan, Chevalley, Koszul, and Weil. New results followed quickly from Borel's point of view. He and Serre [20] analyzed the discrete Abelian subgroups of compact Lie groups to show, among other things, a technique for determining the torsion in the exceptional groups G_2 and F_4.

Another test of the topological methods was the new set of invariants given by the Steenrod operations at each prime. Wu and Thom had demonstrated the importance of these operations mod 2 for characteristic classes by 1950. The computation of the mod p cohomology of compact Lie groups was amply demonstrated in Borel's thesis. After a lecture series by Steenrod in May 1951, Serre and Borel tackled the question of the mod p operations and successfully determined the action of the reduced powers on $U(n)$, the unitary groups, $Sp(n)$, the symplectic groups, and $SO(n)$, the special orthogonal groups [19]. This computation brought the Chern classes into the framework of Wu and Thom, and also settled many cases of the nonexistence of sections of certain fibre spaces given by homogeneous spaces. In particular, Borel and Serre settled a problem of Hopf [98] as to which spheres possessed an almost-complex structure. They show that only S^2 and S^6 can have such a structure, a surprising result at the time.

In a later paper [15], Borel's methods were so refined as to obtain significant integral cohomology results for the entire class of simple compact Lie groups. Applications of Borel's work changed the study of the topology of Lie groups. Samelson's survey of 1952 [155] presents a mix of techniques with which to attack questions of the topology of Lie groups that includes analytic, algebraic, and topological means. Borel's survey of 1955 [16] showed how topological methods, together with the algebra of spectral sequences, were sufficient to uncover and unite the invariants of Lie groups and homogeneous spaces.

The role of Borel's computations in the study of cobordism, of algebraic geometry (Hirzebruch [81]), and of manifold theory is a subject for another history. The removal of a topological impediment, the lack of a method of computation, opened enormous opportunities for the application of algebraic topology to other areas of mathematics.

7. Reception

The method of spectral sequences did not spread rapidly after its initial appearance. Leray's 1946 *Comptes Rendus* notes led Koszul and Cartan in Strasbourg to extract the algebraic essence from which further constructions could be made, notably by Koszul in his thesis, by Cartan in his work on the transgression, by Cartan and Leray for finite groups acting on a space, and by Serre in his note on the cohomology of groups. However, little interest outside of France was evident before Serre's thesis.

Several missed opportunities present themselves – Mac Lane[30] reports visiting Paris in 1947 and discussing sheaves and spectral sequences with Leray. Lyndon's thesis [136] under Mac Lane is founded on a filtration of the cohomology of a group $H^*(G; M)$ relating the associated graded to subquotients of $H^*(Q; H^*(K, M))$ when

$$1 \to K \to G \to Q \to 1$$

is a group extension and M a G-module. Mac Lane admits not making the connection between Lyndon's work and Leray's work – "Leray was obscure!"

Another near miss is the work of Tatsuji Kudo in Japan. In [111], published in the 1950 Osaka Journal, Kudo analyzed a fibre space with CW-complex as base by considering the preimages of the skeleta and the resulting long exact sequences of pairs. In a subsequent paper, published in 1952 [112], Kudo began to work with Leray's ideas and recast his previous analysis in this language. With the appearance of Serre's thesis, the potential of spectral sequences was explored by others in Japan interested in homotopy theory, especially Hiroshi Toda (1928–) [188] whose computations of the homotopy groups of spheres went far beyond the state of the art in 1953.

Whitehead, Massey and others in the United States did try to understand Leray's work after its appearance, in order to get at the source of the "marvelous results he claimed ...".[31] The exact sequence was a fundamental tool of expression by this time. Massey soon gave a new algebraic reformulation of spectral sequences, his *exact couples* [139]. He claims two papers as the sources for his idea. In [199] J.H.C. Whitehead introduced the so-called *Whitehead groups*. One first needs a formalism to handle sequences of Abelian groups: For each n, suppose there are homomorphisms $j: A_n \to C_n$, and $\beta: C_n \to A_{n-1}$ leading to a sequence

$$\cdots \to C_{n+1} \to A_n \to C_n \to A_{n-1} \to \cdots.$$

Suppose further that $j(A_n) = \ker \beta$, but that $\beta(C_n)$ need not be the kernel of j. Whitehead identified a differential $d = j\beta$, satisfying $d \circ d = 0$, and the groups $\Gamma_n = \ker j : A_n \to C_n$, $\Pi_n = A_n/\beta C_{n+1}$ and $H_n = H_n(C_*, d)$, together with the sequence of homomorphisms $[j]: \Pi_n \to H_n$, $[\beta]: H_n \to \Gamma_{n-1}$ and $[k]: \Gamma_n \to \Pi_n$, where $[j]: \Pi_n \to H_n$ is defined by j on a representative $a \in A_n$ of $a + \beta C_{n+1}$ in Π_n followed by the mapping induced by the quotient from cycles of d to H_n; the homomorphism $[\beta]: H_{n+1} \to \Gamma_n$ is induced by taking a cycle $z \in C_n$ of d which satisfies $j\beta(z) = 0$ and mapping it to $\beta(z) \in \Gamma_n$; finally, $[k]: \Gamma_n \to \Pi_n$ is induced by the composite $\Gamma_n \subset A_n \to \Pi_n = A_n/\beta C_{n+1}$. The main result is the exactness of the sequence

$$\cdots \to H_{n+1} \to \Gamma_n \to \Pi_n \to H_n \to \cdots.$$

An exact couple (D, E, i, j, k) is given by a similar algebraic setup – the sequence of homomorphisms, $i: D \to D$, $j: D \to E$, and $k: E \to D$ form a long exact sequence:

$$\cdots \to D \to D \to E \to D \to \cdots.$$

[30] Letter of August 11, 1997.

[31] Whitehead, letter of November 6, 1996.

It follows that the composite $d = j \circ k$ is a differential. Massey shows in [139] that an exact couple can be *derived* to give another exact couple $(iD, H(E,d), i', j', k')$, where the derived homomorphisms arise in a manner analogous to Whitehead's scheme. The algebraic analysis carried out by Chern and Spanier in [41] to prove the Gysin theorem for CW-complexes gave a model for the bigraded case of an exact couple.

Massey developed exact couples around the time Serre's thesis was being written and the appearance of this new algebraic apparatus added to possible applications of these ideas. The simplicity of the algebraic notion also led some to experiment with other constructions, especially with homotopy groups. The paper [139, I, II] ends with a construction of an exact couple with $D = \pi_n(K^p, K^q)$, where K^r denotes the r-skeleton of a CW-complex K. Massey identified parts of the E^2-term of the spectral sequence; for example, $E^2_{2,1} \cong \Gamma(\pi_2(K))$, one of the Whitehead groups which is isomorphic to $H_4(K(\pi_2(K), 2))$. In subsequent papers [139, II–V], Massey introduced a spectral sequence in cohomotopy, similarly based on the skeleta. The method of exact couples offered a tidy algebraic object with which one could construct and investigate many topological invariants. This work was the subject of Massey's talk at the 1953 Cornell conference on fibre spaces.

With the arrival of Serre's thesis at the *Annals of Mathematics* Steenrod[32] spread the word in the United States of some "earth-shaking results on the homotopy groups of spheres". He also sent it out to George Whitehead (then at Brown) and to Henry Whitehead (at Oxford), both among the few world's experts on the homotopy groups of spheres.

In contrast to the perception of Leray's papers, Serre's work was "brilliantly clear" in exposition[32] and its effect was immediate. In France, the methods to compute homotopy groups played an important role in the work of René Thom, then in Strasbourg. Thom's earliest *Comptes Rendus* notes [185, 187] explore the relations among the homologies of a sphere space and a general notion of characteristic classes associated to such bundles. Thom related the Stiefel–Whitney classes to the action of the then new mod 2 Steenrod algebra ([176]). The development of the structure of manifolds and their embeddings led Thom to his results on cobordism theory. The methods of Serre's thesis gave the key to computations of the homotopy groups $\pi_{n+k}(MO(n))$ of the so-called *Thom spaces* which Thom had shown to be isomorphic to the cobordism group of k-dimensional manifolds [187]. The rich field of beautiful mathematics that cobordism opened is the subject for another history.

From Henry Whitehead's group, Peter Hilton (1923–) took immediately to spreading Serre's and Borel's work, in particular, to his seminar in Cambridge where he had arrived from Manchester in 1952. Among some of the early participants were J.F. Adams, M.F. Atiyah, E.C. Zeeman, D.B.A. Epstein, and C.T.C. Wall. At Oxford, Whitehead held lectures on Serre's results and invited him to visit. I.M. James recalls this visit and the impact of Serre's thesis on his work in [103].

In the United States Eilenberg had followed the development of spectral sequences from the beginning. In the *Séminaire Cartan* of 1950/1951, Eilenberg presented his version of spectral sequences in two lectures January 22 and February 5, 1951. In these reports he described another construction of spectral sequences that is featured in the classic book with Cartan *Homological Algebra* [36]. A *Cartan–Eilenberg system* is built on a partially ordered set $(S, <)$ and a functor H that associates to an pair $A < B$ a group $H(A, B)$

[32] Letter from George Whitehead of September 6, 1997.

subject to the axioms that generalize the case of pairs of subspaces of a space, that is, $A < B$ when $B \subset A \subset X$. In the second of the lectures Eilenberg applies this algebraic technology to the case of a fibre space with the filtration resulting from the skeletal filtration of the base space. The Künneth theorem allows one to analyze the subsequent Cartan–Eilenberg system and to identify the E^2-term of the associated spectral sequence with the homology of the base with local coefficients in fibres, as in the case of Serre's thesis.

At Brown, George Whitehead's first Ph.D. student John C. Moore (1926–) immediately took up Serre's methods in his thesis. The main result of the paper [143] (received September 19, 1952) is an extension of Serre's results on $\pi_q(S^n) \otimes \mathbb{F}_p$, in particular, that $\pi_q(S^n) \otimes \mathbb{F}_p = \{0\}$ if $n + 2p - 3 < q < n + 4p - 6$ and $n + 4p - 5 < q < n + 6p - 9$. The technical advances Moore introduced included a version of the Serre spectral sequence for pairs and an isomorphism theorem ([143, Theorem 2.2]) which shows how an isomorphism of spectral sequences can be used to prove an isomorphism of their targets. He goes on to apply these methods to obtain information about triad homotopy groups that had been introduced by Blakers and Massey in [6]. In [144] (received March 31,1953, revised January 14, 1954) Moore extended the computation of homotopy groups to a class of spaces, suggested by Steenrod and already considered by Serre [167], with prescribed homology – a space X is of *homology type* (G, n) if X is simply connected and has trivial homology except in dimension n, where $H_n(X) \cong G$. Moore makes use of the method of Cartan–Serre–Whitehead and the sequence of universal covers of loop spaces found in Serre's thesis in his analysis, as well as the Hopf algebra results of Borel's thesis that had appeared by this time. Moore had frequent contact with Borel and Serre in the fifties. Borel visited the Institute for Advanced Study during the years 1952–1954 and became a permanent member in 1957. Serre visited the Institute for Advanced Study in the years 1955, 1957, and 1959. Moore visited often in Paris – the *Séminaire Cartan* of 1954/1955 is often nicknamed the *Séminaire Cartan–Moore*.

In the former Soviet Union, leadership in topology was changing around the time of Serre's and Borel's work. In Moscow, a seminar was held on Serre's thesis only in 1956 led by Albert S. Schwarz (1934–), M.M. Postnikov (1927–) and V.G. Boltyanskiĭ (1925–). The participants of the seminar included S.P. Novikov, D.B. Fuks, A.G. Vinogradov, and others who went on to make up the next generation of Soviet topologists.

By the late 1950's Leray's sheaf theory and homotopy theory had reached a stage that expository accounts of the subject were possible. The first book length account is the set of lectures by Borel of 1951 at the EPF Zürich [13]. Subsequent influential accounts that treat spectral sequences include the classic *Homological Algebra* of Cartan and Eilenberg [36], *Topologie algébrique et theorie des faisceaux* by R. Godement [74], which treats the final version of sheaf theory developed by Cartan; *Homotopy Theory* by S.-T. Hu [99], and *Homology Theory* by Peter Hilton and Shaun Wylie [77] which treats homology theory especially as it has an impact on questions in homotopy theory. Among the most influential papers extending these ideas is the celebrated 1958 *Tôhuku Journal* paper of Alexandre Grothendieck (1928–) which provided a categorical framework in which spectral sequences arise naturally (and much more). By this time spectral sequences were standard in the topologists' toolbox and had moved into the working language of other branches of mathematics.

8. Closing remarks

In Steenrod's report [181] of the Spring 1953 conference on *Fiber bundles and differential geometry* at Cornell University he writes of the "upheaval within topology itself resulting from the use of fiber space techniques .... The most striking feature of the conference was the frequent use of the same apparatus in two or more widely separated disciplines, with strong suggestions of a probable unification of geometry on some higher level". That apparatus was the spectral sequence. The landscape of homotopy theory had changed radically and central to this change was the appearance and application of spectral sequences, in particular, of Serre's and Borel's Paris theses.

The two problem sets resulting from the Cornell conference signal the rapid acceptance of spectral sequences. The first set by Hirzebruch [82] dealt with differentiable and complex manifolds and is influenced by the work of Borel and the new progress by Thom. The other by Massey singles out spectral sequences among the preferred tools whose development would shape progress in algebraic topology. Furthermore, Massey marks the passing of the immediate post-war state of the subject with an appendix describing the progress on the Eilenberg list of problems.

How was topology different after the introduction of spectral sequences? It is certain that the algebraic content in algebraic topology increased. However, this may be understood more subtly as a kind of algebraization of the subject. Several currents support this development. The program of Élie Cartan to extract the topology of Lie groups and homogeneous spaces from the algebraic properties of Lie algebras opened up a fundamental role for algebra in topology when it was carried on by the work of Chevalley, Weil, Koszul, and Henri Cartan. One of the most profound computations in homotopy theory, Cartan's determination of the homology of the Eilenberg–Mac Lane spaces [25] using his constructions, is clearly based on his use of Weil algebras in [30]. These computations have not been superseded in the literature of algebraic topology and have been the source for many new developments.[33]

Borel's thesis carried on the program initiated by Hopf in his introduction of H-spaces. The topological properties of compact Lie groups and homogeneous spaces were accessible by topological means in a manner that was unified by the application of general theorems like Borel's extension of Hopf's structure theorem for Hopf algebras and general machinery like the transgression and spectral sequences.

Serre's thesis did more than establish the utility of spectral sequences for singular theory. The general nature of the input, the Serre fibration, admitted methods of construction more like the underlying algebra than previously possible, as shown, for example, in the method of killing homotopy groups. In contrast, consider the definition fibre spaces of Hurewicz and Steenrod [102] which relies on a metric structure and Serre's definition [165] which relies on the homotopy lifting property alone. The threads for further algebraization may be found in Serre's and Borel's work – mod $\mathcal{C}$ theory is a precursor to localization and rational homotopy theory; the potential of the singular complex of a space as the fundamental object of study (not the space) is hinted at in Serre's thesis and realized later in the semi-simplicial theory of Moore and Kan; and the study of cobordism via Hopf algebra methods together with the Steenrod algebra action is based on the early computations of Borel for classifying spaces.

[33] See May's article in this volume. Also, Browder's investigations of the Bockstein spectral sequence hinge on input due to Cartan.

The period of development of algebraic topology examined in this brief history is remarkable. The rich atmosphere of difficult problems and untested methods around 1950 was ripe for the sudden realignment that occurred. Among the factors making these changes possible pedagogy played an unexpectedly important role – the spread of the crucial ideas was made possible by the high standard of exposition of the *Séminaire Cartan* and the subsequent clarity of the doctoral theses of Serre and Borel.

Throughout this period, there is a recognizable tension between competing points of view that might be characterized as analytic and combinatorial. The achievements of Leray in relative isolation revolve around a model of algebraic topology exemplified by the work of de Rham, close to the differentiable underpinning and free of the simplicial trappings. The combinatorial approach, exemplified by the work of Eilenberg and Henry Whitehead, beat a path to the homotopy theoretic invariants of spaces – a cruder but still powerful approach to geometric and topological questions. This essential tension is vital in the history of topology – Morse theory before and after the Second World War is a prime example and this history of spectral sequences must be taken as another example.

Acknowledgements

This paper presents a history of recent events. My thanks to the many participants in this story for their recollections, in particular to J.-L. Koszul, H. Cartan, J.-P. Serre, A. Borel, H. Samelson, W.S. Massey, G.W. Whitehead, J.C. Moore, S. MacLane, A.S. Schwarz, and P. Hilton. Conversations with Christian Houzel were most valuable. Thanks to Yves Felix for making Hirsch's selected papers available to me. Many others have contributed to a history of the later development of spectral sequences to which I hope to return in later papers. Also thanks to Stephan Klaus for abstracts on topology from the archives at Oberwohlfach. My thanks to Florence Fasanelli for conversations as well as a copy of her thorough and helpful unpublished Ph.D. thesis on the history of sheaves. The library at the Université Louis Pasteur in Strasbourg was invaluable in the preparation of this paper – my thanks to Christian Kassel and Véronique Bertrand for help there. I also thank Amelia Jones and Jim Stasheff for listening to early versions of this account and for their suggestions for improvement.

Bibliography

[1] J.W. Alexander, *Combinatorial Analysis Situs*, Trans. Amer. Math. Soc. **28** (1926), 301–329.

[2] J.W. Alexander, *On the chains of a complex and their duals; On the ring of a compact metric space*, Proc. Nat. Acad. Sci. USA **21** (1935), 509–511 and 511–512.

[3] J.W. Alexander, *On the connectivity ring of an abstract space*, Ann. of Math. **37** (1936), 698–708.

[4] P. Alexandroff, *General combinatorial topology*, Trans. Amer. Math. Soc. **49** (1941), 41–105.

[5] P. Alexandroff and H. Hopf, *Topologie*, Springer, Berlin (1935).

[6] A. Blakers and W.S. Massey, *The homotopy groups of a triad I–III*, Ann. of Math. (2) **53** (1951), 161–205; **55** (1952), 192–201; **58** (1953), 401–417.

[7] A. Borel, *Oeuvres*, Springer, New York (1983).

[8] A. Borel, *Remarques sur l'homologie filtrée*, J. Math. Pures Appl. (2) **29** (1950), 313–322.

[9] A. Borel, *Impossibilité de fibrer une sphère par un produit de sphères*, C. R. Acad. Sci. Paris **231** (1950), 943–945.

[10] A. Borel, *Sur la cohomologie des variétés de Stiefel et de certains groupes de Lie*, C. R. Acad. Sci. Paris **232** (1951), 1628–1630.
[11] A. Borel, *La transgression dans les espaces fibrés principaux*, C. R. Acad. Sci. Paris **232** (1951), 2392–2394.
[12] A. Borel, *Sur la cohomologie des espaces homogènes de groupes de Lie compacts*, C. R. Acad. Sci. Paris **233** (1951), 569–571.
[13] A. Borel, *Cohomologie des Espaces Localement Compacts, d'après Leray*, Séminaire de topologie algébrique, printemps 1951, EPF Zürich. See also Lecture Notes in Math. vol. 2, 3rd ed., Springer, Berlin (1964).
[14] A. Borel, *Sur la cohomologie des espaces fibrés principaux et des espaces homogènes des groupes de Lie compacts*, Ann. of Math. (2) **57** (1953), 115–207.
[15] A. Borel, *Sur l'homologie et la cohomologie des groupes de Lie compacts connexes*, Amer. J. Math. **76** (1954), 273–342.
[16] A. Borel, *Topology of Lie groups and characteristic classes*, Bull. Amer. Math. Soc. **61** (1955), 397–432.
[17] A. Borel, *Jean Leray and algebraic topology*, to appear in Selected Papers of Jean Leray, Topology, Springer, New York (1999).
[18] A. Borel and J.-P. Serre, *Impossibilité de fibrer un espace Euclidien par des fibres compactes*, C. R. Acad. Sci. Paris **230** (1950), 2258–2260.
[19] A. Borel and J.-P. Serre, *Détermination des p-puissances réduites de Steenrod dans la cohomologie des groupes classiques: Applications*, C. R. Acad. Sci. Paris **233** (1951), 680–682.
[20] A. Borel and J.-P. Serre, *Groupes de Lie et puissances réduites de Steenrod*, Amer. J. Math. **73** (1953), 409–448.
[21] R. Brauer, *Sur les invariants intégraux des variétés des groupes de Lie simple clos*, C. R. Acad. Sci. Paris **201** (1935), 419–421.
[22] É. Cartan, *Sur les nombres de Betti des espaces de groupes clos*, C. R. Acad. Sci. Paris **187** (1928), 196–198.
[23] É. Cartan, *Sur les invariants intégraux de certains espaces homogènes clos et les propriétés topologiques de ces espaces*, Ann. Soc. Math. Polon. **8** (1929), 181–225.
[24] É. Cartan, *La Topologie des Espaces Représentatifs des Groupes de Lie*, Acualités Scientifiques et Industrielles, No. 358, Hermann, Paris (1936).
[25] H. Cartan, *Oeuvres*, Vol. 3, Springer, New York (1979).
[26] H. Cartan, *Méthodes modernes en topologie algébrique*, Comment. Math. Helv. **18** (1945), 1–15.
[27] H. Cartan, *Sur la cohomologie des espaces où opère un groupe: Notions algébriques préliminaires; étude d'un anneau différentiel où opère un groupe*, C. R. Acad. Sci. Paris **226** (1948), 148–150 and 303–305.
[28] H. Cartan, *Sur la notion de carapace en topologie algébrique*, Colloque de Topologie (1947), CNRS, Paris (1949), 1–2.
[29] H. Cartan, *Une théorie axiomatique des carrés de Steenrod*, C. R. Acad. Sci. Paris **230** (1950), 425–427.
[30] H. Cartan, *Notions d'algèbre différentielle; applications aux groupes de Lie et aux variétés où opère un groupe de Lie*, Colloque de Topologie, Bruxelles (1950), CBRM, Liège (1951), 15–27.
[31] H. Cartan, *La transgression dans une groupe de Lie et dans un espace fibré principal*, Colloque de Topologie, Bruxelles (1950), CBRM, Liège (1951) 51–71.
[32] H. Cartan, *Sur les groupes d'Eilenberg–Mac Lane $H(\pi, n)$ I; Méthode des constructions II*, Proc. Nat. Acad. Sci. USA **40** (1954), 467–471 and 704–707.
[33] H. Cartan, *Algebraic Topology*, Lectures, G. Springer and H. Pollack, eds, Harvard Univ. Press, Cambridge, MA (1948).
[34] H. Cartan and J. Leray, *Relations entre anneaux de cohomologie et groupe de Poincaré*, Colloque de Topologie (1947), CNRS, Paris (1949), 83–85.
[35] H. Cartan and J.-P. Serre, *Espaces fibrés et groupes d'homotopie: I Constructions générales; II Applications*, C. R. Acad. Sci. Paris **234** (1952), 288–290, 393–395.
[36] H. Cartan and S. Eilenberg, *Homological Algebra*, Princeton Univ. Press, Princeton, NJ (1956).
[37] E. Čech, *Théorie générale de l'homologie dans un espace quelconque*, Fund. Math. **19** (1932), 149–183.
[38] E. Čech, *Multiplications on a complex*, Ann. of Math. **37** (1936), 681–697.
[39] S.S. Chern, *Characteristic classes of Hermitian manifolds*, Ann. of Math. (2) **47** (1946), 85–121.
[40] S.S. Chern, *On the multiplication in the characteristic ring of a sphere bundle*, Ann. of Math. **49** (1948), 362–372.
[41] S.S. Chern and E.H. Spanier, *The homology structure of fibre bundles*, Proc. Nat. Acad. Sci. USA **36** (1950), 248–255.

[42] C. Chevalley, *Theory of Lie Groups I*, Princeton Univ. Press, Princeton, NJ (1946).

[43] C. Chevalley and S. Eilenberg, *Cohomology theory of Lie groups and Lie algebras*, Trans. Amer. Math. Soc. **63** (1948), 84–124.

[44] B. Eckmann, *Zur Homotopietheorie gefaserter Räume*, Comment. Math. Helv. **14** (1942), 141–192.

[45] B. Eckmann, *Systeme von Richtungsfelder in Sphären und stetige Lösungen komplexer linearer Gleichungen*, Comment. Math. Helv. **15** (1943), 1–26.

[46] B. Eckmann, *Stetige Lösungen linearer Gleichungssysteme*, Comment. Math. Helv. **15** (1943), 318–339.

[47] B. Eckmann, *Der Cohomologie-Ring einer beliebigen Gruppe*, Comment. Math. Helv. **18** (1945/1946), 232–282.

[48] B. Eckmann, *Espaces fibrés et homotopie*, Colloque de Topologie, Bruxelles (1950), CBRM, Liège (1951), 83–99.

[49] B. Eckmann, H. Samelson and G.W. Whitehead, *On fibering spheres by toruses*, Bull. Amer. Math. Soc. **55** (1949), 433–438.

[50] C. Ehresmann, *Sur la théorie des espaces fibrés*, CNRS Colloque Internat. de Topologie Algebrique, Paris (1949), 3–15.

[51] C. Ehresmann and J. Feldbau, *Sur les propriétés d'homotopie des espaces fibrés*, C. R. Acad. Sci. Paris **212** (1941), 945–948.

[52] S. Eilenberg, *On the relation between the fundamental group of a space and the higher homotopy groups*, Fund. Math. **32** (1939), 167–175.

[53] S. Eilenberg, *Cohomology and continuous mappings*, Ann. of Math. (2) **41** (1940), 231–251. See also, *Lectures in Topology*, Univ. of Michigan Press (1941), 57–100.

[54] S. Eilenberg, *Singular homology*, Ann. of Math. (2) **45** (1944), 407–447.

[55] S. Eilenberg, *Homology of spaces with operators I*, Trans. Amer. Math. Soc. **61** (1947), 378–417.

[56] S. Eilenberg, *Topological methods in abstract algebra*, Bull. Amer. Math. Soc. **55** (1949), 3–27.

[57] S. Eilenberg, *On the problems of topology*, Ann. of Math. (2) **50** (1949), 247–260.

[58] S. Eilenberg and S. Mac Lane, *Group extensions and homology*, Ann. of Math. (2) **43** (1942), 758–831.

[59] S. Eilenberg and S. Mac Lane, *Relations between homology and homotopy groups*, Proc. Nat. Acad. Sci. USA **29** (1943), 155–158.

[60] S. Eilenberg and S. Mac Lane, *Relations between homology and homotopy groups of spaces I, II*, Ann. of Math. (2) **46** (1945), 480–509 and **51** (1950), 514–533.

[61] S. Eilenberg and S. Mac Lane, *Determination of the second homology and cohomology groups of a space by means of homotopy invariants*, Proc. Nat. Acad. Sci. USA **32** (1946), 277–280.

[62] S. Eilenberg and S. Mac Lane, *Cohomology and Galois theory I. Normality of algebras and Teichmuller's cocycle*, Trans. Amer. Math. Soc. **64** (1948), 1–20.

[63] S. Eilenberg and S. Mac Lane, *Homology of spaces with operators II*, Trans. Amer. Math. Soc. **65** (1949), 49–99.

[64] S. Eilenberg and S. Mac Lane, *Theory of Abelian groups and homotopy theory I–IV*, Proc. Nat. Acad. Sci. USA **36** (1950), 443–447, 657–663; **37** (1951), 307–310; **38** (1952), 325–329.

[65] S. Eilenberg and N.E. Steenrod, *Axiomatic approach to homology theory*, Proc. Nat. Acad. Sci. USA **31** (1945), 177–180.

[66] S. Eilenberg and N.E. Steenrod, *Foundations of Algebraic Topology*, Princeton Univ. Press, Princeton, NJ (1952).

[67] S. Eilenberg and J. Zilber, *On products of complexes*, Amer. J. Math. **75** (1953), 200–204.

[68] D.K. Faddeev, *On the theory of homology in groups*, Izv. Akad. Nauk SSSR **16** (1952), 17–22.

[69] F. Fasanelli, *The creation of sheaf theory*, Ph.D. thesis (1981).

[70] J. Feldbau, *Sur la classification des espaces fibrés*, C. R. Acad. Sci. Paris **208** (1939), 1621–1623.

[71] H. Freudenthal, *Über die Klassen der Sphärenabbildungen*, Comp. Math. **5** (1937), 299–314.

[72] H. Freudenthal, *Zum Hopfschen Umkehrshomomorphismus*, Ann. of Math. (2) **38** (1937), 847–853.

[73] H. Freudenthal, *Der Einfluss der Fundamentalgruppe auf die Bettischen Gruppen*, Ann. of Math. **47** (1946), 274–316.

[74] R. Godement, *Topologie Algébrique et Théorie des Faisceaux*, Publ. Inst. Math. Strasbourg, XII, Hermann, Paris (1958).

[75] A. Grothendieck, *Sur quelques points d'algèbre homologique*, Tôhoku Math. J. (2) **9** (1958), 119–221.

[76] W. Gysin, *Zur Homologietheorie der Abbildungen und Faserungen der Mannigfaltigkeiten*, Comment. Math. Helv. **14** (1941), 61–121.

[77] P.J. Hilton and S. Wylie, *Homology Theory: An Introduction to Algebraic Topology*, Cambridge Univ. Press, New York (1960).
[78] G. Hirsch, *Un isomorphisme attaché aux structures fibrées*, C. R. Acad. Sci. Paris **227** (1948), 1328–1330.
[79] G. Hirsch, *L'anneau de cohomologie d'un espace fibré et les classes caractéristiques*, C. R. Acad. Sci. Paris **229** (1949), 1297–1299.
[80] G. Hirsch, *Quelques relations entre l'homologie dans les espaces fibrés et les classes caractéristiques relatives à un groupe de structure*, Colloque de Topologie, Bruxelles (1950), CBRM, Liège (1951), 123–136.
[81] F. Hirzebruch, *On Steenrod's reduced powers, the index of inertia and the Todd genus*, Proc. Nat. Acad. Sci. USA **39** (1953), 110–114.
[82] F. Hirzebruch, *Some problems on differentiable and complex manifolds*, Ann. of Math. **60** (1954), 213–236.
[83] G.P. Hochschild, *Local class field theory*, Ann. of Math. **51** (1950), 331–347.
[84] G.P. Hochschild and J.-P. Serre, *Cohomology of group extensions*, Trans. Amer. Math. Soc. **74** (1953), 110–134.
[85] G.P. Hochschild and J.-P. Serre, *Cohomology of Lie algebras*, Ann. of Math. **57** (1953), 591–603.
[86] H. Hopf, *Selecta*, Springer, New York (1964).
[87] H. Hopf, *Abbildungsklassen n-dimensionaler Mannigfaltigkeiten*, Math. Ann. **96** (1926), 209–224.
[88] H. Hopf, *Über die Abbildungen der dreidimensionalen Sphären auf die Kugelfläche*, Math. Ann. **104** (1931), 637–665.
[89] H. Hopf, *Die Klassen der Abbildungen der n-dimensionalen Polyeder auf die n-dimensionalen Sphäre*, Comment. Math. Helv. **5** (1933), 39–54.
[90] H. Hopf, *Über die Abbildungen von Sphären auf Sphären von niedriger Dimension*, Fund. Math. **25** (1935), 427–440.
[91] H. Hopf, *Quelques problèmes de la théorie des représentations continues*, Enseign. Math. **35** (1936), 334.
[92] H. Hopf, *Über die Topologie der Gruppenmannigfaltigkeiten und ihre Verallgemeinerungen*, Ann. of Math. **42** (1941), 22–52.
[93] H. Hopf, *Über den Rang geschlossener Liescher Gruppen*, Comment. Math. Helv. **13** (1940/1941), 119–143.
[94] H. Hopf, *Fundamentalgruppe und zweite Bettische Gruppe*, Comment. Math. Helv. **17** (1942), 257–309.
[95] H. Hopf, *Nachtrag zu der Arbeit "Fundamentalgruppe und zweite Bettische Gruppe"*, Comment. Math. Helv. **15** (1942), 27–32.
[96] H. Hopf, *Die n-dimensionalen Sphären und projektiven Räume in der Topologie*, Proc. of ICM, Vol. I, Cambridge, MA (1950), Amer. Math. Soc., Providence, RI (1951), 193–202.
[97] H. Hopf and H. Samelson, *Ein Satz über die Wirkungsräume geschlossener Liescher Gruppen*, Comment. Math. Helv. **13** (1940/1941), 241–251.
[98] C. Houzel, *A short history: Les débuts de la théorie des faisceaux*, Sheaves on Manifolds, M. Kashiwara and P. Schapira, eds, Springer, New York (1990).
[99] S.T. Hu, *Homotopy Theory*, Academic Press, New York, NY (1959).
[100] W. Hurewicz, *Beiträge zur Topologie der Deformationen: I Höherdimensionalen Homotopiegruppen; II Homotopie- und Homologiegruppen; III Klassen und Homologietypen von Abbildungen; IV Asphärische Räume*, Proc. Akad. Wetensch. Amsterdam **38** (1935), 112–119, 521–528; **39** (1936), 117–126, 215–224.
[101] W. Hurewicz, *On duality theorems*, Bull. Amer. Math. Soc. **47** (1941), 562–563.
[102] W. Hurewicz and N.E. Steenrod, *Homotopy relations in fibre spaces*, Proc. Nat. Acad. Sci. USA **27** (1941), 60–64.
[103] I.M. James, *Reminiscences of a topologist*, Math. Intelligencer **12** (1990), 50–55.
[104] J. Kelley and E. Pitcher, *Exact homomorphisms in homology theory*, Ann. of Math. **48** (1947), 682–709.
[105] A.N. Kolmogoroff, *Über die Dualität im Aufbau der kombinatorischen Topologie*, Mat. Sb. **43** (1936), 97–102.
[106] A.N. Kolmogoroff, *Les groupes de Betti des espaces localement bicompacts; Propriétés des groupes de Betti des espaces localement bicompacts; Les groupes de Betti des espaces métriques; Cycles relatifs. Théorème de dualité de M. Alexander*, C. R. Acad. Sci. Paris **202** (1936), 1144–1147, 1325–1327, 1558–1560, and 1641–1643.
[107] J.-L. Koszul, *Sur les opérateurs de dérivation dans un anneau*, C. R. Acad. Sci. Paris **224** (1947), 217–219.
[108] J.-L. Koszul, *Sur les espaces homogènes*, C. R. Acad. Sci. Paris **224** (1947), 477-479.
[109] J.-L. Koszul, *Homologie et cohomologie des algèbres de Lie*, Bull. Soc. Math. France **78** (1950), 65–127.

[110] J.-L. Koszul, *Sur un type d'algèbres différentielles en rapport avec la transgression*, Colloque de Topologie, Bruxelles (1950), CBRM, Liège (1951), 73–81.

[111] T. Kudo, *Homological properties of fibre bundles*, J. Inst. Poly. Osaka City Univ. (A) **1** (1950), 101–114.

[112] T. Kudo, *Homological structure of fibre bundles*, J. Inst. Poly. Osaka City Univ. (A) **2** (1952), 101–140.

[113] H. Künneth, *Über die Bettischen Zahlen einer Produktmannigfaltigkeit*, Math. Ann. **90** (1923), 65–85.

[114] H. Künneth, *Über die Torsionzahlen von Produktmannigfaltigkeiten*, Math. Ann. **91** (1924), 125–134.

[115] S. Lefschetz, *On singular chains and cycles*, Bull. Amer. Math. Soc. **39** (1933), 124–129.

[116] S. Lefschetz, *Algebraic Topology*, Amer. Math. Soc. Colloq. Series vol. 27, Providence, RI (1942).

[117] J. Leray, *Topologie des espaces de Banach*, C. R. Acad. Sci. Paris **200** (1935), 1082–1084.

[118] J. Leray, *Les complexes d'un espace topologique*, C. R. Acad. Sci. Paris **214** (1942), 781–783.

[119] J. Leray, *L'homologie d'un espace topologique*, C. R. Acad. Sci. Paris **214** (1942), 839–841.

[120] J. Leray, *Les équations dans les espaces topologiques*, C. R. Acad. Sci. Paris **214** (1942), 897–899.

[121] J. Leray, *Transformations et homéomorphies dans les espaces topologiques*, C. R. Acad. Sci. Paris **214** (1942), 938–940.

[122] J. Leray, *Sur la forme des espaces topologiques et sur les points fixes des représentations; Sur la position d'un ensemble fermé de points d'un espace topologique; Sur les équations et les transformations*, J. Math. Pures Appl. (9) **24** (1945), 95–167, 169–199, and 201–248.

[123] J. Leray, *L'anneau d'homologie d'une représentation*, C. R. Acad. Sci. Paris **222** (1946), 1366–1368.

[124] J. Leray, *Structure de l'anneau d'homologie d'une représentation*, C. R. Acad. Sci. Paris **222** (1946), 1419–1422.

[125] J. Leray, *Propriétés de l'anneau d'homologie d'une représentation*, C. R. Acad. Sci. Paris **223** (1946), 395–397.

[126] J. Leray, *Sur l'anneau d'homologie de l'espace homogène d'un groupe clos par un sous-groupe Abélien, connexe, maximum*, C. R. Acad. Sci. Paris **223** (1946), 412–415.

[127] J. Leray, *L'homologie filtrée*, Colloque de Topologie, Paris (1947), CNRS, Paris (1949), 61–82.

[128] J. Leray, *Applications continues commutant avec les éléments d'un groupe de Lie*, C. R. Acad. Sci. Paris **228** (1949), 1748–1786.

[129] J. Leray, *Détermination, dans le cas non exceptionnels, de l'anneau de cohomologie de l'espace homogène quotient d'un groupe de Lie compact par un sous-groupe de même rang*, C. R. Acad. Sci. Paris **228** (1949), 1902–1904.

[130] J. Leray, *L'anneau spectral et l'anneau filtré d'un espace localement compact et d'une application continue*, J. Math. Pures Appl. (9) **29** (1950), 1–139.

[131] J. Leray, *L'homologie d'un espace fibré dont la fibre est connexe*, J. Math. Pures Appl. (9) **29** (1950), 169–213.

[132] J. Leray, *Sur l'homologie des groupes de Lie, des espaces homogènes et des espaces principaux*, Colloque de Topologie, Bruxelles (1950), CBRM, Liège (1951), 101–115.

[133] J. Leray, *La théorie des points fixes et ses applications en analyse*, Proc. Internat. Congress Math., Cambridge, Vol. 2 (1950), 202–208.

[134] J. Leray and J. Schauder, *Topologie et équations fonctionnelles*, Ann. ENS **51** (1934), 43–78.

[135] A. Lichnerowicz, *Un théorème sur l'homologie dans les espaces fibrés*, C. R. Acad. Sci. Paris **227** (1948), 711–712.

[136] R.C. Lyndon, *The cohomology theory of group extensions*, Duke Math. J. **15** (1948), 271–292.

[137] S. Mac Lane, *Homology*, Springer, New York (1963).

[138] S. Mac Lane, *Origins of the cohomology of groups*, Topology and Algebra, Proc. of a Conference in Honor of Beno Eckmann, Enseign. Math. (1978), 191–219.

[139] W.S. Massey, *Exact couples in algebraic topology*, I–II; III–V, Ann. of Math. **56** (1952), 363–396; **57** (1953), 248–286.

[140] W.S. Massey, *Some new algebraic methods in topology*, Bull. Amer. Math. Soc. **60** (1954), 111–123.

[141] W.S. Massey, *Some problems in algebraic topology and the theory of fibre bundles*, Ann. of Math. (2) **62** (1955), 327–359.

[142] D. Montgomery and H. Samelson, *Fiberings with singularities*, Duke Math. J. **13** (1946), 51–56.

[143] J.C. Moore, *Some applications of homology theory to homotopy problems*, Ann. of Math. **58** (1953), 325–350.

[144] J.C. Moore, *On homotopy groups of spaces with a single nonvanishing homology group*, Ann. of Math. **59** (1954), 549–557.

[145] M. Morse, *The Calculus of Variations in the Large*, Amer. Math. Soc. Colloq. Series vol. 18, Providence, RI (1934).
[146] L. Pontrjagin, *Homologies in compact Lie groups*, Mat. Sb. **6** (1939), 389–422.
[147] L. Pontrjagin, *A classification of the mappings of a 3-dimensional complex into the 2-dimensional sphere*, Mat. Sb. **9** (1941), 331–363.
[148] L. Pontrjagin, *A classification of the mappings of a 3-dimensional sphere into an n-dimensional complex*, Dokl. Akad. Nauk USSR **35** (1942), 34–37.
[149] L. Pontrjagin, *Characteristic classes of differential manifolds*, Mat. Sb. **21** (1947), 233–284 (Translated by the AMS in 1950).
[150] L. Pontrjagin, *Homotopy classification of the mappings of an $(n+2)$-dimensional sphere on an n-dimensional*, Dokl. Akad. Nauk USSR **70** (1950), 957–959.
[151] M.M. Postnikov, *Determination of the homology groups of a space by means of the homotopy invariants*, Dokl. Akad. Nauk SSSR **79** (1951), 573–576.
[152] K. Reidemeister, *Homotopieringe und Linsenräume*, Hamburg. Abh. **11** (1935), 102–109.
[153] G. de Rham, *Sur l'Analysis situs des variétés à n dimensions*, J. Math. Pures Appl. (9) **10** (1931), 115–200.
[154] G. de Rham, *Sur la théorie des intersections et les intégrales multiples*, Comment. Math. Helv. **4** (1932), 151–157.
[155] H. Samelson, *Beiträge zur Topologie der Gruppenmannigfaltigkeiten*, Ann. of Math. **42** (1941), 1091–1137.
[156] H. Samelson, *Topology of Lie groups*, Bull. Amer. Math. Soc. **58** (1952), 2–37.
[157] H. Seifert and W. Threlfall, *Lehrbuch der Topologie*, Teubner, Leipzig (1934).
[158] H. Seifert and W. Threlfall, *Variationsrechnung im Grossen (Theorie von Morse)*, Teubner, Leipzig (1939).
[159] *Séminaire H. Cartan*, Ecole Normale Supérieure, 1948/1949: Topologie algébrique; 1949/1950: Espaces fibrés; 1950/1951: Cohomologie des groupes, suites spectrales, et faisceaux; 1954/1955: Algèbres d'Eilenberg–Mac Lane et homotopie, Secrétariat mathématique, Paris (1955/1956).
[160] J.-P. Serre, *Oeuvres*, Vol. 1–3, Springer, Berlin (1986). See particularly Vol. 1.
[161] J.-P. Serre, *Compacité locale des espaces fibrés*, C. R. Acad. Sci. Paris **229** (1949), 1295–1297.
[162] J.-P. Serre, *Trivialité des espaces fibrés. Applications*, C. R. Acad. Sci. Paris **230** (1950), 916–918.
[163] J.-P. Serre, *Cohomologie des extensions de groupes*, C. R. Acad. Sci. Paris **231** (1950), 643–646.
[164] J.-P. Serre, *Homologie singulière des espaces fibrés: I La suite spectrale; II Les espaces de lacets; III Applications homotopiques*, C. R. Acad. Sci. Paris **231** (1950), 1408–1410; **232** (1951), 31–33 and 142–144.
[165] J.-P. Serre, *Homologie singulière des espaces fibrés: Applications*, Ann. of Math. **54** (1951), 425–505.
[166] J.-P. Serre, *Sur les groupes d'Eilenberg–Mac Lane*, C. R. Acad. Sci. Paris **234** (1952), 1243–1245.
[167] J.-P. Serre, *Sur la suspension de Freudenthal*, C. R. Acad. Sci. Paris **234** (1952), 1340–1342.
[168] J.-P. Serre, *Groupes d'homotopie et classes de groupes Abéliens*, Ann. of Math. **58** (1953), 258–294.
[169] J.-P. Serre, *Cohomologie modulo 2 des complexes d'Eilenberg–Mac Lane*, Comment. Math. Helv. **27** (1953), 198–232.
[170] J.-P. Serre, *Quelques calculs de groupes d'homotopie*, C. R. Acad. Sci. Paris **236** (1953), 2475–2477.
[171] J.-P. Serre, *Lettre à Armand Borel, April 16, 1953*, Oeuvres, Vol. 1, 243–250.
[172] E.H. Spanier, *Cohomology theory for general spaces*, Ann. of Math. (2) **49** (1948), 407–427.
[173] N.E. Steenrod, *Topological methods for construction of tensor functions*, Ann. of Math. **43** (1942), 116–131.
[174] N.E. Steenrod, *Homology with local coefficients*, Ann. of Math. (2) **44** (1943), 610–627.
[175] N.E. Steenrod, *Classification of sphere bundles*, Ann. of Math. (2) **45** (1944), 294–311.
[176] N.E. Steenrod, *Products of cocycles and extensions of mappings*, Ann. of Math. (2) **48** (1947), 290–320.
[177] N.E. Steenrod, *Cohomology invariants of mappings*, Ann. of Math. (2) **50** (1949), 954–988.
[178] N.E. Steenrod, *Reduced powers of cohomology classes*, Proc. ICM, Vol. 1, Cambridge (1950), Amer. Math. Soc., Providence, RI (1951), 530.
[179] N.E. Steenrod, *The Topology of Fibre Bundles*, Princeton Univ. Press, Princeton, NJ (1951).
[180] N.E. Steenrod, *Reduced powers of cohomology classs*, Ann. of Math. **56** (1952), 47–67.
[181] N.E. Steenrod, *The conference on fiber bundles and differential geometry in Ithaca*, Bull. Amer. Math. Soc. **59** (1953), 569–570.
[182] N.E. Steenrod, *Reviews of Papers in Algebraic and Differential Topology, Topological Groups, and Homological Algebra*, Parts I and II, Amer. Math. Soc., Providence, RI (1968).
[183] E. Stiefel, *Richtungsfelder und Fernparallelismus in n-dimensionalen Mannigfaltigkeiten*, Comment. Math. Helv. **8** (1936), 3–51.

[184] E. Stiefel, *Über eine Beziehung zwischen geschossenen Lieschen Gruppen und diskontinuerlichen Bewegungsgruppen*, usw., Comment. Math. Helv. **14** (1941/1942), 350–380.

[185] R. Thom, *Sur une partition en cellules associée à une fonction sur une variété*, C. R. Acad. Sci. Paris **228** (1949), 973–975.

[186] R. Thom, *Espaces fibrés en sphères et carrés de Steenrod*, Ann. ENS **69** (1952), 109–181.

[187] R. Thom, *Quelques propriétés globales des variétés différentiables*, Comment. Math. Helv. **28** (1954), 17–86.

[188] H. Toda, *Generalized Whitehead products and homotopy groups of spheres*, J. Inst. Poly. Osaka City Univ. **3** (1952), 43–82.

[189] H.C. Wang, *The homology groups of the fiber bundles over a sphere*, Duke Math. J. **16** (1949), 33–38.

[190] A. Weil, *Oeuvres Scientifiques*, Vol. I and II, Springer, New York (1979).

[191] G.W. Whitehead, *Homotopy groups of spheres*, Proc. of the Internat. Congress of Mathematicians, Vol. 2, Cambridge, MA (1950), Amer. Math. Soc., Providence, RI (1952), 358–362.

[192] G.W. Whitehead, *The $(n+2)$-nd homotopy of the n-sphere*, Ann. of Math. **52** (1950), 245–248.

[193] G.W. Whitehead, *Fiber spaces and Eilenberg homology groups*, Proc. Nat. Acad. Sci. USA **38** (1952), 426–430.

[194] J.H.C. Whitehead, *The Mathematical Works of J.H.C. Whitehead*, Vol. 1–4, Pergamon, New York/London (1962).

[195] J.H.C. Whitehead, *On adding relations to homotopy groups*, Ann. of Math. **42** (1941), 409–428.

[196] J.H.C. Whitehead, *On the groups $\pi_r(V_{n,m})$ and sphere bundles*, Proc. London Math. Soc. **48** (1944), 243–291.

[197] J.H.C. Whitehead, *Combinatorial homotopy I, II*, Bull. Amer. Math. Soc. **55** (1949), 213–245 and 453–496.

[198] J.H.C. Whitehead, *On the realizability of homotopy groups*, Ann. of Math. **50** (1949), 261–263.

[199] J.H.C. Whitehead, *A certain exact sequence*, Ann. of Math. **52** (1950), 51–110.

[200] J.H.C. Whitehead, *Algebraic homotopy theory*, Proc. of the Internat. Congress of Mathematicians, Vol. 2, Cambridge, MA (1950), Amer. Math. Soc., Providence, RI (1952), 354–357.

[201] H. Whitney, *Sphere spaces*, Proc. Nat. Acad. Sci. USA **21** (1935), 464–468.

[202] H. Whitney, *Topological properties of differentiable manifolds*, Bull. Amer. Math. Soc. **43** (1936), 785–805.

[203] H. Whitney, *On maps of an n-sphere to another n-sphere*, Duke Math. J. **3** (1937), 46–50.

[204] H. Whitney, *On products in a complex*, Ann. of Math. **39** (1938), 397–432.

[205] H. Whitney, *Tensor products of Abelian groups*, Duke Math. J. **4** (1938), 495–528.

[206] H. Whitney, *On the theory of sphere bundles*, Proc. Nat. Acad. Sci. USA **26** (1940), 148–153.

[207] H. Whitney, *On the topology of differentiable manifolds*, Lectures in Topology, R. Wilder and W.L. Ayres, eds, Conf. at the University of Michigan (1940), Univ. of Michigan Press, Ann Arbor, MI (1941), 101–141.

[208] R. Wilder and W.L. Ayres (eds), *Lectures in Topology*, Conf. at the University of Michigan (1940), Univ. of Michigan Press, Ann Arbor, MI (1941).

[209] Wu, Wen-Tsün, *On the product of sphere bundles and the duality theorem modulo 2*, Ann. of Math. **49** (1948), 641–653.

[210] Wu, Wen-Tsün, *Classes caractéristiques et i-carrés d'une variété*, C. R. Acad. Sci. Paris **230** (1950), 508–509.

[211] Wu, Wen-Tsün, *Les i-carrés dans une variété Grassmannienne*, C. R. Acad. Sci. Paris **230** (1950), 918–920.

[212] Wu, Wen-Tsün, *Sur les classes caractéristiques des structures fibrées sphériques*, Publ. Inst. Math. Univ. Strasbourg, XI, Paris, Hermann (1952).

[213] Yen, Chih-Tah, *Sur les polynômes de Poincaré des groupes de Lie exceptionnels*, C. R. Acad. Sci. Paris **228** (1949), 628–630.

[214] G.S. Young, *On the factors and fiberings of manifolds*, Proc. Amer. Math. Soc. **1** (1950), 215–223.

This Page Intentionally Left Blank

CHAPTER 24

Stable Algebraic Topology, 1945–1966

J.P. May

Department of Mathematics, University of Chicago, 5734 University Avenue, Chicago, IL 60637-1546, USA
E-mail: may@math.chicago.edu

Stable algebraic topology is one of the most theoretically deep and calculationally powerful branches of mathematics. It is very largely a creation of the second half of the twentieth century. Roughly speaking, a phenomenon in algebraic topology is said to be "stable" if it occurs, at least for large dimensions, in a manner independent of dimension. While there are important precursors of the understanding of stable phenomena, for example in Hopf's introduction of the Hopf invariant [Hopf35, FS], Hurewicz's introduction of homotopy groups [Hur35], and Borsuk's introduction of cohomotopy groups [Bor36], the first manifestation of stability in algebraic topology appeared in Freudenthal's extraordinarily prescient 1937 paper [Fr37, Est], in which he proved that the homotopy groups of spheres are stable in a range of dimensions.

Probably more should be said about precursors, but I will skip ahead and begin with the foundational work that started during World War II but first reached print in 1945. Aside from the gradual development of homology theory, which of course dates back at least to Poincaré, some of the fundamental precursors are treated elsewhere in this volume [Ma, BG, Mc, We]. However, another reason for not attempting such background is that I am not a historian of mathematics, not even as a hobby. I am a working mathematician who is bemused by the extraordinarily rapid, and perhaps therefore jagged, development of my branch of the subject. I am less interested in who did what when than in how that correlated with the progression of ideas.

My theme is the transition from classical algebraic topology to stable algebraic topology, with emphasis on the emergence of cobordism, K-theory, generalized homology and cohomology, the stable homotopy category, and modern calculational techniques. The history is surprising, not at all as I imagined it. For one example, we shall see that the introduction of spectra was quite independent of the introduction of generalized cohomology theories. While some key strands developed in isolation, we shall see that there was a sudden coalescence around 1960: this was when the subject as we know it today began to take shape, although in far from its final form: I doubt that we are there yet even now.

Younger readers are urged to remember the difficulty of communication in those days. Even in 1964, when I wrote my thesis, the only way to make copies was to type using carbon paper: mimeographing was inconvenient and the Xerox machine had not been in-

HISTORY OF TOPOLOGY
Edited by I.M. James

vented, let alone fax or e-mail. Moreover, English had not yet become the lingua franca. Many relevant papers are in French or German (which I read) and some are in Russian, Spanish, or Japanese (which I do not read); further, the Iron Curtain hindered communication, and translation from the Russian was spotty. On the other hand, the number of people working in topology was quite small: most of them knew each other from conferences, and correspondence was regular. Moreover, the time between submission and publication of papers was shorter than it is today, usually no more than a year.

I have profited from a perusal of all of Steenrod's very helpful compendium [StMR] of Mathematical Reviews in algebraic and differential topology published between 1940 and 1967. Relatively few papers before the mid 1950's concern stable algebraic topology, whereas an extraordinary stream of fundamental papers was published in the succeeding decade. That stream has since become a torrent. I will focus on the period covered in [StMR], especially the years 1950 through 1966, which is an arbitrary but convenient cutoff date. For the later part of that period, I have switched focus a little, trying to give a fairly complete indication of the actual mathematical content of all of the most important relevant papers of the period. I shall also point out various more recent directions that can be seen in embryonic form during the period covered, but I shall not give references to the modern literature except in cases of direct follow up and completion of earlier work. I plan to try to bring the story up to date in a later paper, but lack of time has prevented me from attempting that now.

References to mathematical contributions give the year of publication, the only exception being that books based on lecture notes are dated by the year the lectures were given. References to historical writings are given without dates.

1. Setting up the foundations

A great deal of modern mathematics, by no means just algebraic topology, would quite literally be unthinkable without the language of categories, functors, and natural transformations introduced by Eilenberg and MacLane in their 1945 paper [EM45b]. It was perhaps inevitable that some such language would have appeared eventually. It was certainly not inevitable that such an early systematization would have proven so remarkably durable and appropriate; it is hard to imagine that this language will ever be supplanted.

With this language at hand, Eilenberg and Steenrod were able to formulate their axiomatization of ordinary homology and cohomology theory. The axioms were announced in 1945 [ES45], but their celebrated book "The foundations of algebraic topology" did not appear until 1952 [ES52], by which time its essential ideas were well known to workers in the field. It should be recalled that Eilenberg had set the stage with his fundamentally important 1940 paper [Eil40], in which he defined singular homology and cohomology as we know them today.

I will say a little about the axioms shortly, but another aspect of their work deserves immediate comment. They clearly and unambiguously separated the algebra from the topology. This was part of the separation of homological algebra from algebraic topology as distinct subjects. As discussed by Weibel [We], the subject of homological algebra was set on firm foundations in the comparably fundamental book "Homological algebra" of Cartan and Eilenberg [CE56].

Two things are conspicuously missing from Eilenberg–Steenrod. We think of it today as an axiomatization of the homology and cohomology of finite CW complexes, but in fact CW complexes are nowhere mentioned. The definitive treatment of CW complexes had been published by J.H.C. Whitehead in 1948 [Whi48], but they were not yet in regular use. Many later authors continued to refer to polyhedra where we would refer to finite CW complexes, and I shall sometimes take the liberty of describing their results in terms of finite CW complexes.

Even more surprisingly, Eilenberg–MacLane spaces are nowhere mentioned. These spaces had been introduced in 1943 [EM43, EM45a], and the relation

$$\widetilde{H}^n(X;\pi) \cong [X, K(\pi,n)] \tag{1.1}$$

was certainly known to Eilenberg and Steenrod. It seems that they did not believe it to be important. Nowadays, the proof of this relation is seen as one the most immediate and natural applications of the axiomatization.

However, there was something missing for the derivation of this relation. Despite their elementary nature, the theory of cofiber sequences and the dual theory of fiber sequences were surprisingly late to be formulated explicitly. They were implicit, at least, in Barratt's papers on "track groups" [Ba55], but they were not clearly articulated until the papers of Puppe [Pu58] and Nomura [Nom60]. The concomitant principle of Eckmann–Hilton duality also dates from the late 1950's [Eck57, EH58] (see also [Hil65]). The language of fiber and cofiber sequences pervades modern homotopy theory, and its late development contrasts vividly with the earlier introduction of categorical language. Probably not coincidentally, the key categorical notion of an adjoint functor was also only introduced in the late 1950's, by Kan [Kan58].

Although a little peripheral to the present subject, a third fundamental text of the early 1950's, Steenrod's "The topology of fiber bundles" [St51] nevertheless must be mentioned. In the first flowering of stable algebraic topology, with the introduction of cobordism and K-theory, the solidly established theory of fiber bundles was absolutely central to the translation of problems in geometric topology to problems in stable algebraic topology.

2. The Eilenberg–Steenrod axioms

The functoriality, naturality of connecting homomorphism, exactness, and homotopy axioms need no comment now, although their economy and clarity would not have been predicted from earlier work in the subject. Remember that these are axioms on the homology or cohomology of pairs of spaces. The crucial and subtle axiom is excision. A triad $(X; A, B)$ is *excisive* if X is the union of the interiors of A and B. In homology, the excision map $H_*(B, A\cap B) \to H_*(X, A)$ must be an isomorphism. One subtlety is that I have stated the axiom in the form that Eilenberg and Steenrod verify it for singular homology. With a view towards other theories, they state the axiom under the stronger hypothesis that B is closed in X.

Conveniently for later developments, the dimension axiom was stated last. The fundamental theorem is that homology and cohomology with a given coefficient group is unique on triangulable pairs or, more generally, on finite CW pairs.

Several important extensions of the axioms came later. First, one wants an axiom that characterizes ordinary homology and cohomology on general CW pairs. For that Milnor [Mil62a] added the additivity axiom. It asserts that homology converts disjoint unions to direct sums and cohomology converts disjoint unions to direct products. It implies that the homology of a CW complex X is the colimit of the homologies of its skeleta X^n. In cohomology, it implies $\lim^1$ exact sequences

$$0 \to \lim{}^1 H^{q-1}(X^n) \to H^q(X) \to \lim H^q(X^n) \to 0. \tag{2.1}$$

This allows the extension of the uniqueness theorem to infinite CW pairs.

One next wants an axiom that distinguishes singular theories from other theories on general pairs of spaces. I do not know who first formulated it; it appears in [Swi75] and may be due to Adams. This is the weak equivalence axiom. It asserts that a weak equivalence of pairs induces an isomorphism on homology and cohomology. Any space is weakly equivalent to a CW complex, any pair of spaces is weakly equivalent to a CW pair, and any excisive triad is weakly equivalent to a triad that consists of a CW complex X and a pair of subcomplexes A and B with union X. Here $B/A \cap B \cong X/A$ as CW complexes, which neatly explains the excision axiom. The weak equivalence axiom reduces computation of the homology and cohomology of general pairs to their computation on CW pairs. Thus it implies the uniqueness theorem for homology and cohomology on general pairs.

Finally, one wants an axiom system for the reduced homology and cohomology of based spaces. The earliest published account is in the 1958 paper [DT58] of Dold and Thom, who ascribe it to Puppe. They use it to prove that the homotopy groups of the infinite symmetric products $SP^\infty X$ of based spaces X can be computed as the reduced integral homology groups of X. There are several slightly later papers [Ke59, BP60, Hu60] devoted to single space axioms for the homology and cohomology of both based spaces and, curiously, unbased spaces.

For the reduced homology of nondegenerately based spaces, the axioms just require functors $\tilde{k}_q$ together with natural suspension isomorphisms

$$\Sigma_* : \tilde{k}_q(X) \cong \tilde{k}_{q+1}(\Sigma X) \tag{2.2}$$

that satisfy the exactness, wedge, and weak equivalence axioms. Here the exactness axiom requires the sequences

$$\tilde{k}_q(X) \xrightarrow{f_*} \tilde{k}_q(Y) \to \tilde{k}_q(Cf) \tag{2.3}$$

to be exact for a map $f : X \to Y$ with cofiber $Cf = Y \cup_f CX$. The wedge axiom requires the functors $\tilde{k}_q$ to carry wedges (1-point unions) to direct sums. The weak equivalence axiom requires a weak equivalence to induce isomorphisms on all homology groups. Given such a reduced homology theory, one obtains an unreduced homology theory by setting $k_q(X) = \tilde{k}_q(X_+)$, where X_+ is the union of X and a disjoint basepoint, and

$$k_q(X, A) = \tilde{k}(Cf),$$

where $f : A \to X$ is the inclusion. For an unreduced homology theory k_*, one obtains a reduced homology theory by setting $\tilde{k}_q(X) = k_q(X, *)$. Thus reduced and unreduced homology theories are equivalent notions. The same is true for cohomology theories. The summary in this paragraph makes no reference to the dimension axiom and applies in general.

In view of (2.2), all of ordinary homology and cohomology theory is actually part of stable algebraic topology. As an informal rule of thumb, when thinking in terms of classical algebraic topology, one uses unreduced theories. When thinking in terms of stable algebraic topology, one wants the suspension axiom to hold without qualification in all degrees and one therefore works with reduced theories. In fact, in a great deal of recent work, it is an accepted convention that k_* *means* reduced homology, and one writes $k_*(X_+)$ for unreduced homology. I shall not take that point of view here, however.

This summary of the axioms is skewed towards singular homology and cohomology. The viewpoint of someone working in, say, algebraic geometry would be quite different. However, there are two footnotes to the axioms that are not well known and may be worth mentioning. To characterize Čech cohomology on compact Hausdorff spaces, Eilenberg and Steenrod add the continuity axiom. Keesee [Kee51] observed that this axiom implies the homotopy axiom.

More substantively, let us go back to (1.1) above. If X has the homotopy type of a CW complex, then the square brackets denote homotopy classes of based maps. Huber [Hu61] proved that if X is a paracompact Hausdorff space, then the *Čech* cohomology group $\check{H}^n(X;\pi)$ is isomorphic to the set of homotopy classes of maps $X \to K(\pi, n)$. In contrast, for the general representation of *singular* cohomology in the form (1.1), we must understand $[X, K(\pi, n)]$ to be the set of maps in the category that is obtained from the homotopy category of based spaces by adjoining formal inverses to the weak equivalences; equivalently, we must replace X by a CW complex weakly equivalent to it before taking homotopy classes of maps.

3. Stable and unstable homotopy groups

Another important precursor of stable algebraic topology was a substantial increase in the understanding of the relationship between stable and unstable homotopy groups and of certain fundamental exact sequences relating homotopy groups in different dimensions. I am here thinking of what was achieved by bare hands work, in the early to mid 1950's, using CW complexes and homotopical methods rather than the contemporaneous and overlapping progress that came with the introduction of spectral sequences.

We have seen that the critical axiom for homology is excision. In the early 1950's, Blakers and Massey [BM51, BM52, BM53] made a systematic study of excision in homotopy theory, proving that homotopy groups satisfy the excision axiom in a range of dimensions. This gave a new proof of the Freudenthal suspension theorem and considerably clarified the conceptual relationship between homology and homotopy. The proofs were quite difficult, and it soon became fashionable to prove versions of their results using homology and spectral sequences. However, Boardman later came up with a quite accessible direct homotopical proof, which is presented in [Swi75], for example. It is worth emphasizing that the homotopical proof gives a stronger result than can be obtained by homological methods.

The Freudenthal suspension theorem establishes the stable range for homotopy groups, roughly twice the connectivity of a space. It was shown by G.W. Whitehead [Wh53] that there is a metastable range for the homotopy groups of spheres. The suspension homomorphism E fits into the EHP exact sequence

$$\cdots \to \pi_q(S^n) \xrightarrow{E} \pi_{q+1}(S^{n+1}) \xrightarrow{H} \pi_{q+1}(S^{2n+1}) \xrightarrow{P} \pi_{q-1}(S^n) \xrightarrow{E} \pi_q(S^{n+1}) \to \cdots$$

when $q \leqslant 3n-2$. Here H is a (generalized) Hopf invariant that Whitehead had introduced earlier [Wh50] and P is the (J.H.C.) Whitehead product. There were many extensions and refinements of these results. For example, Hilton [Hil51] gave a definition of the Hopf invariant in the next range of dimensions, $q < 4n$, in the sequence above. The extrapolation of calculations and understanding in stable homotopy theory to calculations and understanding in the metastable range, and further, has been a major theme ever since.

James [Ja55, Ja56a, Ja56b, Ja57] and Toda [To62a] went much further with this. James proved that, on 2-primary components, there is an EHP exact sequence that is valid for all values of q, and Toda proved an appropriate analogue for odd primes. James introduced the James construction JX for the purpose. Here JX is the free topological monoid generated by a based space X. For a connected CW complex X, James proved that JX is homotopy equivalent to $\Omega \Sigma X$. The space JX comes with a natural filtration, and its simple combinatorial structure allows direct construction of suitable Hopf invariant maps. Milnor [Mil56b] proved that ΣJX splits up to homotopy as the wedge of the suspensions of its filtration quotients. These arguments were the prototypes for a great deal of later work in which combinatorial approximations to the n-fold loop spaces $\Omega^n \Sigma^n X$ have been used to obtain stable decompositions of such spaces, leading to a great deal of new calculational information in stable homotopy theory. However, this goes beyond the present story.

The power and limitations of such direct homotopical methods of calculation are well illustrated in Toda's series of papers [To58a, To58b, To58c, To59] and monograph [To62b]; while cohomology operations, spectral sequences, and the method of killing homotopy groups are used extensively, most of the work in these calculations of the groups $\pi_{n+k}(S^n)$ for small k consists of direct elementwise inductive arguments in the EHP sequence. Later work of this sort gave these groups for a few more values of k, but it was apparent that this was not the route towards major progress in the determination of the homotopy groups of spheres.

4. Spectral sequences and calculations in homology and homotopy

Although the credit for the invention of spectral sequences belongs to Leray [Le49, Mc], for algebraic topology the decisive introduction of spectral sequences is due to Serre [Se51]. For a fibration $p: E \to B$ with connected base space B and fiber F, the Serre spectral sequence in homology has $E^2_{p,q} = H_p(B; H_q(F; \pi))$, where local coefficients are understood, and it converges in total degree $p+q$ to $H_*(E; \pi)$. The analogous cohomology spectral sequence with coefficients in a commutative ring π is a spectral sequence of differential algebras, and it converges to the associated graded algebra of $H^*(E; \pi)$ with respect to a suitable filtration.

With the Serre spectral sequence, algebraic topology emerged as a subject in which substantial calculations could be made. While its applications go far beyond our purview, many of the calculations that it made possible and ideas to which it led were essential prerequisites to the emergence of stable algebraic topology.

Work of Borel [Bo53a, Bo53b] and others gave a systematic understanding of the homology and cohomology of the classical Lie groups and of their classifying spaces and homogeneous spaces. The basic characteristic classes had all been defined earlier, but the precise detailed analysis of the various cohomology algebras and their induced maps was vital to future progress.

Serre's introduction of class theory [Se53a], and his use of the spectral sequence to prove the finiteness of the homotopy groups of spheres, save for $\pi_n(S^n)$ and $\pi_{4n-1}(S^{2n})$, were to change the way people thought about algebraic topology. Earlier calculations had generally had as their goal the understanding of homology and cohomology with integer or with real coefficients. In the years since, calculations have largely focused on mod p homology and cohomology, especially in stable algebraic topology where the rational theory is essentially trivial. Moreover, this change in point of view led ultimately to the study of all of homotopy theory in terms of localized and completed spaces.

The method of killing homotopy groups introduced by Cartan and Serre [CS52a, CS52b] was also profoundly influential. It provided the first systematic route to the computations of homotopy groups. The idea is easy enough. Let X be a simple space. Inductively, by killing homotopy groups and passing to homotopy fibers, one can construct a sequence of fibrations

$$p_n : X[n+1, \infty) \to X[n, \infty)$$

with fibre $K(\pi_n(X), n-1)$, where $X[n, \infty)$ is $(n-1)$-connected and its higher homotopy groups are those of X. The initial map p_1 is just the universal covering of X. Assuming that one knows the first n homotopy groups of X, one should have enough inductive control on the space $X[n, \infty)$ to use the Serre spectral sequence to compute $H_{n+1}(X[n+1, \infty))$, which by the Hurewicz isomorphism is $\pi_{n+1}(X)$. This is closely related to Postnikov systems [Pos51a, Pos51b, Pos51c], which were not yet available to Cartan and Serre and so were implicitly reinvented by them. If $i_n : X \to X_n$ is the n-th term of the Postnikov tower of X, then i_n induces an isomorphism on π_q for $q \leqslant n$ and the higher homotopy groups of X_n are zero; $X[n+1, \infty)$ is the homotopy fiber of i_n.

An interesting companion to this method was given in Moore's study [Mo54] of the homotopy groups of spaces with a single nonvanishing homology group, which are now called Moore spaces. This work led later to the introduction of the mod p homotopy groups of spaces. Cohomotopy groups with coefficients were introduced and studied earlier, by Peterson [Pe56a, Pe56b]. Moore also gave a functorial, semi-simplicial, construction of Postnikov systems, in [Ca54-55] and [Mo58], which are sometimes called Moore–Postnikov systems as a result. This and related work of Moore in [Ca54-55], Heller in [He55], and especially Kan in [Kan55] and many later papers (see [May67]), began the modern systematic use of simplicial methods in algebraic topology.

5. Steenrod operations, $K(\pi, n)$'s, and characteristic classes

For the method of killing homotopy groups to be useful, one must know something about the cohomology of Eilenberg–MacLane spaces. The problem of calculating these cohomology groups was intensively studied by Eilenberg and MacLane, notably in [EM50], and was solved a few years later by Cartan [Ca54-55], using methods of homological algebra. However, Cartan's original answer was not in the form we know it today. In fact, in mod p cohomology for odd primes p, it is still not obvious how to correlate Cartan's calculations with the definitive calculations in terms of Steenrod operations.

I will not say anything about the invention and development of the basic properties of the Steenrod operations [St47, St52, St53a, St53b, St57, ST57] since that is interestingly discussed in [Ma, Wh1]. Steenrod and Epstein [SE62] published a systematic account of the results. Epstein [Ep66] later showed how to construct Steenrod operations in a general context of homological algebra. In fact, simply by separating the algebra from the topology, Steenrod's original definition can be adapted to a variety of situations in both topology and algebra [May70].

An essential point is that the Steenrod operations are stable, in the sense that the following diagrams commute, where $\mathbb{Z}_2$ is the field $\mathbb{Z}/2\mathbb{Z}$.

$$\begin{array}{ccc} \widetilde{H}^q(X,\mathbb{Z}_2) & \xrightarrow{Sq^i} & \widetilde{H}^{q+i}(X;\mathbb{Z}_2) \\ \Sigma^* \downarrow & & \downarrow \Sigma^* \\ \widetilde{H}^{q+1}(\Sigma X;\mathbb{Z}_2) & \xrightarrow[Sq^i]{} & \widetilde{H}^{q+1+i}(\Sigma X;\mathbb{Z}_2). \end{array} \tag{5.1}$$

The analogous diagram commutes for odd primes, where P^i has degree $2i(p-1)$.

Serre [Se53b] computed $H^*(K(\pi_2, n); \mathbb{Z}_2)$, where π_2 is cyclic of order 2, in modern terms: it is the free commutative algebra on suitable composites of Steenrod operations acting on the fundamental class $\iota_n \in H^n(K(\pi_2, n); \mathbb{Z}_2)$. The analogue for odd primes was worked out by Cartan in [Ca54-55], in later exposés that are in fact independent of his original calculations published in the same place.

Formulas for the iteration of the Steenrod operations were first proven by Adem [Adem52] at the prime 2 and by Adem and Cartan [Adem53, Adem57, Ca55], independently, at odd primes. However, it was Cartan who first defined the Steenrod algebra A_p and determined its basis of admissible monomials.

In the paper [Se53b], Serre also formulated the modern viewpoint on cohomology operations. A cohomology operation ϕ of degree i is a natural transformation $k^q \to \ell^{q+i}$ for some fixed q, where k^* and ℓ^* are any cohomology theories. When k^* is ordinary cohomology with coefficients in π and ℓ^* is ordinary cohomology with coefficients in ρ, ϕ is determined by naturality by the element $\phi(\iota_q) \in H^{q+i}(K(\pi, q); \rho)$. Observe that, by (1.1), this element may be viewed as a homotopy class of maps $K(\pi, q) \to K(\rho, q+i)$.

A crucial point quickly understood was the calculation of the Steenrod operations in the cohomologies of Lie groups and their classifying spaces and homogeneous spaces. In particular, already in 1950 [Wu50a, Wu53], Wu proved his basic formula for the calculation of the Steenrod operations on the Stiefel–Whitney classes:

$$Sq^r(w_s) = \sum_t \binom{s-r+t-1}{t} w_{r-t} w_{s+t} \quad \text{for } s > r \geqslant 0. \tag{5.2}$$

Borel and Serre made a systematic study shortly afterwards [BS51, BS53].

Also in 1950 [Wu50b], Wu proved his formula giving an algorithm for the calculation of the Stiefel–Whitney classes of the tangent bundle of a manifold directly in terms of its cup products; see Section 12. Wu was a close collaborator of Thom, and his work was dependent on work of Thom, announced in part in 1950 [Thom50a, Thom50b] and published in 1952 [Thom52]. In that paper, Thom proved the Thom isomorphism theorem and used it to give the now familiar description of Stiefel–Whitney classes in terms of Steenrod operations. Since [Thom52] was later overshadowed by Thom's great work on cobordism, it is well worth describing some of its original contributions.

Thom considered locally trivial fiber bundles $p: E \to B$ with fiber S^{k-1}, with no assumptions about the group of the bundle. Working sheaf theoretically and resolutely avoiding the use of spectral sequences, which were available to him, Thom proved the Thom isomorphism

$$\phi: H^q(B) \to H^{q+k}(Mp, E), \tag{5.3}$$

where Mp is the mapping cylinder of p. He worked with twisted integer coefficients, thus allowing for nonoriented fibrations, before studying the mod 2 case. Observe that, in the motivating example of the unit sphere bundle $E = S(E')$ of a k-dimensional vector bundle $p': E' \to B$ with a Riemannian metric, the quotient space Mp/E is homeomorphic to the quotient space $D(E')/S(E')$, where $D(E')$ is the unit disk bundle of E'. This quotient space is called the Thom space of p' and now usually denoted Tp' or $T(E')$.

Using mod 2 coefficients in the Thom isomorphism, Thom defined the Stiefel–Whitney classes of E by

$$w_i = \phi^{-1} Sq^i \phi(1), \tag{5.4}$$

and he proved that, in the case of vector bundles, these are the classical Stiefel–Whitney classes of E. He rederived the properties of Stiefel–Whitney classes, in particular the Whitney duality theorem, from the new definition. This gave an elegant new proof of Whitney's result [Whit41] that the Stiefel–Whitney classes of the normal bundle of an immersion f are invariants of the induced map f^* on mod 2 cohomology. In particular, they are independent of the choice of the differentiable structures on the manifolds in question. It is worth emphasizing that Whitney's foundational work in [Whit41] and other papers, for example on embeddings and immersions of smooth manifolds, was an essential prerequisite to virtually all of the later applications of algebraic topology to geometric topology.

Thom then generalized to obtain results of this form for purely topological immersions, with no hypothesis of differentiability. It should be remembered that this paper appeared four years before Milnor's discovery of exotic differential structures on spheres [Mil56a]. For an embedding f, he went further and showed that the homotopy type of a tubular neighborhood of f is independent of the differentiable structure on the ambient manifold. He then introduced the notion of fiber homotopy equivalence and proved that the fiber homotopy type of the tangent bundle of a manifold is independent of its differentiable structure. He observed that the Stiefel–Whitney classes are invariant under fiber homotopy equivalence, and asked what other such classes there might be. The determination of all characteristic classes for spherical fibrations evolved over the following two decades. That is a long story, intertwined with the theory of iterated loop spaces, and is well beyond our present scope.

6. The introduction of cobordism

In the last chapter of [Thom52], Thom set up the modern theory of Poincaré duality for manifolds with boundary and explained the now familiar necessary Euler characteristic and index conditions for a differentiable manifold to be the boundary of a compact differentiable manifold. The emphasis he placed on the index was a precursor of things to come. He also recalled Pontryagin's fundamental observation [Pon42, Pon47] that, for M to be such a boundary, it is necessary that all of its characteristic numbers be zero. He went on to observe that the vanishing of Stiefel–Whitney numbers is still a necessary condition when M is not assumed to be differentiable. He observed that "la recherche de conditions suffisantes est un problème beaucoup plus difficile".

Two years later, as announced in [Thom53a, Thom53b, Thom53c] and published in his wonderful 1954 paper [Thom54], he had solved this problem for smooth compact manifolds. The importance to modern topology, both geometric and algebraic, of his introduction and calculation of cobordism cannot be exaggerated. For example, Milnor's construction of exotic differentiable structures on S^7 begins with Thom's theory and in particular with Thom's result that every smooth compact 7-manifold is a boundary.

Cobordism theory was not wholly unprecedented. In 1950, Pontryagin [Pon50] showed that the stable homotopy groups of spheres, in low dimension at least, are isomorphic to the framed cobordism groups of smooth manifolds. His motivation was to obtain methods for the computation of stable homotopy groups, and he used this technique to prove that $\pi_{n+2}(S^n) \cong \mathbb{Z}/2\mathbb{Z}$, thus correcting an earlier mistake of his. While that motivation seems misguided in retrospect, it was an imaginative attack on the problem. Pontryagin's paper was in Russian, never translated, and it is not quoted by Thom. However, Thom does quote earlier papers of Pontryagin [Pon42, Pon47] in which the idea of pulling back the zero-section in Grassmannians along a smooth approximation to a classifying map plays a prominent role.

Thom's paper [Thom54] reads a little surprisingly today. Its main focus is not cobordism, which does not appear until the last chapter, but rather the realization of homology classes of manifolds by submanifolds. It seems that it was this that first motivated Thom to a detailed analysis of the cohomology and homotopy of Thom complexes, not just in the stable range relevant to cobordism but also in the unstable range. Moreover, the existence of a stable range for the homotopy groups of $TSO(k)$ and $TO(k)$ is proven by direct methods of algebraic topology, rather than as a consequence of the isomorphism between homotopy groups and cobordism groups.

For a closed subgroup G of $O(k)$, Thom lets $T(G)$ be the Thom space of the universal bundle $E_G \to B_G$ with fiber S^{k-1}. He considers a compact oriented manifold V^n and asks when a homology class $x \in H_{n-k}(V)$ is realizable as the image of the fundamental class of submanifold W^{n-k} of codimension k. He dualizes the question as follows. For any space X, say that a class $y \in H^k(X)$ is G-realizable if there is a map $f: X \to T(G)$ such that $f^*(\mu) = y$, where $\mu \in H^k(T(G))$ is the Thom class. Let $y \in H^k(V)$ be Poincaré dual to x. Then "le théorème fondamental" asserts that x is realizable by a submanifold W such that the structure group of the normal bundle of W in V can be reduced to G if and only if y is G-realizable. Of course, the analogue with mod 2 coefficients does not need orientability. As we shall see in Section 16, Atiyah explained this result conceptually almost a decade later.

Taking G to be the trivial group, it follows from a result of Serre [Se53a] that x is realizable if k is odd or if $n < 2k$ and that Nx is realizable for some integer N that depends only on k and n. However, the main focus is on $G = O(k)$ and $G = SO(k)$. Here Thom shows directly that $\pi_{k+i}(TO(k))$ is independent of k when $i < k$, and similarly for $TSO(k)$. Moreover, crucially, he proves that $TO(k)$ has the same $2k$-type as a precisely specified product of Eilenberg–MacLane spaces $K(\mathbb{Z}_2, k+i)$. The Wu formula (5.2) is the key to the calculation. He goes on to study $H^*(TO(k); \mathbb{Z}_2)$ in low dimensions beyond the stable range for $k \leqslant 3$. For the realizability problem, he deduces that $x \in H_i(V^n; \mathbb{Z}_2)$ is realizable for $i < [n/2]$, with further information in low codegrees $n - i$.

The problem for $TSO(k)$ is much harder, and $\pi_{k+i}(TSO(k))$ is only determined completely for $i \leqslant 7$; more detailed information is obtained for $k \leqslant 4$. However, Thom shows that $TSO(k)$ has the rational cohomology type of an explicitly specified product of Eilenberg–MacLane spaces $K(\mathbb{Z}, k+i)$. For the realizability problem, he deduces that some integer multiple of any $x \in H_i(V^n; \mathbb{Z})$ is realizable, and that any x is realizable if $i \leqslant 5$ or $n \leqslant 8$.

Before turning to cobordism, Thom studies the problem posed by Steenrod of determining which homology classes $x \in H_r(K)$ of a finite polyhedron K are realizable as $f_*(z)$, where z is the fundamental class of a compact manifold M^r and $f : M^r \to K$ is a map. By embedding K as a retract of a manifold with boundary M and taking the double V of M to obtain a manifold without boundary, Thom reduces this question to the realizability question already studied. He thereby proves that, in mod 2 homology, every class x is realizable. In retrospect, of course, this presages unoriented bordism and its relationship to ordinary mod 2 homology. Similarly, he proves that, in integral homology, some integer multiple of every class x is realizable. Remarkably, he then proves that every class x is realizable if $r \leqslant 6$, but that there are unrealizable classes in all dimensions $r \geqslant 7$.

Only after all of this does he prove the cobordism theorems. Let $\mathcal{N}_n$ be the set of cobordism classes of smooth compact n-manifolds, where two n-manifolds are cobordant if their disjoint union is the boundary of a smooth compact $(n+1)$-manifold with boundary. Define Ω_n similarly for oriented n-manifolds. Under disjoint union, $\mathcal{N}_n$ is a $\mathbb{Z}_2$-vector space and Ω_n is an Abelian group; any boundary is the zero element. Under Cartesian product, $\mathcal{N}_*$ and Ω_* are graded rings. Moreover, the index defines a ring homomorphism $I : \Omega_* \to \mathbb{Z}$. The fundamental geometric theorem is the Thom isomorphism: $\mathcal{N}_n$ is isomorphic to the stable homotopy group $\pi_{k+n}(TO(k))$ and Ω_n is isomorphic to the stable homotopy group $\pi_{k+n}(TSO(k))$.

While modern proofs are easier reading than Thom's, the basic ideas are the same. In slightly modernized terms, an isomorphism $\phi : \mathcal{N}_n \to \pi_{k+n}(TO(k))$ is constructed as follows. Embed a given n-manifold M in $\mathbb{R}^{k+n}$ for k large, let ν be the normal bundle of the embedding, and construct a tubular neighborhood V of M in $\mathbb{R}^{k+n}$. Define a map f from S^{k+n} to the Thom space $T(\nu)$ by identifying V with the total space of ν and mapping points not in V to the basepoint. This is the Pontryagin–Thom construction. Classify ν and compose f with the induced map of Thom spaces $T(\nu) \to TO(k)$ to obtain $\phi(M)$, checking that the homotopy class of the composite is independent of the choice of M in its cobordism class and of the embedding. To construct an inverse isomorphism ψ to ϕ, view the classifying space $BO(k)$ as a Grassmannian manifold of sufficiently high dimension. Up to homotopy, any map $g : S^{k+n} \to TO(k)$ can be smoothly approximated by a map that is transverse to the zero-section. Define $\psi(g)$ to be the cobordism class of the inverse image of the zero section, checking that this class is independent of the homotopy class of g. Transversality is the crux of the proof, and Thom was the first to develop this notion.

From here, the earlier calculations in the paper immediately identify the groups $\mathcal{N}_n$. Using this identification, Thom proves that two manifolds are cobordant if and only if they have the same Stiefel–Whitney numbers. By calculating the Stiefel–Whitney numbers of products, this allows him to determine the ring structure of $\mathcal{N}_*$: it is a polynomial algebra on one generator of dimension n for each $n \geqslant 2$ not of the form $2^j - 1$. The even dimensional generators can be chosen to be the cobordism classes of the real projective spaces $\mathbb{RP}^{2n}$.

Similarly, the groups Ω_n are identified modulo torsion by the earlier calculations. Using this, Thom proves that if all Pontryagin numbers of an oriented manifold are zero, then the disjoint union of some number of copies of that manifold is a boundary. This allows determination of the ring $\Omega_* \otimes \mathbb{Q}$: it is a polynomial algebra on generators of dimension $4n$ for $n \geqslant 1$. The generators can be chosen to be the cobordism classes of the complex projective spaces $\mathbb{CP}^{2n}$.

Dold [Dold56a] soon after identified odd-dimensional generators of $\mathcal{N}_*$. The Wu formula for the computation of Stiefel–Whitney classes of manifolds give restrictions on which collections of Stiefel–Whitney numbers actually correspond to the cobordism class of a manifold, and Dold [Dold56b] proved that these relations are complete: a collection of Stiefel–Whitney numbers that satisfies the Wu relations corresponds to a manifold. In modern invariant terms, the Stiefel–Whitney numbers of manifolds define a monomorphism $\mathcal{N}_* \to \mathrm{Hom}(H^*(BO; \mathbb{Z}_2), \mathbb{Z}_2)$, and its image consists of those homomorphisms that annihilate the subgroup generated by the Wu relations.

7. The route from cobordism towards K-theory

Hirzebruch [Hirz53] had already introduced multiplicative sequences of characteristic classes before Thom's paper. However, cobordism theory provided exactly the right framework for their study and allowed him to prove the index theorem [Hirz56]: the index of a smooth oriented $4n$-manifold M is the characteristic number $\langle L(\tau), [M]\rangle$, where L is the L-genus and τ is the tangent bundle of M. Here $L(\tau)$ is a polynomial in the Pontryagin classes of M determined in Hirzebruch's formalism by the power series $L(x) = x/\tanh(x)$. Using Thom's observation that the index defines a ring homomorphism $\Omega_* \to \mathbb{Z}$, Hirzebruch's formalism shows that the index formula must hold for some power series L, and $L(x)$ is the only power series that gives the correct answer on complex projective spaces.

The purpose of Hirzebruch's monograph [Hirz56] was to prove the Riemann–Roch theorem for algebraic varieties of arbitrary dimension. It would take us too far afield to say much about this, and a quite detailed summary may be found in Dieudonné [Dieu, pp. 580–595]. Suffice it to say that Hirzebruch's essential strategy was to reduce the Riemann–Roch theorem to the index theorem. One key ingredient in the reduction should be mentioned, namely a method for splitting vector bundles that led later to the splitting principle in K-theory.

Another nice discussion of [Hirz56] may be found in Bott's review [Bott61] of the second part of Borel and Hirzebruch's deeply influential work [BH58, BH59, BH60]. The Riemann–Roch theorem showed that the characteristic number $\langle T(\tau_c), [M]\rangle$ of any projective nonsingular variety M is an integer, namely the arithmetic genus of M; here τ_c is the complex tangent bundle of M and T is the Todd genus, which is determined by the power series $T(x) = x/1 - \mathrm{e}^{-x}$. Borel and Hirzebruch sought and proved an analogous integrality theorem for arbitrary differentiable manifolds. The $\hat{A}$-genus is related to the Todd genus by the formula $T(x) = \mathrm{e}^{x/2}\hat{A}(x)$, and it satisfies $\hat{A}(x) = \hat{A}(-x)$. As Bott explains clearly, this makes it plausible that the $\hat{A}$-genus should satisfy a similar integrality relation on arbitrary compact manifolds, as Borel and Hirzebruch prove. More precisely, they prove it up to a factor of 2 that was later eliminated by Milnor's proof (implicit in [Mil60]) that the Todd genus of an almost complex manifold is an integer.

Milnor and Kervaire [Mil58b, KM60] gave an important application of the integrality of the $\hat{A}$-genus. In 1942 [Wh42], G.W. Whitehead introduced the stable J-homomorphism

$$J : \pi_q\big(SO(n)\big) \to \pi_{q+n}\big(S^n\big),$$

n large. Writing $\pi_q^s = \pi_{q+n}(S^n)$ for the q-th stable homotopy group of spheres and letting n go to infinity, this can be written $J : \pi_q(SO) \to \pi_q^s$. Milnor and Kervaire used the integrality theorem to prove that, when $q = 4k-1$, the order j_n of the image of J is divisible by the denominator of $B_k/4k$, where B_k is the k-th Bernoulli number. This result gave the first sign of regularity in the stable homotopy groups of spheres, and their proof showed that the J-homomorphism is of considerable relevance to geometric topology. In fact, although this is a result in stable homotopy theory, they derive it from a generalization of a theorem of Rohlin in differential topology. Rohlin's theorem [Ro51, Ro52] states that the Pontrjagin number $p_1(M)$ of a compact oriented smooth 4-manifold M with $w_2(M) = 0$ is divisible by 48. Milnor and Kervaire mimic his arguments to prove that the Pontrjagin number $p_n(M)$ of an almost parallelizable smooth $4n$-manifold is divisible by $(2n-1)!\,j_n a_n$, where a_n is 2 if n is even and 1 if n is odd, with equality for at least one such manifold M.

For the historical story, one striking feature of the work of Borel and Hirzebruch is its systematic use of multiplicative functions $F_{\mathbb{C}}(X) \to H^{**}(X;\mathbb{R})$ and $F_{\mathbb{R}}(X) \to H^{**}(X;\mathbb{R})$, where $F_{\mathbb{R}}(X)$ and $F_{\mathbb{C}}(X)$ are the semi-groups of equivalence classes of complex and real vector bundles over X and $H^{**}(X;\mathbb{R})$ is the direct product of the real cohomology groups of X. A multiplicative function is one that converts sums to products. The authors are tantalizingly close to K-theory. Two things are missing: the Grothendieck construction and Bott periodicity.

The first was introduced by Grothendieck [BS58], who needed it to formulate his generalized, relative, version of the Riemann–Roch theorem in algebraic geometry. Grothendieck is the inventor of the general subject of K-theory, and his ideas played a centrally important role in the introduction of topological K-theory.

As to the second, as Bott notes in his review, the work of Borel and Hirzebruch led them to an exact sequence

$$0 \to \mathbb{Z}_{n!} \to \pi_{2n}\big(U(n)\big) \to \pi_{2n}\big(U(n+1)\big) \to 0. \tag{7.1}$$

More precisely, they proved the sequence to be exact modulo 2-torsion. As Bott writes:

> The exact sequence conflicted, at the time of its discovery, with computations of homotopy theorists and led to a spirited controversy. At present it is known the sequence is exact even with regard to the prime 2.

What he neglects to say is that the sequence also follows from Bott periodicity, and the conflict for some time held up publication of that result.

8. Bott periodicity and K-theory

One version of the Bott periodicity theorem asserts that there is a homotopy equivalence $BU \to \Omega SU$. The periodicity is clearer in the equivalent reformulation $BU \times \mathbb{Z} \simeq \Omega^2(BU \times \mathbb{Z})$. The real analogue gives $BO \times \mathbb{Z} \simeq \Omega^8(BO \times \mathbb{Z})$. Bott's original proof of these beautiful results is based on the use of Morse theory. Before proving the periodicity theorem, Bott had clearly demonstrated the power of Morse theory by using it to prove that there is no torsion in the integral homology of ΩG for any simply connected compact Lie group G [Bott56]. Bott announced the periodicity theorem in [Bott57], and he gave two somewhat different proofs, both based on Morse theory, in [Bott58, Bott59a].

It immediately became a challenge to reprove the periodicity theorems using the standard methods of algebraic topology. In the complex case, a proof was given by Toda [To62b], together with a rederivation of the Borel–Hirzebruch exact sequence (7.1), but his proof did not show that BU and ΩSU have the same homotopy type. The space BU is an H-space under Whitney sum, and Bott's proofs led to simple and explicit H-maps that give the equivalences. In the real case, there are actually six maps that must be proven to be equivalences. These explicit maps were exploited by Dyer and Lashof [DL61] and Moore (written up by Cartan [Ca54-55]) to give direct calculational proofs. Actually, there is a curious simplification to be made: comparison of the proofs in [DL61] and [Ca54-55] shows that each finds particular difficulty in proving one of the required equivalences, but they find difficulty with different maps: combining the best of both proofs gives a quite tractable argument.

Finally, in their announcement [AH59], submitted in May, 1959, Atiyah and Hirzebruch introduce the functor $K(X)$ for a finite CW complex X: it is the Grothendieck construction on the semi-group $F_{\mathbb{C}}(X)$, and it is a ring with multiplication induced by the tensor product of vector bundles. They define $KO(X)$ similarly. They noticed a striking reinterpretation of Bott periodicity: tensor product of bundles induces a natural isomorphism β that fits into the commutative diagram

$$\begin{array}{ccc} K(X)\otimes K(S^0) & \xrightarrow{\ \beta\ } & K(X\times S^2) \\ \downarrow{\scriptstyle ch} & & \downarrow{\scriptstyle ch} \\ H^{**}(X;\mathbb{Q})\otimes H^{**}(S^2;\mathbb{Q}) & \xrightarrow[\alpha]{} & H^{**}(X\times S^2;\mathbb{Q}), \end{array}$$

where ch is the Chern character and α is the cup product isomorphism.

They observe that, for connected X, the kernel $\widetilde{K}(X)$ of the dimension map $\varepsilon : K(X) \to \mathbb{Z}$ can be identified with the set of homotopy classes of maps $X \to BU$. In principle, modulo a $\lim^1$ argument not yet available, this leads to a homotopy equivalence from BU to the basepoint component of $\Omega^2 BU$. However, their reinterpretation of Bott periodicity was by no means an obvious one. In [Bott58], Bott related his explicit maps to tensor products of bundles and so proved that his original version of the periodicity theorem really did imply the version noticed by Atiyah and Hirzebruch. Moreover, he gave the analogous reinterpretation in the real case, where a direct proof of the new version was less simple.

Jumping ahead to 1963 for a moment, Atiyah and Bott together [AB64] then found a direct and elementary analytic proof of the complex case of the periodicity isomorphism in its tensor product formulation, using clutching functions to describe bundles over $X \times S^2$ explicitly. Their proof actually gives a more general result, namely a Thom isomorphism, and important refinements and generalizations are given in their lecture notes [At64, Bott63]. The analytic proof is relevant to the Atiyah–Singer index theorem, which was already announced in 1963 [AS63] and which generalizes Hirzebruch's index theorem. The first published proof appeared in 1965 [Pa65], based on seminars in 1963/64.

In their 1959 announcement [AH59] and also in [Hirz59], Atiyah and Hirzebruch give a Riemann–Roch theorem relative to a suitable map $f : M \to N$ of differential manifolds; see Section 12 for the statement. They observe that their theorem can be rewritten for holomorphic maps between complex manifolds in the same form as Grothendieck's version of the Riemann–Roch theorem. Their results imply a new proof of the integrality of the $\hat{A}$-genus, together with a sharpening in the case of $Spin$-manifolds of dimension congruent to $4 \mod 8$ that had been conjectured by Borel and Hirzebruch. They also rederive and give a conceptual sharpening of Milnor's result on the J-homomorphism.

In [AH59], nothing is said about $K(X)$ being part of a generalized cohomology theory. Moreover, it is clear that the authors as yet have no hint of K-homology and Poincaré duality: their statement of the Riemann–Roch theorem involves a pushforward map $f_!$, as it must, but that map was not well understood. They remark that "It is probable that $f_!$ is actually induced by a functorial homomorphism $K(Y) \to K(X)$".

Rather than proceed directly to 1960 and the first published account of K-theory as a generalized cohomology theory, I shall interpolate a discussion of several quite different lines of work that were going on in the late 1950's.

As preamble, Milnor [BM58, Mil58c] saw immediately, in February 1958, that Bott's results led to the solution of two longstanding problems; [BM58] is a pair of letters between Milnor and Bott on this subject, and [Mil58c] fills in the details. The relevant result of Bott is that the image of the Hurewicz homomorphism $\pi_{2n}(BU) \to H_{2n}(BU)$ is divisible by exactly $(n-1)!$. This is closely related to the exact sequence (7.1). What Milnor deduces from this is:

(1) The vector space $\mathbb{R}^n$ possesses a bilinear product without zero divisors only for n equal to $1, 2, 4$, or 8.

(2) The sphere S^{n-1} is parallelizable only for $n-1$ equal to $1, 3$, or 7.

The latter result was also proven at about the same time by Kervaire [Ker58].

9. The Adams spectral sequence and Hopf invariant one

Milnor's results just cited are also among the many implications of Adams' celebrated theorem that $\pi_{2n-1}(S^n)$ contains an element of Hopf invariant one if and only if n is 1, 2, 4, or 8 [Ad60]. This result was announced in [Ad58b], which was submitted in April 1958. This work was a sequel to and completion of work begun in [Ad58a], submitted in June 1957, in which Adams first attacked the Hopf invariant one problem and introduced the Adams spectral sequence.

Fix a prime p, let A be the mod p Steenrod algebra, and let X be a space. In its original form in [Ad58a], the Adams spectral sequence satisfies

$$E_2^{s,t} = \mathrm{Ext}_A^{s,t}\big(H^*(X), \mathbb{Z}_p\big),$$

where s is the homological degree, t is the internal degree, and $t - s$ is the total degree, so that $E_2^{s,t} = 0$ if $s < 0$ or $t < s$. The differentials are of the form

$$d_r : E_r^{s,t} \to E_r^{s+r,t+r-1}.$$

There is a filtration of the stable homotopy groups $\pi_n^s(X)$ such that

$$E_\infty^{s,n+s} = F^s\pi_n^s(X)/F^{s+1,n+s+1}\pi_n^s(X).$$

The intersection of the filtrations consists of the elements of $\pi_n^s(X)$ that are of finite order prime to p. When $X = S^0$, $\{E_r^{*,*}\}$ is a spectral sequence of differential $\mathbb{Z}_p$-algebras and converges as an algebra to the associated graded algebra of the ring of stable homotopy groups of spheres under the composition product.

The Adams spectral sequence can be thought of in several ways: it is a sophisticated reformulation and generalization of the Cartan–Serre method of killing homotopy groups, and it is an extension and systematization of the method of studying homotopy groups by considering higher order cohomology operations.

The idea of higher order operations first appeared with Steenrod's introduction of functional cohomology operations [St49]. Let $f : Y \to X$ be a map. Steenrod showed how to construct an element $x \cup_f x'$ in $H^*(Y)$ from a pair of elements x, x' in $H^*(X)$ such that $x \cup x' = 0$ and $f^*(x') = 0$. He defined functional mod 2 Steenrod operations similarly. These operations are defined on a subspace of $H^*(X)$, and they are well-defined up to indeterminacy. Adem [Adem56] made a systematic study of functional cohomology operations associated to stable cohomology operations, and Peterson [Pe57] gave a presentation in terms of Postnikov systems with stable k-invariants. Although a few low dimensional examples had appeared earlier, Adem [Adem58] gave the first systematic study of secondary cohomology operations, building on his earlier proof of the Adem relations for the iterated Steenrod operations. He related secondary and functional cohomology operations in [Adem59]. Peterson and Stein [PS59] then gave a treatment of secondary and functional operations in terms of two-stage Postnikov systems.

It was this kind of treatment that Adams had in mind. Secondary and higher operations come from relations between relations, and homological algebra is the natural tool for the study of relations between relations. The essential idea of the construction of the Adams spectral sequence is to construct a realization of a free resolution of the A-module $H^*(X)$ (in a range of dimensions) by means of a resolution of the space X. This gives a kind of exact couple of spaces that leads to an exact couple giving the desired spectral sequence on passage to homotopy groups. Implicitly, as became much clearer with a later reformulation in terms of the homology of spectra rather than the cohomology of spaces, the fundamental points are the representation (1.1) of cohomology and the calculation of the cohomology of Eilenberg–MacLane spaces in terms of Steenrod operations.

The relationship to the Hopf invariant one problem comes about as follows. There is an element of Hopf invariant one in $\pi_{2n-1}(S^n)$ if and only if there is a (stable) two-cell complex such that the Steenrod operation Sq^n connects the bottom cell to the top cell in mod 2 cohomology. If n is not a power of two, then Sq^n is decomposable as a linear combination of iterated Steenrod operations, by the Adem relations, and no such two-cell complex is possible. Now, for any connected $\mathbb{Z}_p$-algebra A, $\mathrm{Ext}_A^{1,t}(\mathbb{Z}_p, \mathbb{Z}_p)$ is isomorphic to the dual of the vector space of degree t indecomposable elements of A. Take A to be the mod 2 Steenrod algebra and consider the Adams spectral sequence for $X = S^0$. Then we have elements $h_i \in E_2^{1,2^i}$ dual to the Steenrod operations Sq^{2^i}. It is direct from the construction of the spectral sequence that there is an element of Hopf invariant one detected by Sq^{2^i} if and only if h_i is a permanent cycle in the spectral sequence.

The element h_0 corresponds to the Bockstein $Sq^1 = \beta$, and multiplication by h_0 in the spectral sequence detects multiplication by 2 in the stable homotopy groups of spheres. Adams computes enough of $E_2^{s,*}$, $s = 2$ and $s = 3$, to see that the elements $h_0 h_i^2$ are nonzero in E_2 for $i \geqslant 3$. The only way that $h_0 h_i^2$ can be a boundary is if $d_2(h_{i+1}) = h_0 h_i^2$. If $i \geqslant 3$ and both h_i and h_{i+1} are permanent cycles, we conclude that h_i represents an odd dimensional homotopy class x_i such that $2x_i^2$ is nonzero. This is impossible since π_*^s is a graded commutative ring. This implies the main theorem of [Ad58a]: if both $\pi_{2n-1}(S^n)$ and $\pi_{4n-1}(S^{2n})$ contain elements of Hopf invariant one, then $n \leqslant 4$, which was tantalizingly close to the expected answer.

This line of argument does not work to solve the problem. However, the method of proof implies that Sq^{16}, although indecomposable in A, admits a decomposition in terms of composites of primary and secondary operations, taking into account the relevant domains of definition and indeterminacy. In [Ad60], Adams constructs such a decomposition of Sq^{2^i} for all $i \geqslant 4$. While the argument makes no use of the Adams spectral sequence, it implies the differential $d_2(h_{i+1}) = h_0 h_i^2$ for $i \geqslant 3$.

The arguments in [Ad60] are very long, and I won't attempt a complete summary. They require a more thorough exposition of the foundations of graded homological algebra than was needed in [Ad58a], and this work has been used ever since. They also require an axiomatization and construction of secondary cohomology operations in terms of universal examples, together with a detailed study of how to relate the homological algebra to the analysis of the operations. Finally, particular operations relevant to the problem at hand are constructed, a putative decomposition formula for Sq^{2^n} is proven formally by means of the general theory, and the coefficient of Sq^{2^n} in the decomposition is proven to be nonzero by explicit calculation in a specific example.

There are two crucially important ingredients in the work that must be singled out. First, the work of Milnor and Moore [MM65] on graded Hopf algebras plays a key role in the relevant homological algebra. Although [MM65] was not published until 1965, a mimeographed version was distributed much earlier and was an essential prerequisite to the higher level of algebraic sophistication that Adams introduced into algebraic topology.

Second, Adams needed to make some calculations of E_2 beyond those of [Ad58a], and for this purpose he made substantial use of Milnor's remarkable analysis of the structure of the Steenrod algebra [Mil58a]. This analysis has played a central role in a great many later calculations in stable algebraic topology. The Steenrod algebra A is a Hopf algebra. Its coproduct is determined by the Cartan formula and is cocommutative. Therefore the dual Hopf algebra, denoted A_*, is commutative as an algebra. Milnor proved that it is a free commutative algebra in the graded sense. Explicitly, for an odd prime p, it can be written as a tensor product

$$A_* = E\{\tau_i \mid i \geqslant 0\} \otimes P\{\xi_i \mid i \geqslant 1\} \tag{9.1}$$

of an exterior algebra on odd degree generators τ_i and a polynomial algebra on even degree generators ξ_i. Moreover, the coproduct on the generators admits a simple explicit formula, in principle equivalent to the Adem relations but far more algebraically tractable. The dual B of $P\{\xi_i \mid i \geqslant 1\}$ can be identified both with the subalgebra of A generated by the Steenrod operations P^i and with the quotient of A by the two-sided ideal generated by the Bockstein β. Note that, in quotient form, B also makes sense when $p = 2$. We shall come back to it later.

Shortly after Adams' work, the techniques he developed were adapted to solve the analogue of the Hopf invariant one problem at odd primes p, showing that there can be a two-cell complex with P^n connecting the bottom cell to the top cell in mod p cohomology if and only if $n = 1$. This work was done independently by Liulevicius [Liu62a] and by Shimada and Yamanoshita [SY61].

Using the structure theory for mod p Hopf algebras of Milnor and Moore and Milnor's analysis of the Steenrod algebra, I later developed tools in homological algebra that allowed the use of the Adams spectral sequence for explicit computation of the stable homotopy groups of spheres in a range of dimensions considerably greater than had been known previously [May65a, May65b, May66]. Correspondence initiated in the course of this work led Adams and myself to a long friendship, and I have given a brief account of all of Adams' work in [May2] and a eulogy and personal reminiscences in [May1].

10. S-duality and the introduction of spectra

Setting up the Adams spectral sequence as Adams did it originally is a tedious business, the reason being that one is trying to do stable work with unstable objects: one should be using "spectra" rather than spaces. Similarly, the representability of ordinary cohomology and the introduction of cobordism and K-theory must eventually have forced the introduction of spectra, which appear naturally as sequences of Eilenberg–MacLane spaces, as sequences of Thom spaces, and as sequences of spaces featuring in the Bott periodicity theorem.

Nevertheless, the fact is that the introduction of spectra had nothing whatever to do with these lines of work. Rather, it grew out of the work on S-duality of Spanier and Whitehead. I will be brief about this since it is also treated in [BG] in this volume.

In 1949, Spanier [Sp50] reconsidered Borsuk's cohomotopy groups [Bor36]. For a (nice) compact pair (X, A), where $\dim X < 2n - 1$, Spanier defined $\pi^n(X, A)$ to be the set of homotopy classes of maps $(X, A) \to (S^n, *)$. As in Borsuk [Bor36], these are Abelian groups, and Spanier showed that these cohomotopy groups satisfy *all* of the Eilenberg–Steenrod axioms for a cohomology theory, except that they are only defined in a range of non-negative degrees depending on the dimension of X. He also showed that the cohomotopy groups map naturally to the integral Čech cohomology groups and that, for a CW complex X with subcomplex A, $\pi^n(X^m \cup A, X^{m-1} \cup A)$ is isomorphic to the cellular cochain group $C^m(X, A; \pi_m(S^n))$. These were puzzling results. The real explanation, that these cohomotopy groups are the terms in a positive range of dimensions of a cohomology theory whose coefficients are nonzero in negative dimensions, would come later. With hindsight, the cellular cochain isomorphism just mentioned is the first hint of the Atiyah–Hirzebruch spectral sequence for stable cohomotopy theory. Spanier also observed that the Hurewicz isomorphism theorem for $[S^n, X]$ and the Hopf classification theorem for $[X, S^n]$ are dual to one another.

To make a home for such duality phenomena in all dimensions, Spanier and Whitehead devised the S-category in [SW53, SW57]. Its objects are based spaces, and the set $\{X, Y\}$ of S-maps $X \to Y$ is

$$\{X, Y\} = \operatorname{colim}_{n \geqslant 0} \left[\Sigma^n X, \Sigma^n Y\right].$$

That is, homotopy classes of based maps $f : \Sigma^n X \to \Sigma^n Y$ and $g : \Sigma^q X \to \Sigma^q Y$ define the same S-map if $\Sigma^k f$ and $\Sigma^{n-q+k} g$ are homotopic for some $k \geqslant 0$. The S-category is additive, and $\Sigma : \{X, Y\} \to \{\Sigma X, \Sigma Y\}$ is a bijection.

Although obscured by their language of "carriers", in retrospect a most unfortunate choice of technical details, Spanier and Whitehead introduce graded morphisms by setting $\{X, Y\}_q = \{\Sigma^q X, Y\}$ if $q \geqslant 0$ and $\{X, \Sigma^{-q} Y\}$ if $q < 0$. They prove that, for CW complexes X and Y with X finite, the Abelian groups $\{X, Y\}_q$ satisfy all except the dimension axiom of the Eilenberg–Steenrod axioms for a homology theory in Y when X is fixed and for a cohomology theory in X when Y is fixed. They even set up the Atiyah–Hirzebruch spectral sequences for stable homotopy and stable cohomotopy.

However, they do not take the step of describing their results in a language of homology and cohomology theories, and none of their later papers return to this point of view. With their definitions, the wedge axiom would not be satisfied in cohomology for infinite X, and only homology and cohomology theories represented by suspension spectra of spaces would be obtained. Thus this would not have been the right way to set up generalized homology and cohomology theories, and that was far from their intention. The useful version of the Spanier–Whitehead category is its full subcategory of *finite* CW complexes. This category is far too small to form a satisfactory foundation for stable homotopy theory, but it is appropriate for the study of duality between finite CW complexes, which is the main point of the papers [SW55, SW58] and the expository notes [Whi56, Sp56, Sp58].

The 1956 note [Sp56] of Spanier, reviewed by Hilton, gives a nice description of dual theorems in algebraic topology and seems to have been a forerunner of Eckmann–Hilton duality. The 1956 survey of Whitehead [Whi56] looks more towards the past, based as it was on Whitehead's presidential address to the London Mathematical Society. Prior to this point, it had been common practice to discuss duality in ordinary homology and cohomology in terms of Pontryagin duality of groups. Whitehead gives an interesting exposition of this point of view on duality, the role of colimits in understanding singular homology and Čech cohomology, and various other aspects of duality theory in algebraic topology. At that stage in our story, it is not very surprising that Whitehead understands the Eilenberg–Steenrod axioms solely in terms of ordinary homology and cohomology theories.

In retrospect, it is more surprising that Spanier in his 1959 paper [Sp59b] still understands the axioms this way. In a footnote, he refers to the Eilenberg–Steenrod axioms to specify what he means by homology and cohomology, and of course he means all of the axioms. There is no hint of generalized homology and cohomology theories in the paper, although one of its main points is the convenience and importance of spectra in the study of duality theory. Nevertheless, the work of Spanier and Whitehead, especially the work in [Sp59b], was soon to lead to duality theorems in generalized homology and cohomology.

Before saying more about [Sp59b], I should mention the interesting paper [Sp59a] that Spanier wrote a year earlier. In it, he returns to the Dold–Thom description [DT58] of integral homology as the homotopy groups of the infinite symmetric product, and he shows how this can be related to the S-category and Spanier–Whitehead duality. Function spaces are used heavily in the comparison, and it seems that their use may have led to the idea of spectra.

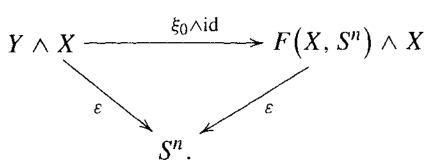
In any case, Spanier's student Lima introduced spectra in his 1958 thesis, published in [Lima59]. In Lima's work, a spectrum is a sequence of based finite CW complexes L_i and S-maps $\lambda_i : \Sigma L_i \to L_{i+1}$. Lima also considers inverse spectra, with structure maps reversed. He uses spectra to give an extension of the S-category and an extension of Spanier–Whitehead duality from polyhedra embedded in spheres to general compact subspaces of spheres. In a sequel, Lima [Lima60] develops Postnikov systems in his category of spectra. He also gives a curious dual theory whose dual Postnikov invariants lie in homology groups with coefficients in cohomotopy groups.

In Spanier's paper [Sp59a], he redefines spectra X to be sequences of based spaces T_i and based maps, not S-maps, $\sigma_i : \Sigma T_i \to T_{i+1}$ that satisfy certain connectivity and convergence conditions. These conditions have the effect of giving his spectra a stable range analogous to the one implied for the suspension spectrum $\{\Sigma^i X\}$ of a based space X by the generalized Freudenthal suspension theorem, which was first proven in [SW57]. His intent is to recast Spanier–Whitehead duality in terms of smash products $X \wedge Y$ and function spectra $\mathbb{F}(X, Y)$, where X and Y are based spaces and $\mathbb{F}(X, Y)$ has i-th space the function space $F(X, \Sigma^i Y)$. Curiously, he does not define general function spectra $\mathbb{F}(X, T)$. He writes $\mathbb{F}(X)$ for $\mathbb{F}(X, S^0)$ and calls it the functional dual of X, and he observes that $H^{-q}(X) \cong H_q(\mathbb{F}(X))$. He defines stable maps $\{X, T\}$ from a space to a spectrum and shows that there are canonical duality isomorphisms

$$\{X, \mathbb{F}(Y, S^n)\} \cong \{X \wedge Y, S^n\} \cong \{Y, \mathbb{F}(X, S^n)\}.$$

(Actually, his statement of this has $\mathbb{F}(-, S^n)$ replaced with the n-fold suspension of the functional dual, but his definition of suspension disagrees with the modern one.) While the asymmetry between spaces and spectra is clearly unsatisfactory, this was a step from the S-category towards the true stable homotopy category.

He then redefines what it means for spaces X and Y to be n-dual to one another. Let $i_n \in \widetilde{H}^n(S^n)$ be the fundamental class. A map $\varepsilon : Y \wedge X \to S^n$ is said to be an *n-duality map* if the homomorphism $f_\varepsilon : \widetilde{H}_q(Y) \to \widetilde{H}^{n-q}(X)$ defined by $f_\varepsilon(y) = \varepsilon^*(i_n)/y$ is an isomorphism, where / is the slant product. He proves that ε determines and is determined by a weak equivalence ξ from the suspension spectrum of Y to $\mathbb{F}(X, S^n)$ such that the following diagram of spaces commutes in the S-category:



This gives an intrinsic characterization of the n-dual of X that leads to all of the properties proven in the earlier work of Spanier and Whitehead [SW55]. The earlier work shows that if X is embedded in S^{n+1} and Y is embedded in the complement of X in such a way that the inclusion $Y \to S^{n+1} - X$ induces an isomorphism of all homology groups, then there is a duality map $\varepsilon : Y \wedge X \to S^n$. This unfortunately means that Spanier's new notion of an n-duality is what in the earlier work was called an $(n+1)$-duality. The new notion relegates the role of the embeddings to the verification of a more conceptual defining property and makes it much simpler to determine when spaces X and Y are n-dual to one another. It is equivalent to the modern homotopical definition of a duality map in the stable homotopy category.

All of this work of Spanier and Whitehead was independent of the work on cobordism, integrality theorems, and K-theory that was going on at the same time. In [MS60], submitted a month after [Sp59a], Milnor and Spanier show that if a smooth compact n-manifold M is embedded in the pair $\mathbb{R}^{n+k}$ with normal bundle ν, then the Thom space $T(\nu)$ is $(n+k)$-dual (new style) to M_+. Moreover, they show that if k is sufficiently large, then ν is fiber homotopy trivial if and only if there is an S-map $S^n \to M$ of degree one. They also make the nice observation that Adams' solution to the Hopf invariant one problem implies that the tangent bundle of a homotopy n-sphere is fiber homotopy trivial if and only if n is 1, 3, or 7.

A year later, in [At61c], Atiyah made a systematic study of the relationship between Thom complexes and S-duality. In particular, he proved the Atiyah duality theorem, which identifies the $(n+k)$-dual of the cofibration sequence $\partial M_+ \to M_+ \to M/\partial M$ of a smooth compact n-manifold M with boundary ∂M as the cofibration sequence

$$T\big(\nu(\partial M)\big) \to T\big(\nu(M)\big) \to T\big(\nu(M)\big)/T\big(\nu(\partial M)\big)$$

associated to the normal bundles of a proper embedding of the pair $(M, \partial M)$ in $(\mathbb{R}^{n+k-1} \times [0, \infty), \mathbb{R}^{n+k-1} \times \{0\})$. He also proved that, for any bundle ξ over a smooth compact manifold M without boundary, the Thom complex $T(\xi)$ is S-dual to the Thom complex $T(\nu \oplus \xi^{\perp})$, where $\xi \oplus \xi^{\perp}$ is trivial. We will return to this paper when we discuss the J-homomorphism.

11. Oriented cobordism and complex cobordism

With the aid of the Adams spectral sequence, the work of Thom on the oriented cobordism ring could be completed. Although slightly ahistorical, the language of spectra will clarify how this came about. Using the structural maps $\sigma : \Sigma T_n \to T_{n+1}$, the homotopy, homology, and cohomology of a spectrum $T = \{T_n\}$ can be defined as follows:

$$\pi_q(T) = \operatorname{colim} \pi_{n+q}(T_n), \tag{11.1}$$

$$H_q(T) = \operatorname{colim} \widetilde{H}_{n+q}(T_n) \tag{11.2}$$

and

$$H^q(T) = \lim \widetilde{H}^{n+q}(T_n), \tag{11.3}$$

where the last definition is only correct when $\lim^1 \widetilde{H}^{n+q-1}(T_n) = 0$. As Adams noted in 1959 [Ad59], the Adams spectral sequence generalizes readily to a spectral sequence for the computation of $\pi_*(T)$ in terms of the mod p cohomology $H^*(T)$, regarded as a module over the Steenrod algebra A. The E_2-term is given by

$$E_2^{s,t} = \operatorname{Ext}_A^{s,t}\big(H^*(T), \mathbb{Z}_p\big),$$

and everything said earlier applies, with simpler proofs, in this more general setting.

For each of the familiar sequences of classical groups $G(n)$, namely $G = O, SO, U, SU$, Sp, and $Spin$, the Thom spaces $TG(n)$ of the universal bundles give a Thom spectrum MG. A uniform method of attack on the problem of computing $\pi_*(MG)$ is to first compute the mod p cohomology of MG for each prime p and then compute the mod p Adams spectral sequence.

A key reason that Thom was able to compute $\mathcal{N}_*$ completely was that the mod 2 cohomology $H^*(MO)$ is a free module over the mod 2 Steenrod algebra A. A quick direct proof of this fact, using Hopf algebra techniques, was given by Liulevicius [Liu62b] in 1962.

For an Abelian group π, the sequence of spaces $K(\pi, n)$ gives a spectrum $H\pi$ such that $\pi_0(H\pi) = \pi$ and the remaining homotopy groups of $H\pi$ are zero. The mod p cohomology of $H\mathbb{Z}_p$ is the mod p Steenrod algebra, as Cartan had implicitly shown [Ca55]. The representation of cohomology (1.1) generalizes to spectra. Representing generators of $H^*(MO)$ as maps from MO to suspensions of $H\mathbb{Z}_2$, one obtains a map from MO to a product of suspensions of $H\mathbb{Z}_2$ that induces an isomorphism on mod 2 cohomology. Since one knows that $\pi_*(MO)$ is a $\mathbb{Z}_2$-vector space, one readily deduces that this map is an equivalence of spectra, allowing one to read off $\pi_*(MO)$. However, a good homotopy category of spectra in which to make such a deduction only appeared later.

Using spectra and the Adams spectral sequence, Milnor [Mil60] in 1959 proved that $\Omega_* = \pi_*(MSO)$ has no odd torsion. This was proven independently by Averbuh [Av59] and, a little later, Novikov [Nov60]. These are announcements. Averbuh's proofs never appeared and Novikov's proofs [Nov62] seem never to have been translated from the Russian.

Also in 1959 [Wall60], but without using spectra or the Adams spectral sequence, Wall determined the 2-torsion in Ω_*. In particular, he proved that Ω_* has no elements of order 4 and that two oriented manifolds are cobordant if and only if they have the same Stiefel–Whitney and Pontryagin numbers. These results were both conjectured by Thom [Thom54]. A nice deduction from the explicit form of the generators Wall found is that the square of any manifold is cobordant to an oriented manifold, and he remarked the desirability of a direct geometric proof; we shall return to this in Sections 16 and 17.

After calculating the 2-torsion in Ω_* by other means, Wall used this calculation to prove that the mod 2 cohomology $H^*(MSO)$ is the direct sum of suspensions of copies of A and of A/ASq^1. He remarks "It seems that a direct proof ... would be extremely difficult", but he found such a direct proof not long afterwards [Wall62]. That allows a more direct calculation of Ω_*. In fact, the mod 2 cohomology of $H\mathbb{Z}$ is A/ASq^1. As Browder, Liulevicius, and Peterson observed later [BLP66], it follows that there is a map f from the spectrum MSO to a product of suspensions of copies of $H\mathbb{Z}$ and $H\mathbb{Z}_2$ that induces an isomorphism on mod 2 cohomology. In a good homotopy category of spectra, one readily deduces that f is a 2-local equivalence. Of course, the foundations for such an argument only came later, but the calculation of homotopy groups is easily made by use of the Adams spectral sequence.

Milnor [Mil60] and Novikov [Nov60, Nov62] also introduced and calculated complex cobordism $\pi_*(MU)$. Although the geometric interpretation was not included in Milnor [Mil60], this is the cobordism theory of weakly almost complex manifolds, namely manifolds with a complex structure on their stable normal bundles. The explicit calculation, carried out one prime at a time and then collated algebraically, showed that $\pi_*(MU)$ is a polynomial ring on one generator of degree $2i$ for each $i \geqslant 1$. Interestingly, there is no known geometric reason why the complex cobordism ring should be concentrated in even degrees. The analogue for symplectic cobordism is false. The cited papers of Milnor and Novikov raise the question of determining $\pi_*(MG)$ for other classical groups G and give some information. We will return to this in Sections 16 and 17.

12. K-theory, cohomology, and characteristic classes

In their 1960 paper [AH61a], Atiyah and Hirzebruch explicitly introduce K-theory as a generalized cohomology theory. Whether or not the idea of taking a generalized cohomology theory seriously occurred to anyone before, this paper is the first published account. They restrict attention to finite CW complexes X for convenience, but they are fully aware of both represented K-theory and inverse limit K-theory, namely the inverse limit of $K^*(X^n)$ as X^n runs over the skeleta of X. Using Bott periodicity, they prove that $\mathbb{Z}$-graded K-theory satisfies all of the Eilenberg–Steenrod axioms except the dimension axiom and they introduce $\mathbb{Z}_2$-graded K-theory. Regarding ordinary rational cohomology as $\mathbb{Z}_2$-graded by sums of even and odd degree elements, they prove that the Chern character extends to a multiplicative map of cohomology theories $ch\colon K^*(X) \to H^{**}(X;\mathbb{Q})$ which becomes an isomorphism when the domain is tensored with $\mathbb{Q}$.

They also introduce what is now called the Atiyah–Hirzebruch spectral sequence. It satisfies

$$E_2^{p,q} = H^p\big(X; K^q(pt)\big),$$

and it converges to $K^*(X)$. Since it is compatible with Bott periodicity, it may be regraded so as to eliminate the grading q. It collapses, $E_2 = E_\infty$, if $H^*(X;\mathbb{Z})$ is concentrated in even degrees or, using the Chern character, if $H^*(X;\mathbb{Z})$ has no torsion. They state without proof that d_3 can be identified with the integral operation Sq^3, and they give partial information about the product structure. They also state without proof that the spectral sequence generalizes to a Serre type spectral sequence for the K-theory of fibre bundles.

The Riemann–Roch theorem of their earlier paper [AH59] is generalized to the cohomology theory K^*, but still with no hint of K-homology and a genuine pushforward map in K-theory. The theorem states that if $f : M \to N$ is a continuous map between compact oriented differentiable manifolds and if there is a given element $c_1(f) \in H^2(M;\mathbb{Z})$ such that $c_1(f) \equiv w_2(M) - f^*w_2(N) \bmod 2$, then, for $x \in K^*(M)$,

$$f_!\big(ch(x)e^{c_1(f)/2} \cdot \hat{A}(M)\big) = ch\big(f_!(x)\big) \cdot \hat{A}(N) \qquad (12.1)$$

in $H^*(N;\mathbb{Q})$. On the left $f_!$ is the pushforward in rational cohomology determined by Poincaré duality and f_*; a posteriori, $f_!$ is defined similarly in K-theory.

Using both the Riemann–Roch theorem and the spectral sequence, they study the K-theory of certain differentiable fiber bundles and compute $K^*(G/H)$ explicitly when H is a closed connected subgroup of maximal rank in a compact connected Lie group G. Moreover, when $H^*(G;\mathbb{Z})$ has no torsion, they prove that the natural map $R(H) \to K(G/H)$ is surjective. Calculations with the maximal rank condition dropped came much later.

Taking $\mathcal{K}(BG)$ to be the inverse limit K-theory of BG, they define a homomorphism $\alpha : R(G)_I^\wedge \to \mathcal{K}(BG)$ and prove that it is an isomorphism when G is a compact *connected*

Lie group. They also prove that $\mathcal{K}^1(BG) = 0$ for such G. The proof is by direct calculation when T is a torus and by comparison with the result for a maximal torus in general. They conjecture that this result remains true for any compact Lie group G.

In [At61b], which appeared in 1961, Atiyah proves the same result for *finite* groups G. The proof is by direct calculation when G is cyclic, by induction up a composition series when G is solvable, and by application of the Brauer induction theorem to pass from solvable groups to general finite groups. The second step depends on a Hochschild–Serre type spectral sequence that satisfies $E_2^{p,q} = H^p(G/N; \mathcal{K}^q(BN))$ and converges to $\mathcal{K}^*(BG)$, where N is a normal subgroup of G. The last step depends on the transfer homomorphisms in K-theory associated to finite covers. Atiyah claims in a footnote that the result does remain true for general compact Lie groups. However, a proof did not appear until the 1969 paper [AS69] of Atiyah and Segal, which is based on the use of equivariant K-theory. This was developed in lectures at Harvard and Oxford in 1965, but the first published accounts appeared later [At66a, Seg68].

In 1961 [AH61c], Atiyah and Hirzebruch make use of real K-theory KO to obtain a number of interesting results on characteristic classes in ordinary mod p cohomology. These are less well-known than they ought to be, perhaps because [AH61c] is written in German; some of its results were later reworked by Dyer [Dyer69]. Atiyah and Hirzebruch greatly extend and clarify observations Hirzebruch had already made in 1953 [Hirz53], and they improve results in the expository paper [AH61c], also in German, which was written a bit earlier and contains a nice general overview of the authors' results on K-theory, including some that I will not discuss here.

In [AH61c], using Milnor's analysis of the Steenrod algebra, Atiyah and Hirzebruch first determine the group of natural ring isomorphisms $\lambda : H^{**}(X) \to H^{**}(X)$. The obvious examples are $\lambda = Sq \equiv \sum Sq^r$ if $p = 2$ and $\lambda = P \equiv \sum P^r$ if $p > 2$. For a $\mathbb{Z}_p$-oriented vector bundle ξ with Thom isomorphism ϕ, they define $\underline{\lambda}(\xi) = \phi^{-1}\lambda\phi(1)$. Thus $\underline{Sq}$ is the total Stiefel–Whitney class and $\underline{P}$ is the total Wu class. They observe that, for a finite CW complex X, $\underline{\lambda}$ extends to a natural homomorphism from $KO(X)$ to the group $G^{**}(X)$ of elements of $H^{**}(X)$ with zeroth component 1 and, if $p > 2$, odd components zero, where the multiplication in $G^{**}(X)$ is given by the cup product. Write $Wu(\lambda, \xi) = \lambda^{-1}\underline{\lambda}(\xi)$. Then, when $p = 2$,

$$Wu(Sq, \xi) = \sum_{i \geqslant 0} 2^i T_i\big(w_1(\xi), \ldots, w_i(\xi)\big),$$

where the T_i are the Todd polynomials. Here the right side makes sense since $2^i T_i$ is a rational polynomial with denominator prime to 2. When $p > 2$, let $f = p^{1/p-1}$ and let P_i be the i-th Pontryagin class. Then

$$Wu(P, \xi) = \sum_{i \geqslant 0} f^{2i} L_i\big(P_1(\xi), \ldots, P_i(\xi)\big) = \sum_{i \geqslant 0} f^{2i} \hat{A}_i\big(P_1(\xi), \ldots, P_i(\xi)\big).$$

In both cases, there is an implied analogue for complex bundles, with Chern classes appearing on the right-hand sides of the equations.

These formulas suggest a relationship between the differential Riemann–Roch theorem and Wu's formulas for the characteristic classes of manifolds. Let $f : M \to N$ be a continuous map between differentiable manifolds M and N. Atiyah and Hirzebruch prove that, for any $x \in H^*(M)$,

$$f_!\big(\lambda(x)\cdot Wu(\lambda^{-1},\tau_M)\big) = \lambda\big(f_!(x)\big)\cdot Wu(\lambda^{-1},\tau_N), \tag{12.2}$$

where $f_!$ is the pushforward map determined by Poincaré duality and f_*. When N is a point, this reduces to

$$\big\langle \lambda(y), [M]\big\rangle = \big\langle \big(y\cdot Wu(\lambda,\tau_M)\big), [M]\big\rangle.$$

Taking $\lambda = Sq$ if $p = 2$ or $\lambda = P$ if $p > 2$, this is Wu's formula for the determination of the Stiefel–Whitney or L-classes of M in terms of Steenrod operations and cup products in $H^*(M)$.

It should be remarked at this point that Adams [Ad61b] proved the Wu relations for not necessarily differentiable manifolds in 1961. In 1960 [Ad61a], he proved an integrality theorem for the Chern character. Atiyah and Hirzebruch [AH61c] observe that (12.2) is an analogue of the differentiable Riemann–Roch theorem (12.1), and they show that this is more than just an analogy by using Adams' integrality theorem to derive important cases of (12.2) from (12.1). In a noteworthy remark, they point out that one can ask for such a Riemann–Roch type theorem whenever one has a natural transformation from one generalized cohomology theory to another, provided that both theories satisfy an analogue of Poincaré duality that allows pushforwards to be defined. This still precedes Poincaré duality in K-theory.

Even without K-homology, Atiyah in 1962 [At62] found an ingenious and influential proof of a Künneth theorem for K-theory, obtaining a short exact sequence of the expected form

$$0 \to K^*(X)\otimes K^*(Y) \xrightarrow{\alpha} K^*(X\times Y) \xrightarrow{\beta} \mathrm{Tor}\big(K^*(X), K^*(Y)\big) \to 0.$$

13. Generalized homology and cohomology theories

The work of G.W. Whitehead [Wh60, Wh62a] and Brown [Br63, Br65] defined and characterized represented generalized homology and cohomology theories in close to their modern forms. We have seen that K-homology is nowhere mentioned in the work of Atiyah and Hirzebruch. However, Whitehead's announcement [Wh60] of his definition of represented homology was already submitted in February 1960, and appeared that year, although the full paper [Wh62a] was not submitted until May, 1961, and appeared in 1962. More surprisingly, [Wh62a] makes no mention of either K-theory or bordism and contains no references to Atiyah and Hirzebruch, although the Bott spectrum is mentioned briefly. There seems to have been little mutual influence.

It seems that the main influence on Whitehead was his own earlier work on the homotopy groups of smash products of spaces [Wh56] and the work on duality of Spanier and J.H.C. Whitehead [SW55] and its further development by Spanier [Sp59b]. Whitehead defines a spectrum E to be a sequence of spaces E_i and maps $\sigma_i : \Sigma E_i \to E_{i+1}$, dropping the convergence conditions that Spanier imposed. He says that E is an Ω-spectrum if the adjoint maps $\widetilde{\sigma} : E_i \to \Omega E_{i+1}$ are homotopy equivalences. Actually, he insists on spaces

E_i for all integers i, rather than for $i \geqslant 0$ as is now more usual. He defines a map $f : E \to E'$ to be a sequence of maps $f_i : E_i \to E_i'$ such that the diagrams

$$\begin{array}{ccc} \Sigma E_i & \xrightarrow{\sigma_i} & E_{i+1} \\ {\scriptstyle \Sigma f_i}\downarrow & & \downarrow{\scriptstyle f_{i+1}} \\ \Sigma E_i' & \xrightarrow{\sigma_i'} & E_{i+1}' \end{array} \tag{13.1}$$

commute *up to homotopy*, and he says that two maps f and g are homotopic if $f_i \simeq g_i$ for all i.

Taking the obvious steps beyond Spanier [Sp59b], Whitehead defines the function spectrum $\mathbb{F}(X, E)$ and the smash products $E \wedge X \cong X \wedge E$ between a based space X and a spectrum E. As an unfortunate choice, he restricts X to be compact in these definitions, and his homology and cohomology theories are only defined on finite CW complexes. Remember that the additivity axiom came a bit later. In particular, these definitions give $\Omega E = \mathbb{F}(S^1, E)$ and $\Sigma E = E \wedge S^1$ (except that he writes the suspension coordinate on the left). Defining the homotopy groups of spectra as in (11.1), he proves that suspension gives an isomorphism $\Sigma_* : \pi_q(E) \to \pi_{q+1}(\Sigma E)$.

For finite based CW complexes X and a spectrum E, Whitehead defines

$$\widetilde{H}_q(X; E) = \pi_q(E \wedge X). \tag{13.2}$$

This is suggested by the more obvious cohomological analogue

$$\widetilde{H}^q(X; E) = \pi_{-q}\big(F(X, E)\big). \tag{13.3}$$

In retrospect, this definition of homology is correct for general CW complexes X, but this definition of cohomology is only correct for general CW complexes X when E is an Ω-spectrum.

Much of [Wh62a] is concerned with products in generalized homology and cohomology theories. These are induced by pairings $(D, E) \to F$ of spectra, which are specified by maps

$$D_m \wedge E_n \to F_{m+n}$$

that are suitably compatible up to homotopy with the structure maps σ of D, E, and F. Starting from such pairings of spectra, Whitehead defines and studies the properties of external products

$$\widetilde{H}_m(X; D) \otimes \widetilde{H}_n(Y; E) \to \widetilde{H}_{m+n}(X \wedge Y; F),$$
$$\widetilde{H}^m(X; D) \otimes \widetilde{H}^n(Y; E) \to \widetilde{H}^{m+n}(X \wedge Y; F)$$

and slant products

$$\backslash : \widetilde{H}_n(X \wedge Y; D) \otimes \widetilde{H}^m(X; E) \to \widetilde{H}_{n-m}(Y; F),$$
$$/ : \widetilde{H}_n(X \wedge Y; D) \otimes \widetilde{H}_m(Y; E) \to \widetilde{H}^{n-m}(X; F).$$

He obtains cup and cap products by pulling back along diagonal maps. By now, all of this is familiar standard practice.

Similarly, the familiar duality theorems are proven. Whitehead defines a ring spectrum E in terms of a product $(E, E) \to E$ and unit $S \to E$, where S is the sphere spectrum, namely, the suspension spectrum of S^0. He defines an E-orientation of a compact connected n-manifold M in terms of a fundamental class in $\widetilde{H}_n(M; E)$, and he proves a version of Alexander duality for dual pairs embedded in M. This specializes to give Poincaré duality for M. Taking $M = S^{n+1}$, which is E-oriented for any E, it specializes to give Spanier–Whitehead duality in any theory.

When [Wh62a] was written, Brown [Br63] had already proven his celebrated representation theorem. That paper also gave an incorrect first approximation to Milnor's additivity axiom [Mil62a]. In fact, James and Whitehead [JW58] had exhibited homology theories that fail to satisfy the additivity axiom and whose existence contradicted one of Brown's results. The correction of [Br63] noted this and pointed out simpler axioms for the representability theorem. Brown later published the improved version in a general categorical setting [Br65]. That version is one of the foundation stones of modern abstract homotopy theory.

Let k be a contravariant set-valued homotopy functor defined on based CW complexes. The functor k is said to satisfy the Mayer–Vietoris axiom if, for a pair of subcomplexes A and B of a CW complex X with union X and intersection C, the natural map from $k(X)$ to the pullback of the pair of maps $k(A) \to k(C)$ and $k(B) \to k(C)$ is surjective; k is said to satisfy the wedge axiom if it converts wedges to products. Brown in [Br65] proves that $k(X)$ is then naturally isomorphic to $[X, Y]$ for some CW complex Y. If k is only defined on finite CW complexes, Brown reaches the same conclusion but with a countability assumption on the $k(S^q)$. Adams [Ad71a] later showed that the countability assumption can be removed when the functor k is group-valued.

Applied to the term $\tilde{k}^n(-)$ of a (reduced) generalized cohomology theory $\tilde{k}^*$, Brown's theorem gives a CW complex E_n such that $\tilde{k}^n(X) \cong [X, E_n]$ for all CW complexes X. The suspension axiom on the theory leads to homotopy equivalences $E_n \to \Omega E_{n+1}$. Thus a cohomology theory $\tilde{k}^*$ gives rise to an Ω-spectrum E. Whitehead [Wh62a] followed up by using Spanier's version [Sp59b] of duality theory to show that a homology theory gives rise to a cohomology theory on finite CW complexes. Applying Brown's theorem for finite CW complexes (and using Adams' variant to avoid countability hypotheses), it follows that a homology theory on finite CW complexes is also representable by a spectrum.

Since the Brown representation is natural, a map of cohomology theories gives rise to a map of Ω-spectra. Defining the category of cohomology theories on spaces in the evident way, we see that it is equivalent to the homotopy category of Ω-spectra E whose spaces E_n are homotopy equivalent to CW complexes. We call this the Whitehead category of Ω-spectra. Milnor's basic result [Mil59] that the loop space of a space of the homotopy type of a CW complex has the homotopy type of a CW complex is relevant here.

Via the suspension spectrum functor and a functor that converts spectra to Ω-spectra, one can check that the S-category of finite CW complexes embeds as a full subcategory of the Whitehead category. Thus the Whitehead category is an approximation to stable homotopy theory that substantially improves on the S-category by providing the proper home for cohomology theories on spaces. However, as we shall see in Section 21, this is not yet the genuine stable homotopy category.

In the summer of 1962, there was an International Congress in Stockholm, preceded by a colloquium on algebraic topology at Aarhus. The proceedings of the latter contain brief expositions of generalized cohomology by Dold [Dold62], Dyer [Dyer62], and Whitehead [Wh62b]. Dold was the first to make the important observation that rational cohomology theories are products of ordinary cohomology theories, and he gave the first general exposition of the Atiyah–Hirzebruch spectral sequence. Making systematic use of Brown's representability theorem, his later book [Dold66], in German, gave a complete treatment of these matters and much else. Dyer was the first to write down a general treatment of the Riemann–Roch theorem, although already in 1962 he described the result as a folk theorem known to Adams, Atiyah, Hirzebruch and others. His later book [Dyer69] gave a complete treatment, along with an exposition of much of the work of Atiyah and Hirzebruch described in the previous section. He still avoids use of K_*, but this appears implicitly in the form of Atiyah duality, which allows an appropriate definition of pushforward maps.

Not everything in cohomology theory was to be done using its represented form. For example, working directly from the axioms, Araki and Toda [AT65] made a systematic study of products in mod q cohomology theories and of Bockstein spectral sequences in generalized cohomology. Nevertheless, most work was to be simplified and clarified by working with represented theories.

14. Vector fields on spheres and $J(X)$

In the proceedings of the 1962 Aarhus and Stockholm conferences, Adams [Ad62d] described his solution of the vector fields on spheres problem [Ad62b, Ad62c] and outlined his work on the groups $J(X)$, which appeared gradually in [Ad63, Ad65a, Ad65b, Ad66a]. I summarized these papers in [May2], emphasizing their impact on later work and the reformulations that became possible with later technology. These applications of K-theory have been of central importance to the development of stable algebraic topology.

The key new idea was the introduction of the Adams operations ψ^k in real and complex K-theory. These play a role in K-theory that is of comparable importance to the role played by Steenrod operations in ordinary mod p cohomology. It was clear from Grothendieck's work [Gro57] how to extend the exterior power operations λ^k from vector bundles to K-theory. The "Newton polynomials" Q_k that express the power operations $x_1^k + \cdots + x_n^k$ in a polynomial ring $\mathbb{Z}[x_1, \ldots, x_n]$ as polynomials in the elementary symmetric polynomials σ_k were familiar to topologists from their role in the study of characteristic classes. Adams' ingenious idea was to define

$$\psi^k(x) = Q_k\big(\lambda^1(x), \ldots, \lambda^n(x)\big).$$

Here X is a finite CW complex, $x \in K(X)$, and n is large.

Either by a representation theoretical argument, as in [Ad62c], or by use of the splitting principle and reduction to the case of line bundles, one finds that the ψ^k are natural ring homomorphisms that commute with each other. They are easily evaluated on line bundles and on the K-theory of spheres, and their relationship to the Chern character and the Bott isomorphism are easily determined. They greatly enhance the calculational power of K-theory.

Adams discovered these operations after first trying to solve the vector fields on spheres problem by use of secondary and higher operations in ordinary cohomology in [Ad62a], a paper that was obsolete by the time it appeared. The idea that a problem that required higher order operations in ordinary cohomology could be solved using primary operations in K-theory had a strong impact on the directions taken by stable algebraic topology.

The vector fields on spheres problem asks how many linearly independent vector fields there are on S^{n-1}. The answer is $\rho(n)-1$. Here $\rho(n) = 2^c+8d$, where $n = (2a+1)2^b$ and $b = c+4d, 0 \leqslant c \leqslant 3$. It had long been known [Eck42] that there exist $\rho(n)-1$ such fields. Adams proved that there are no more. Work of James [Ja58a, Ja58b, Ja59] had reduced the problem to a question about the reducibility of a certain complex. Up to suspension, Atiyah [At61c] identified the S-dual of that complex with a stunted projective space. This reduced the problem to the question of the coreducibility of $X = \mathbb{RP}^{m+\rho(n)}/\mathbb{RP}^{m-1}$ for a suitable m. Here coreducibility means that there is a map $f : X \to S^m$ that has degree 1 when restricted to the bottom cell S^m of X. Adams proves that X is not coreducible, thus solving the problem.

For the proof, Adams starts with the calculation of $K(\mathbb{CP}^n)$ and $K(\mathbb{CP}^n/\mathbb{CP}^m)$, which was first carried out by Atiyah and Todd [AT60]. He next calculates $K(\mathbb{RP}^n)$ and $K(\mathbb{RP}^n/\mathbb{RP}^m)$. Finally he calculates $KO(\mathbb{RP}^n)$ and $KO(\mathbb{RP}^n/\mathbb{RP}^m)$. In each case, he obtains complete information on the ring structure and the Adams operations. The main tools are just the Atiyah–Hirzebruch spectral sequence and the Chern character. For X as above, the existence of a coreduction f and the naturality relation $f^*\psi^k = \psi^k f^*$ lead to a contradiction.

For a connected finite CW complex X, define $J(X)$ to be $\mathbb{Z} \oplus \tilde{J}(X)$, where $\tilde{J}(X)$ is the quotient of $\widetilde{K}(X)$ obtained by identifying two stable equivalence classes of vector bundles if they are stably fiber homotopy equivalent. Let $J : K(X) \to J(X)$ be the evident quotient map. Atiyah in [At61b] (where $J(X)$ means what we and Adams call $\tilde{J}(X)$) proved that the bundle $O(n)/O(n-k) \to S^{n-1}, n \geqslant 2k$, admits a section if and only if n is a multiple of the order of $J(1-\xi)$, where ξ is the canonical line bundle over $\mathbb{RP}^{k-1}$. Thus the vector fields problem can be viewed as a special case of the problem of determining $J(X)$. In fact, as Bott first observed [Bott62, Bott63], Adams' calculations in [Ad62c] imply that $KO(\mathbb{RP}^n) \cong J(\mathbb{RP}^n)$. While Adams was aware of the relationship between the vector fields problem and the study of J, he chose not to discuss this in [Ad62c]; he published a proof of the cited isomorphism in [Ad65a].

The results just discussed have complex analogues, using $U(n)/U(n-k)$ and $\mathbb{CP}^{k-1}$. The bundle $\pi_{n,k} : U(n)/U(n-k) \to S^{2n-1}$ admits a section if and only if n is divisible by a certain number M_k. The necessity was proven first, by Atiyah and Todd [AT60], and the sufficiency was then proven by Adams and Walker [AW64]. For the proof, they compute $KO(\mathbb{CP}^n)$ and $KO(\mathbb{CP}^n/\mathbb{CP}^m)$, use the methods and results of [Ad63, Ad65a] to study $J : KO(\mathbb{CP}^n) \to J(\mathbb{CP}^n)$, and deduce that the order of $J(1-\xi)$ is M_k, where ξ is the canonical line bundle over $\mathbb{CP}^{k-1}$.

Many of the results of Atiyah [At61b] and Adams [Ad62c] on stunted projective spaces have analogues for stunted lens spaces, and these were worked out by Kambe, Matsunaga, and Toda [Ka66, KMT66].

The papers [Ad63, Ad65a, Ad65b, Ad66a] carry out the general study of $J(X)$ for a connected finite CW complex X. The overall plan is to define two further, more computable, quotients $J'(X)$ and $J''(X)$ of $K(X)$ such that the quotient homomorphisms from $K(X)$ factor to give epimorphisms $J''(X) \to J(X) \to J'(X)$ and then to prove that $J'(X) = J''(X)$. Thus $J'(X)$ is a lower bound and $J''(X)$ an upper bound for $J(X)$, and these two bounds coincide.

That $J''(X)$ really is an upper bound depends on the Adams conjecture:

> If k is an integer, X is a finite CW complex and $y \in KO(X)$, then there exists a non-negative integer $e = e(k, y)$ such that $k^e(\psi^k - 1)y$ maps to zero in $J(X)$.

Adams [Ad63] proved this when y is a linear combination of $O(1)$ or $O(2)$ bundles and when $X = S^{2n}$ and y is a complex bundle. His proof is based on the "Dold theorem mod k", which asserts that if $f : \eta \to \xi$ is a fiberwise map of sphere bundles of degree $\pm k$ on each fiber, then $k^e\eta$ and $k^e\xi$ are fiber homotopy equivalent for some $e > 0$. For $k = 1$, this is a result of Dold [Dold63].

The groups $J'(X)$ and $J''(X)$ are defined and calculated in favorable cases in [Ad65a]. In particular, the image of J in π^s_{4k-1} is shown to be either the denominator of $B_k/4k$, as expected, or twice it; the expected answer would follow from the Adams conjecture. The group $J''(X)$ is $KO(X)/W(X)$, where $W(X)$ is the subgroup generated by all elements $k^{e(k)}(\psi^k - 1)y$ for a suitable function e. The content of the Adams conjecture is that $J''(X)$ is indeed an upper bound for $J(X)$.

To define $J'(X)$, Adams needs certain operations ρ^k which he calls "cannibalistic classes". They are related to the ψ^k as the Stiefel–Whitney classes are related to the Steenrod operations. That is, $\rho^k = \phi^{-1}\psi^k\phi(1)$, where ϕ is the KO-theory Thom isomorphism. This definition and calculations based on it require good control on KO-orientations of vector bundles. While Adams developed some of this himself, the published version of [Ad66a] relies on the paper [ABS64] of Atiyah, Bott, and Shapiro, and I shall say more about that in the next section. This definition only works for Spin($8n$)-bundles, in which case the operations ρ^k were introduced by Atiyah (unpublished) and Bott [Bott62, Bott63], who denoted them θ_k. Adams shows that the operations can be extended to all of $KO(X)$ if one localizes the target groups away from k. If sphere bundles η and ξ are fiber homotopy equivalent, then $\rho^k(\xi) = \rho^k(\eta)[\psi^k(1+y)/(1+y)]$ for some $y \in \widetilde{KO}(X)$, independent of k. The group $J'(X)$ is $KO(X)/V(X)$, where $V(X)$ is the subgroup of these elements x such that $\rho^k(x) = \psi^k(1+y)/(1+y)$ in $KO(X) \otimes \mathbb{Z}[1/k]$ for all $k \neq 0$ and some $y \in \widetilde{KO}(X)$.

Adams gives the proof that $J'(X) = J''(X)$ in [Ad65b]. This entails a good deal of representation theory, some of it involving the extension to the real case of arguments used by Atiyah and Hirzebruch [AH61a] in their comparison between $R(G)^\wedge_I$ and $K(BG)$ for a compact connected Lie group G. This is used to construct a certain diagram between K-groups, the motivation for which is the heuristic idea that $1 + y = \rho^k x$ is a solution of the equation $\rho^\ell(\psi^k - 1)x = \psi^\ell(1+y)/(1+y)$. This diagram is then proven to be a weak pullback by calculational analysis. To get a more precise hold on $J'(X)$, Adams proves that the ψ^k are periodic in the sense that, for any positive integer m, there is an exponent e, depending only on X, such that, for any $x \in KO(X)$, $\psi^k(x) \equiv \psi^\ell(x) \mod m$ if $k \equiv \ell \mod m^e$. He uses this to characterize which elements $(\nu_k) \in \prod_{k\neq 0}(1 + \widetilde{KO}(X)[1/k])$ are of the form $\nu_k = \rho^k(x)\psi^k[(1+y)/(1+y)]$ for some $x \in \widetilde{KSpin}(X)$ and $y \in \widetilde{KO}(X)$.

Modulo the Adams conjecture, Adams proves in [Ad66a] that $J(S^n)$ is a direct summand of π^s_n. He does this by studying invariants d and e that are associated to maps $f : S^{q+r} \to S^q$; there are two variants, real and complex. The real invariant $d_{\mathbb{R}}(f)$ is just the induced homomorphism f^* on $\widetilde{KO}$, and it is zero unless $r \equiv 1$ or $2 \mod 8$, when it detects certain well-known direct summands $\mathbb{Z}_2$ of π^S_*. When $d_{\mathbb{R}}(f) = 0$ and $d_{\mathbb{R}}(\Sigma f) = 0$, the cofiber sequence $S^q \to Cf \to S^{q+r+1}$ gives a short exact sequence on application of $\widetilde{KO}$, and $e_{\mathbb{R}}(f)$ is the resulting element of the appropriate Ext^1 group of extensions.

Here Ext^1 is taken with respect to an Abelian category of Abelian groups with Adams operations that commute with each other and satisfy the periodicity relations. Building in that much structure allows direct computation of the relevant Ext^1 group, which in the cases of interest is an explicitly determined subgroup of $\mathbb{Q}/\mathbb{Z}$. Adams' algebraic formalism leads to an analysis of how $e_{\mathbb{R}}$ relates Toda brackets in homotopy theory to Massey products in Ext groups, and these relations are the key to many of Adams' detailed calculations.

The real e-invariant is essential to the proof of the splitting of π_*^s. The complex e-invariant $e_{\mathbb{C}}$ admits a more elementary description in terms of the Chern character and was introduced and studied independently by Dyer [Dyer63] and Toda [To63]. Adams, Dyer, and Toda all show that $e_{\mathbb{C}}$ can be used to reprove the Hopf invariant one theorem, at all primes p. Adams [Ad66a] also uses $e_{\mathbb{C}}$ to prove that if Y is the mod p^f Moore space, p odd, with bottom cell in a suitable odd dimension, and if $r = 2(p-1)p^{f-1}$, then there is a map $A : \Sigma^r Y \to Y$ that induces an isomorphism on $\widetilde{K}$. Iterating A s times, by use of suspensions, and first including the bottom cell and then projecting on the top cell, there result elements $\alpha_s \in \pi_{rs-1}^s$, and Adams uses $e_{\mathbb{C}}$ to prove that these maps are all essential. This generalized and clarified a construction of Toda [To58a] and was a forerunner of a great deal of recent work on periodicity phenomena in stable homotopy theory. When $f = 1$, Toda himself [To63] showed how to use $e_{\mathbb{C}}$ to detect these elements as Toda brackets.

Once the Adams conjecture was proven, various classifying spaces not available to Adams were constructed, and the theories of localization and completion were developed, the proof that $J'(X) = J''(X)$ could later be carried out in a more conceptual homotopy theoretic way. The speculative last section of [Ad65b] anticipated much of this. Adams showed that, once appropriate foundations were in place, one would be able to deduce that, for any KO-oriented spherical fibration ξ of dimension $8n$, the sequence $\rho^k(\xi) = \phi^{-1}\psi^k\phi(1)$ would be of the form cited above. This would imply that, for any x in the group $\widetilde{K}(F; KO)(X)$ of KO-oriented stable spherical fibrations, there is an element $x' \in \widetilde{K}Spin(X)$ such that $\rho^k(x) = \rho^k(x')$ for all k. In retrospect, this was headed towards localized splittings of the classifying space for KO-oriented spherical fibrations, with one factor being $BSpin$ and the other a space $BCokerJ$ whose homotopy groups are essentially the cokernel of $J : \pi_*(BSpin) \to \pi_*^s$.

Adams asked, among other things, whether or not the $J(X)$ specify a natural direct summand of some other functor of X, and he observed that, since the $J(X)$ do not give a term in a cohomology theory on X, they cannot be direct summands of a term of a cohomology theory. We now fully understand the answers to his questions. The process of reaching that understanding was to have major impact on geometric topology and algebraic K-theory, as well as on many topics within algebraic topology.

15. Further applications and refinements of K-theory

The need for K and KO orientations of suitable vector bundles was apparent from the moment K-theory was introduced. Such orientations were essential to the work of Adams just discussed and were first studied in detail by Bott [Bott62, Bott63]. However, the definitive treatment was given in the beautiful paper [ABS64] of Atiyah, Bott, and Shapiro, which was written by the first two authors after Shapiro's untimely death.

The authors first give a comprehensive algebraic treatment of Clifford algebras and their relationship to spinor groups. Let C_k be the Clifford algebra of the standard negative definite quadratic form $-\sum x_i^2$ on $\mathbb{R}^k$ and let $M(C_k)$ be the free Abelian group generated by the irreducible $\mathbb{Z}_2$-graded C_k-modules. The inclusion of C_k in C_{k+1} induces a homomorphism $M(C_{k+1}) \to M(C_k)$. Let A_k be its cokernel. Then the groups A_k are periodic of period 8 and are isomorphic to the homotopy groups $\pi_k(BO)$. Their complex analogues A_k^c are isomorphic to the homotopy groups of BU. Under tensor product, the A_k and A_k^c form graded rings isomorphic to the positive dimensional homotopy groups of KO and KU. These facts are far too striking to be mere coincidences.

They next give an account of relative K-theory in bundle theoretic terms, proving that, for any n, a suitably defined set $L_n(X, Y)$ of equivalence classes of sequences of vector bundles over X, exact over Y and of length any fixed $n \geqslant 1$, maps isomorphically to $K(X, Y)$ under an Euler characteristic they construct. The proof depends on a difference bundle construction that is important in many applications.

Combining ideas, they view the algebraic theory as a theory of bundles over a point and generalize it to a theory of bundles over X. Starting from a fixed Euclidean vector bundle V over X, they construct an associated Clifford bundle $C(V)$ over X whose fiber over x is the Clifford algebra $C(V_x)$. They define $M(V)$ to be the Grothendieck group of $\mathbb{Z}_2$-graded $C(V)$-modules over X and define $A(V)$ to be the cokernel of the homomorphism $M(V \oplus 1) \to M(V)$. Using their explicit description of relative K-theory, an elementary construction gives a natural homomorphism

$$\chi_V : A(V) \to \widetilde{KO}\big(B(V), S(V)\big) \cong \widetilde{KO}(TV).$$

It is multiplicative on external sums of bundles in the sense that

$$\chi_V(E) \cdot \chi_W(F) = \chi_{V \oplus W}(E \otimes F).$$

If V is the associated bundle $V = P \times_{Spin(k)} \mathbb{R}^k$ of a principal $Spin(k)$-bundle P and M is a C_k-module, then $E = P \times_{Spin(k)} M$ is a $C(V)$-module. This gives a homomorphism $\beta_P : A_k \to A(V)$ and thus a composite homomorphism $\alpha_P = \xi_V \beta_P : A_k \to \widetilde{KO}(TV)$. Taking X to be a point and P to be trivial, there results a homomorphism of rings

$$\alpha : A_* \to \sum_{k \geqslant 0} KO^{-k}(pt).$$

The beautiful theorem now is that α and its complex analogue are isomorphisms of rings. This suggests that a proof of Bott periodicity based on the use of Clifford algebras should be possible. Using Banach algebras, Wood [Wood65] and Karoubi [Kar66, Kar68] later found such proofs.

Now consider a $Spin$-bundle $V \cong P \times_{Spin(n)} \mathbb{R}^n$, where $n = 8k$. Define $\mu_V = \alpha_P(\lambda^k) \in \widetilde{KO}(TV)$. Then μ_V restricts on fibers to the canonical generator of the free $KO^*(pt)$-module $KO^*(S^n)$. That is, it is an orientation of V, and so it induces a Thom isomorphism $\phi : KO^*(X) \to \widetilde{KO}^*(TV)$. It follows that a $Spin(8k)$-bundle V is KO-orientable if and only $w_1(V) = 0$ and $w_2(V) = 0$. The orientation is multiplicative in the sense that $\mu_{V \oplus W} = \mu_V \cdot \mu_W$. The authors prove that the orientation they construct coincides with that constructed earlier by Bott [Bott62, Bott63]. Similarly, they obtain an orientation $\mu_V^c \in \widetilde{KU}(TV)$ for a $Spin^c$-bundle of dimension $n = 2k$. They state that the agreement of their orientations with Bott's gives additional good properties, but they do not say what these properties are.

In [Ad65a], Adams explained some of these properties, since he needed them for computation. Note first that, since $U(k) \to SO(2k)$ lifts canonically to $Spin^c(2k)$, the orientations of $Spin^c$-bundles give orientations of complex bundles. The complexification of the orientation of a *Spin*-bundle V is the orientation of $V \otimes \mathbb{C}$. According to Adams, the Todd and $\hat{A}$ classes are given in terms of the K-theory and rational cohomology Thom isomorphisms by the formulas

$$e^{c_1(V)}T^{-1}(\xi) = \phi^{-1} ch \mu_V^c$$

for a complex bundle V and

$$\hat{A}^{-1}(V) = \phi^{-1} ch \mu_{V\otimes\mathbb{C}}^c$$

for a *Spin*-bundle V. According to Adams "It is well known that this is the way $\hat{A}$ enters the theory of characteristic classes". That is, $\hat{A}(M) \equiv \hat{A}(\tau) = \phi^{-1} ch \mu_{\nu\otimes\mathbb{C}}^c$, where τ is the tangent bundle of a manifold M with normal bundle ν.

We have noted the analogy between Adams operations and Steenrod operations. In the 1966 paper [At66a], Atiyah went further and showed that this analogy could be made into a precise mathematical relationship, at least for complex K-theory. He redefined the Adams operations by constructing a homomorphism of rings

$$j : R_* = \sum_k \mathrm{Hom}_{\mathbb{Z}}\left(R(\Sigma_k), \mathbb{Z}\right) \to \mathrm{Op}(K).$$

Here Σ_k is the k-th symmetric group, $R(\Sigma_k)$ its character ring, and $\mathrm{Op}(K)$ is the ring of natural transformations from the functor K to itself. This makes essential use of equivariant K-theory and the isomorphism $K_G(X) \cong K(X) \otimes R(G)$ for a finite group G and a space X regarded as a G-space with trivial action. The k-th tensor power of a vector bundle over X is a Σ_k-bundle over X, and this gives a k-th power map $K(X) \to K(X) \otimes R(\Sigma_k)$; composing with homomorphisms $R(\Sigma_k) \to \mathbb{Z}$, we obtain the k-th component of j. As a matter of algebra, there is a copy of the polynomial algebra generated by certain elements that deserve to be denoted ψ^k sitting inside R_*, and the images of the ψ^k under j are the Adams operations.

Making essential use of the construction of relative K-theory in [ABS64], this allows Atiyah to relate the Adams operations to Steenrod operations by a direct comparison of definitions. The K-theory of a CW complex X is filtered by $K_q(X) = \mathrm{Ker}(K(X) \to K(X^q))$ with associated graded group $E_0^* K(X)$. Suppose that $H_*(X)$ has no torsion and let p be a prime. The Atiyah–Hirzebruch spectral sequence implies an isomorphism $H^{2q}(X; \mathbb{Z}_p) \cong E_0^{2q} K(X) \otimes \mathbb{Z}_p$. Atiyah proves that, for $x \in K_{2q}(X)$, there are elements $x_i \in K_{2q+2i(p-1)}(X)$ such that $\psi^p(x) = \sum_{i=0}^q p^{q-i} x_i$. Writing $\bar{x}$ for the mod p reduction of x and letting $P^i = Sq^{2i}$ when $p = 2$, he then proves the remarkable formula $P^i(\bar{x}) = \bar{x}_i$. The idea of introducing Steenrod operations into generalized homology theories along the lines that Atiyah worked out in the case of K-theory has had many subsequent applications.

In another influential 1966 paper, Atiyah [At66b] introduced Real K-theory KR, which must not be confused with real K-theory KO. In the paper, real vector bundles mean one thing over "real spaces" and another thing over "spaces", which has bedeviled readers ever since: we distinguish Real from real, never starting a sentence with either. A Real space is just a space with a $\mathbb{Z}_2$-action, or involution, denoted $x \to \bar{x}$. A Real vector bundle $p : E \to X$ is a complex vector bundle E with involution such that $\overline{cy} = \bar{c}\,\bar{y}$ and $\overline{p(y)} = p(\bar{y})$ for $c \in \mathbb{C}$ and $y \in E$. There is a Grothendieck ring $KR(X)$ of Real vector bundles over a compact Real space X.

Atiyah shows that the elementary proof of the periodicity theorem in complex K-theory that he and Bott gave in [AB64] transcribes directly to give a periodicity theorem in KR-theory. The wonderful thing is that this general theorem specializes and combines with information on coefficient groups deduced from Clifford algebras to give a new proof of the periodicity theorem for real K-theory. An essential point is to introduce a bigraded version of KR-theory, as was first done by Karoubi [Kar66] in a more general context. In more modern terms, KR is a theory graded on the real representation ring $RO(\mathbb{Z}_2)$, and it is the first example of an $RO(G)$-graded cohomology theory. Such theories now play a central role in equivariant algebraic topology.

In Atiyah's notation, define groups

$$KR^{p,q}(X, A) = KR\big(X \times B^{p,q}, X \times S^{p,q} \cup A \times B^{p,q}\big),$$

where $B^{p,q}$ and $S^{p,q}$ are the unit disk and sphere in $\mathbb{R}^q \oplus i\mathbb{R}^p$. In the absolute case, these are the components of a bigraded ring. There is a Bott element $\beta \in KR^{1,1}(B^{1,1}, S^{1,1})$, and multiplication by β is an isomorphism. Setting $KR^p(X, A) = KR^{p,0}(X, A)$, it follows that $KR^{p,q}(X, A) \cong KR^{p-q}(X, A)$, and it turns out that this is periodic of period 8. When the involution on X is trivial, $KR(X) \cong KO(X)$, and this gives real Bott periodicity. Complex K-theory K and self-conjugate K-theory KSC, which is defined in terms of complex bundles E with an isomorphism from E to its conjugate, are also obtained from KR-theory by suitable specialization. This leads to long exact sequences relating real, complex, and self-conjugate K-theory that have been of considerable use ever since. The self-conjugate theory had been introduced by Green [Gr64] and Anderson [An64], who first discovered these exact sequences. The ideas in [At66b] have found a variety of recent applications. This is the paper of which Adams wrote in his review: "This is a paper of 19 pages that cannot adequately be summarized in less than 20".

In contrast, we come now to the definitive proof by K-theory of the Hopf invariant one theorem, for all primes p, that was given in the paper [AA66] of Adams and Atiyah. They give a complete proof of the Hopf invariant one theorem for $p = 2$ in just over a page (see also [May1]). The essential idea is to apply the relation $\psi^2\psi^3 = \psi^3\psi^2$ in the K-theory of a two-cell complex $S^n \cup_f e^{2n}$, n even. If the Hopf invariant of f is one, then a simple calculation shows that this relation leads to a contradiction unless n is 2, 4, or 8. The proof at odd primes takes only a little longer.

16. Bordism and cobordism theories

We now back up and return to the story of cobordism. Immediately after the introduction of K-theory, in 1960, Atiyah [At61a] introduced the oriented bordism and cobordism theories, denoted $MSO_*(X)$ and $MSO^*(X)$, for finite CW complexes X. Just as K^* was the first explicitly specified generalized cohomology theory, MSO_* was the first explicitly specified generalized homology theory.

For a finite CW pair (X, A) and any integer q, Atiyah defines

$$MSO^q(X, A) = \operatorname{colim}\left[\Sigma^{n-q} X/A, TSO(n)\right] \tag{16.1}$$

and verifies that these groups satisfy all of the Eilenberg–Steenrod axioms except the dimension axiom. This is the theory represented by the spectrum *MSO*, but Atiyah's work precedes Whitehead's paper [Wh62a], and that language was not yet available.

He defines oriented bordism groups geometrically. He proceeds a little more generally than is currently fashionable, but with good motivation. He considers the category $\mathcal{B}$ of pairs (X, α), where X is a finite CW complex (say) and α is a principal $\mathbb{Z}_2$-bundle over X, that is, a not necessarily connected double cover. Maps and homotopies of maps in $\mathcal{B}$ are bundle maps and bundle homotopies. For a smooth manifold M (with boundary), let γ denote the orientation bundle of M. Then $MSO_q(X, \alpha)$ is defined to be the set of "bordism classes" of maps $f : (M, \gamma) \to (X, \alpha)$, where M is a q-dimensional closed manifold. Here f is bordant to $f' : (M', \gamma) \to (X, \alpha)$ if there is a manifold W such that $\partial W = M \coprod M'$ together with a map $g : (W, \gamma) \to (X, \alpha)$ that restricts to f on M and to f' on M'. When α is trivial, f is just a map $M \to X$, where M is an oriented q-manifold, and Atiyah writes $MSO_q(X)$ for the resulting oriented bordism group. He observes that $MG_q(X)$ can be defined similarly for the other classical groups G.

One virtue of the more general definition is the observation that, for large n,

$$MSO_q(\mathbb{RP}^n, \xi) \cong \mathcal{N}_q, \tag{16.2}$$

where $\xi : S^n \to \mathbb{RP}^n$ is the canonical double cover. More deeply, Atiyah proves that, for an n-manifold M without boundary $MSO_q(M, \gamma)$ is isomorphic in the stable range $2q < n$ to a certain group $L_q(M)$ introduced by Thom [Thom54] and used in the proof of his "théorème fondamental". This allows Atiyah to show that Thom's theorem directly implies Poincaré duality: for a finite CW pair (X, A) such that $X - A$ is a closed oriented n-manifold

$$MSO^q(X, A) \cong MSO_{n-q}(X - Y, \gamma). \tag{16.3}$$

Taking Y to be empty and X to be oriented, this specializes to

$$MSO^q(X) \cong MSO_{n-q}(X).$$

Although he does not go into detail, Atiyah was aware of the expected interpretation in terms of cup and cap products induced from the maps

$$TSO(m) \wedge TSO(n) \to TSO(m+n).$$

For n large and even, so that $\gamma = \xi$, (16.2) and (16.3) imply that

$$\mathcal{N}_q \cong MSO^{2n-q}(\mathbb{RP}^{2n}). \tag{16.4}$$

One of Atiyah's main motivations was to understand certain exact sequences relating oriented and unoriented cobordism groups, in particular the exact sequence

$$\Omega_n \xrightarrow{2} \Omega_n \to \mathcal{N}_n, \tag{16.5}$$

due originally to Rohlin [Ro53, Ro58] and also proven by Dold [Dold60]. These exact sequences play a central role in Wall's computation of Ω_*. Using (16.4), Atiyah shows that they are just long exact sequences obtained by applying the theory MSO^* to pairs of projective spaces.

Conner and Floyd [CF64a] followed up Atiyah's work with a thorough exposition and many interesting applications of the theories MO_* and MSO_*. Atiyah did not give a geometric definition of the relative groups $MSO_*(X, A)$. Conner and Floyd do so carefully, and they prove that $MSO_*(X, A)$ so defined satisfies

$$MSO_q(X, A) \cong \pi_{n+q}\big(X/A \wedge TSO(n)\big) \quad \text{if } n \geqslant q + 2.$$

This shows that the geometrically defined theory agrees with the theory given by Whitehead's prescription. They construct the bordism Atiyah–Hirzebruch spectral sequence converging from $H_*(X, A; \Omega_*)$ to $MSO_*(X, A)$. For the unoriented theory, they show that

$$MO_*(X, A) \cong H_*(X, A; \mathbb{Z}_2) \otimes \mathcal{N}_*, \tag{16.6}$$

as we see from the splitting of MO as a product of Eilenberg–MacLane spectra. Similarly, they show that, modulo the Serre class of odd order Abelian groups,

$$MSO_*(X, A) \cong H_*(X, A; \Omega_*).$$

Using this, they reinterpret and generalize Thom's work on the Steenrod representation problem. For example, they show that the natural map $MSO_*(X, A) \to H_*(X, A; \mathbb{Z})$ is an epimorphism if and only if the oriented bordism spectral sequence for (X, A) collapses and that this holds if $H_*(X, A; \mathbb{Z})$ has no odd torsion. They also generalize (16.5) to an exact sequence

$$MSO_n(X, A) \xrightarrow{2} MSO_n(X, A) \to MO_n(X, A).$$

However, the main point of Conner and Floyd's monograph [CF64a] was the use of cobordism for the study of transformation groups of manifolds. The cohomological study of group actions was initiated in the remarkable early work of P.A. Smith [Sm38]. The use of cohomological methods in the study of transformation groups was systematized in the seminar [Bo60] of Borel and others, including Floyd. In its introduction, Borel had pointed out the desirability of making more effective use of differentiability assumptions than had been possible previously. Conner and Floyd introduced equivariant cobordism as a follow up, and they found many very interesting applications of it to the study of fixed point spaces of differentiable group actions. I shall only indicate a little of what they do.

They define oriented and unoriented geometric equivariant cobordism groups for any finite group G with respect to group actions on manifolds with isotropy groups constrained to lie in any set of subgroups of G closed under conjugacy. Write $\mathcal{N}_*^G$ and Ω_*^G for these groups when all subgroups are allowed as isotropy groups. Conner and Floyd focus on the case of free actions (trivial isotropy group). Here the geometric description of bordism theory directly implies that the cobordism groups of smooth compact manifolds with free G-actions are isomorphic to the bordism groups $MO_*(BG)$. Restricting to oriented manifolds and orientation preserving actions, the resulting cobordism groups are isomorphic to the bordism groups $MSO_*(BG)$. This opens the way to calculations. As in Atiyah's work on $K^*(BG)$, transfer homomorphisms play a significant role.

In the unoriented case, $MO_*(BG)$ is calculated in terms of $H_*(G; \mathbb{Z}_2)$ by (16.6). As an elementary application, Conner and Floyd give a geometric proof of Wall's observation that the square of a manifold is cobordant to an oriented manifold. However, the main applications concern the fixed point space F of a nontrivial smooth involution on a closed n-manifold M, which for clarity we assume to be connected. Let F^m be the union of the components of F of dimension m. If the Stiefel–Whitney classes of the normal bundle of F^m in M are trivial for $0 \leqslant m < n$, then F^m is a boundary for $0 \leqslant m < n$. This is a substantial generalization of the fact that F cannot have exactly one fixed point, a fact that, with its odd primary analogue, motivated their entire study.

Remarkably, although they did not have a description of $\mathcal{N}_*^{\mathbb{Z}_2}$ as the homotopy groups of a space, Conner and Floyd were able to compute these cobordism groups in terms of bordism groups; precisely, they obtained a split short exact sequence

$$0 \to \mathcal{N}_n^{\mathbb{Z}_2} \to \sum_{m=0}^{n} MO_m\bigl(BO(n-m)\bigr) \to MO_{n-1}(B\mathbb{Z}_2) \to 0.$$

For an odd prime p, Conner and Floyd calculate the bordism groups $MSO_*(B\mathbb{Z}_p)$ completely and give partial information on $MSO_*(B\mathbb{Z}_{p^k})$ for $k > 1$. They also study $MSO_*((B(\mathbb{Z}_p)^k))$, ending with a conjecture on annihilator ideals that was only proven much later. In this connection, they obtained partial information on a Künneth theorem for the computation of $MSO_*(X \times Y)$. Landweber [Lan66] later gave the complete result, along with the easier analogue for MU_*. Conner and Floyd went on to study the equivariant complex bordism groups $MU_*(BG)$ for free G-actions in [CF64b]. This work has been very influential in the development of both equivariant geometric topology and equivariant stable algebraic topology, which recently has become a major subject in its own right.

17. Further work on cobordism and its relation to K-theory

We have seen that Milnor [Mil60, Mil62b] and Novikov [Nov60, Nov65] raised the problem of determining the cobordism groups $\Omega_*^G \cong \pi_*(MG)$ of G-manifolds for $G = SU, Sp$ and *Spin*. They were aware that only the question of 2-torsion was at issue. Liulevicius [Liu64] described $H^*(MG, \mathbb{Z}_2)$ as a coalgebra over the Steenrod algebra for various G and began the study of the relevant mod 2 Adams spectral sequences. In particular, he calculated E_2 and showed that $E_2 \neq E_\infty$ for MSU and MSp. He also computed $\pi_*(MSp)$ in low dimensions. The calculation of the 2-torsion in $\pi_*(MSp)$ has been studied extensively over the last 30 years, and a complete answer is still out of sight. I shall say no more about that here. However, the remaining cases were all completely understood by the end of 1966.

The literature in this area burgeoned in the mid 1960's, and I will mention only some of the main developments. Stong [Sto68], unfortunately out of print, gives an excellent and thorough survey of results through 1967, with a complete bibliography. Foundationally, he starts from the systematic treatment of the geometric interpretation of $\pi_*(MG)$ that was given by Lashof in 1963 [Las63].

As a preamble to explicit calculations, Milnor [Mil65] and others gave some attractive conceptual results concerning the squares of manifolds. As a consequence of their work on fixed points of involutions in [CF64b], Conner and Floyd had observed that if $V_{\mathbb{R}}$ is the real form of a complex algebraic variety $V_{\mathbb{C}}$ and both are nonsingular, then $V_{\mathbb{C}}$ is unoriented cobordant to $V_{\mathbb{R}} \times V_{\mathbb{R}}$. Milnor [Mil65] showed that this implies that an unoriented cobordism class contains a complex manifold if and only if it contains a square. He also explained in terms of Stiefel–Whitney numbers when a manifold is unoriented cobordant to a complex manifold. Further, he conjectured and proved in low dimensions that the square of an orientable manifold is unoriented cobordant to a *Spin*-manifold. P.G. Anderson [And66] proved that the square of a torsion element of Ω_* is unoriented cobordant to an SU-manifold, and he deduced Milnor's conjecture from that. Stong [Sto66b] later gave a simpler proof.

In their monograph [CF66a], Conner and Floyd worked out the analogue of their development of geometric and represented oriented cobordism theory in the complex case, together with its SU variant. Although the details are a good deal more complicated, they follow the methods used by Wall [Wall60] and Atiyah [At61a] in the case of oriented cobordism to determine the additive structure of the SU-cobordism ring Ω_*^{SU}. The essential point is to determine the torsion, and they prove that the torsion subgroup of Ω_q^{SU} is zero unless $q = 8n + 1$ or $q = 8n + 2$, in which cases it is a $\mathbb{Z}_2$-vector space whose dimension is the number of partitions of n. Wall [Wall60] later completed the determination of the multiplicative structure of Ω_*^{SU}.

In concurrent work, Anderson, Brown, and Peterson [ABP66a] calculated the mod 2 Adams spectral sequence for $\pi_*(MSU)$. They use a result of Conner and Floyd [CF66a] to determine the differential d_2, and they deduce that $E_3 = E_\infty$. This is a more sophisticated application of the Adams spectral sequence than had appeared in earlier work, and it was the first significant example in which the Adams spectral sequence was determined completely despite the presence of nontrivial differentials. Moreover, they prove that an SU-manifold is a boundary if and only if all of its Chern numbers and certain of its (normal) KO-characteristic numbers are zero.

To define KO-characteristic numbers, they make one of the first explicit uses of Poincaré duality in KO-theory, relying on the Atiyah–Bott–Shapiro orientation to obtain canonical KO-fundamental classes of SU-manifolds. Another interesting feature of their work is the complete determination of the image of the framed cobordism groups Ω_*^{fr}, that is the stable homotopy groups of spheres, in Ω_*^{SU}. This allows them to connect up their calculations with the Kervaire surgery invariant and the realization of Poincaré duality spaces as SU-manifolds up to homotopy equivalence.

Soon afterwards, Anderson, Brown, and Peterson [ABP66b] followed up their work on Ω_*^{SU} with a calculation of Ω_*^{Spin}, which is a good deal harder. Let $bo\langle n\rangle$ denote the spectrum obtained from the real Bott spectrum by killing its homotopy groups in dimensions less than n. They construct a map f from *MSpin* to an appropriate product of copies of spectra $bo\langle 2n\rangle$ and suspensions of $H\mathbb{Z}_2$ and prove that f induces an isomorphism on mod 2 cohomology. A posteriori, f is a 2-local equivalence.

The essential input that makes this calculation possible is Stong's calculation [Sto63] of the mod 2 cohomology of the $bo\langle 2n\rangle$ as modules over the Steenrod algebra. These modules are of the form $A/A(Sq^1, Sq^2)$ or A/ASq^3, and this allows calculation of the relevant Adams spectral sequences. However, a good deal of work, most of it dealing with the algebra of modules over the Steenrod algebra, is needed to go from this input to the final conclusion. Incidentally, working on the space level, Adams had earlier calculated the mod p cohomologies of the $bu\langle 2n\rangle$ for all primes p [Ad61a].

Similarly to the case of Ω_*^{SU}, a *Spin*-manifold is a boundary if and only if all of its Stiefel–Whitney numbers and certain of its *KO*-characteristic numbers are zero. Moreover, a manifold is cobordant to a *Spin*-manifold if and only all of its Stiefel–Whitney numbers involving w_1 or w_2 are zero. The image of Ω_*^{fr} in Ω_*^{Spin} is determined by comparison with the case of Ω_*^{SU}. A result of Stong [Sto66a] determines the ring structure on the torsion free part of Ω_*^{Spin}.

In their monograph [CF66b], Conner and Floyd give a general exposition of the relationship between K-theory and cobordism, starting from a variant of the orientation theory of Atiyah, Bott, and Shapiro. They construct compatible natural transformations of multiplicative cohomology theories

$$\mu_c : \widetilde{MU}^*(X) \to \widetilde{K}^*(X)$$

and

$$\mu_r : \widetilde{MSU}^*(X) \to \widetilde{KO}^*(X)$$

on finite CW complexes X. Thinking of an element of $\widetilde{MU}^n(X)$ as a homotopy class of maps $f : S^{2k-n} \wedge X \to TU(k)$, k large, they obtain $\mu_c(f)$ by transporting the Thom class along the composite

$$\widetilde{K}\big(TU(k)\big) \xrightarrow{f^*} \widetilde{K}\big(S^{2k-n} \wedge X\big) \cong \widetilde{K}^n(X).$$

Up to sign, $\mu_c : \Omega_*^U \to \mathbb{Z}$ gives the Todd genus $T[M]$ of U-manifolds. Since μ_c is a ring homomorphism, it gives $\mathbb{Z}$ a structure of MU^*-module, where $MU^n = \Omega_{-n}^U$. Conner and Floyd prove the remarkable facts that complex cobordism determines complex K-theory and symplectic cobordism determines real K-theory. Precisely, the maps μ_c and μ_r induce isomorphisms

$$MU^*(X, A) \otimes_{MU^*} K^*(pt) \cong K^*(X, A)$$

and

$$MSp^*(X, A) \otimes_{MSp^*} KO^*(pt) \cong KO^*(X, A)$$

on finite CW pairs (X, A). The reason that *MSp* comes in is clear from the proof, which makes heavy use of the Atiyah–Hirzebruch spectral sequence and relies on the fact that $H^*(BSp)$ is concentrated in even degrees. Along the way, considerable information about characteristic classes in cobordism theories is obtained.

There is a last part of [CF66b] that deserves to be better known than it is. In slightly modernized terms, Conner and Floyd consider the cofiber MU/S of the unit $S \to MU$. They give a cobordism interpretation of $\pi_*(MU/S)$ in terms of U-manifolds with stably framed boundary, or (U,fr)-manifolds. The cofiber sequence gives rise to a short exact sequence

$$0 \to \Omega_{2n}^{U} \to \Omega_{2n}^{U,fr} \to \Omega_{2n-1}^{fr} \to 0$$

for each $n > 0$. The Todd genus defines a homomorphism $T : \Omega_*^{U,fr} \to \mathbb{Q}$, and it turns out that there is a closed U-manifold with the same Chern numbers as a given (U,fr)-manifold M if and only if $T(M)$ is an integer. Therefore T induces a homomorphism $\pi_{2n-1}(S) \cong \Omega_{2n-1}^{fr} \to \mathbb{Q}/\mathbb{Z}$. Conner and Floyd show that this homomorphism coincides with Adams' complex e-invariant. This allows the use of Adams' complete determination of the behavior of $e_{\mathbb{C}}$ to obtain geometric information. Using SU in place of U, they obtain a similar interpretation of Adams' real e-invariant $\pi_{8n+3}(S) \to \mathbb{Q}/\mathbb{Z}$, and they use explicit manifold constructions modelled on Toda brackets appearing in Adams' work to reprove the result of Anderson, Brown, and Peterson on the image of Ω_*^{fr} in Ω_*^{SU}.

The work of Conner and Floyd uses a basic theorem of Hattori [Ha66] and Stong [Sto66b]. The tangential characteristic numbers of a U-manifold M^{2n} determine a homomorphism $H^{2n}(BU; \mathbb{Q}) \to \mathbb{Q}$. Let I^{2n} be the subgroup of $H^{2n}(BU; \mathbb{Q})$ consisting of all elements that are mapped into $\mathbb{Z}$ by all such homomorphisms. The Riemann–Roch theorem of Atiyah and Hirzebruch shows that the $2n$ component of $ch(x)T$ is in I^{2n} for all $x \in K(BU)$, where T is the universal Todd class. Atiyah and Hirzebruch [AH61c] conjectured that these Riemann–Roch integrality relations are complete, in the sense that every element of I^{2n} is of this form.

This is the theorem of Hattori and Stong. It can be rephrased in several ways. Stong uses methods of cobordism to show that all homomorphisms $\Omega_n^U \to \mathbb{Z}$ are integral linear combinations of certain homomorphisms given by K-theory characteristic numbers. Hattori shows that the conjecture is equivalent to the assertion that, for large k, the homomorphism

$$\alpha : \widetilde{K}\big(TU(k)\big) \to \operatorname{Hom}\big(\pi_{2n+2k}\big(TU(k)\big), \widetilde{K}\big(S^{2n+2k}\big)\big)$$

given by $\alpha(y)(x) = x^*(y)$ is an epimorphism. He proves that the K-theory Hurewicz homomorphism

$$\pi_{2n+2k}\big(TU(k)\big) \to \widetilde{K}_{2n+2k}\big(TU(k)\big),$$

which is induced by the unit $S \to K$ of the K-theory spectrum, is a split monomorphism. He then deduces the required epimorphism property by use of Poincaré duality in K-theory. Adams and Liulevicius [AL72] later gave a spectrum level reinterpretation and proof of Hattori's theorem, viewing it as a result about the connective K-theory Hurewicz homomorphism of MU.

18. High dimensional geometric topology

The period that I have been discussing was of course also a period of great developments in high dimensional geometric topology. There was a closer interaction between algebraic and geometric topology throughout the period than there is today, and some of the most important work in both fields was done by the same people. Cobordism itself is intrinsically one such area of interaction. It would be out of place to discuss such related topics as h-cobordism and s-cobordism here. However, some geometric work was so closely intertwined with the main story or was to be so important to later developments that it really must be mentioned, if only very briefly.

First, there is the work of Kervaire and Milnor [KM63] on groups of homotopy spheres. This gives one of the most striking reductions of a problem in geometric topology to a problem in stable homotopy theory, albeit in this case to the essentially unsolvable one of computing the cokernel of the J-homomorphism.

As we have already mentioned, the starting point of modern differential topology was Milnor's discovery [Mil56b] of exotic differentiable structures on S^7. Kervaire and Milnor classify the differentiable structures on spheres in terms of the stable homotopy groups of spheres and the J-homomorphism. Let Θ_n be the group of h-cobordism classes of homotopy n-spheres under connected sum. By Smale's h-cobordism theorem [Sm62], Θ_n is the set of diffeomorphism classes of differentiable structures on S^n when $n \neq 3$ or 4. Kervaire and Milnor show that every homotopy sphere is stably parallelizable. The proof uses Adams' result [Ad65a] that $J : \pi_*(SO) \to \pi_*^s$ maps the torsion classes monomorphically. They then show that the homotopy spheres that bound a parallelizable manifold form a subgroup bP_{n+1} of Θ_n such that Θ_n / bP_{n+1} embeds as a subgroup of $\pi_n^s / J(\pi_n(SO))$. This embedding is an isomorphism if $n = 4k + 1$.

In the 1960's, geometric topologists began to take seriously the classification of piecewise linear and topological manifolds, and the appropriate theories of bundles and classifying spaces were developed. A few of the important relevant papers are those of Hirsch [Hir61], Milnor [Mil64], Kister [Ki64], Lashof and Rothenberg [LR65], and Haefliger and Wall [HW65]. We point out one conclusion that is particularly relevant to our theme, namely a theorem of Hirsch and Mazur that is explained in [LR65] and that is closely related to the work of Kervaire and Milnor just discussed. If M is a smoothable combinatorial manifold, then the set of concordance classes of smoothings of M is in bijective correspondence with the set of homotopy classes of maps $M \to PL/O$. Part of the explanation is that $PL(n)/O(n) \to PL/O$ induces an isomorphism of homotopy groups through dimension n, which converts an unstable problem into a stable one. Williamson [Wi66] proved the analogue of Thom's theorem for PL-manifolds, showing that the cobordism ring Ω_*^{PL} of oriented PL-manifolds is isomorphic to $\pi_*(MSPL)$, and similarly in the unoriented case.

These results raised the question of computing characteristic classes for PL and topological bundles and of computing the PL-cobordism groups. These calculational questions, which turn out to be closely related to the questions raised by Adams in [Ad65b], would later motivate a substantial amount of work in stable algebraic topology. A 1965 paper of Hsiang and Wall [HsW65] discussed the orientability of nonsmooth manifolds with respect to generalized cohomology theories. A year or two later, Sullivan discovered [Sull70] that PL-bundles admit canonical KO-orientations (away from 2). That fact has played an important role in answering such questions.

Although almost nothing was known about these questions in 1966, a useful conceptual guide to later calculations was published that year by Browder, Liulevicius, and Peterson [BLP66]. By then, classifying spaces $BF(n)$ for spherical fibrations were also on hand, by work of Stasheff [Sta63c] and later Dold [Dold66]. The authors consider a system of spaces $BG(n)$, where $G(n)$ may have no a priori meaning, and maps $BO(n) \to BG(n) \to$

$BF(n)$, $BG(n) \to BG(n+1)$ and $BG(m) \times BG(n) \to BG(m+n)$ satisfying some evident compatibility conditions. They define the Thom space $TG(n)$ by use of the pullback of the universal spherical fibration over $BF(n)$ and have a Thom spectrum MG. They have a Thom isomorphism $H^*(BG) \to H^*(MG)$ in mod 2 cohomology, and they define Stiefel–Whitney classes as usual.

With this set up, they observe that theorems of Milnor and Moore [MM65] imply that $H^*(MG)$ is a free A-module and there is a Hopf algebra $C(G)$ over A such that

$$H^*(BG) \cong H^*(BO) \otimes C(G)$$

as Hopf algebras over A and

$$\pi_*(MG) \cong \pi_*(MO) \otimes C(G)^*$$

as algebras. Letting BSG be the 2-fold cover of BG determined by w_1, they also observe that $H^*(MSG)$ is the direct sum of a free A-module and suspensions of copies of A/ASq^1, so that, a posteriori, MSG splits 2-locally as a product of corresponding Eilenberg–MacLane spectra, just as MSO does. Moreover

$$H^*(BSG) \cong H^*(BSO) \otimes C(G)$$

as Hopf algebras over A, and at least the additive structure of $\pi_*(MSG)$, modulo odd torsion, is determined by $C(G)$ and the Bockstein spectral sequence of $H^*(BSG)$.

Intuitively, this means that the mod 2 characteristic classes for "G-bundles" completely determine the unoriented G-cobordism ring and the 2-local part of the oriented G-cobordism ring. The proofs require no geometry, but when one has a manifold interpretation of $\pi_*(MG)$, for example when $G = PL$, it follows directly that a G-manifold is a boundary if and only if all characteristic numbers defined in terms of $H^*(BG)$ are zero.

For an odd prime p, they prove that $H^*(BSG; \mathbb{Z}_p)$ is a free B-module, where B is the sub Hopf algebra of A generated by the P^i, but the calculation of the odd torsion in $\pi_*(MG)$ requires use of the Adams spectral sequence and is thus of a quite different character than the determination of the 2-torsion.

19. Iterated loop space theory

So far I have focused on the mainstream of developments through 1966, but there are some other directions of work that were later to become important to stable algebraic topology. This section describes one stream of work that was later to merge with the mainstream. Although the connection was not yet visible in 1966 and won't be made visible here, the relevant later work was to provide key tools for the calculations called for in the previous section.

Let X be an H-space. One can ask whether or not X has a classifying space Y, so that $X \simeq \Omega Y$. If so, one can ask whether Y is an H-space. If so, one can ask whether Y has a classifying space. Iterating, one can ask whether X is an n-fold loop space, or even an infinite loop space. One wants the answers to be in terms of internal structure on the space

X. The answers are closely related to an understanding of the spaces $\Omega^n \Sigma^n X$, which play a role roughly dual to the role of Eilenberg–MacLane spaces in ordinary homotopy theory. Such questions were later to be a major part of stable algebraic topology, but some important precursors were on hand by 1966.

Recall that a topological monoid X, that is an associative H-space with unit, has a classifying space BX and $X \simeq \Omega BX$ if $\pi_0(X)$ is a group under the induced multiplication. This result has a fairly long history, which would be out of place here.

In 1957, Sugawara [Su57a] gave a fibration-theoretic necessary and sufficient condition for a space X to be an H-space, or to be a homotopy associative H-space. In the same year [Su57b], he followed up by giving necessary and sufficient conditions for X to have a classifying space. Obviously, X must be homotopy associative, but that is not sufficient. Sugawara described an infinite sequence of higher homotopies that must be present on loop spaces and showed that the existence of such homotopies is sufficient. Three years later [Su60], he took the next step and displayed an infinite sequence of higher commutativity homotopies such that a loop space ΩY has such homotopies if and only if Y is an H-space. Stasheff [Sta63a, Sta63b] later reformulated Sugawara's higher associativity homotopies in a much more accessible fashion, introducing A_n and A_∞-spaces. The latter are Sugawara's H-spaces with all higher associativity homotopies, and Stasheff reproved the result that such an H-space has a classifying space.

Systematic computations of $H_*(\Omega^n \Sigma^n X; \mathbb{Z}_p)$ began in 1956 with the work of Araki and Kudo [AK56]. Using higher commutativity homotopies, they mimic Steenrod's original construction of the Steenrod squares in mod 2 cohomology in terms of $\cup_i$-products to obtain mod 2 homology operations for n-fold loop spaces. They use these operations to compute $H_*(\Omega^n S^{n+k}; \mathbb{Z}_2)$. To compute $H_*(\Omega^n \Sigma^n X; \mathbb{Z}_2)$ for general spaces X, bracket operations of two variables are needed. These were introduced by Browder [Br60]. He reproved the results of Kudo and Araki by mimicking Steenrod's construction of Steenrod operations in terms of the homology of the cyclic group $\mathbb{Z}_2$, and he computed $H_*(\Omega^n \Sigma^n X; \mathbb{Z}_2)$ as a functor of $H_*(X; \mathbb{Z}_2)$. The functoriality is a calculational fact, not something true for general theories. For example, $K(\Omega^n \Sigma^n X)$ is not a functor of $K(X)$. It is related to Dold's earlier, but noncalculational, result [Dold58b] that the homologies of the symmetric products of X are determined by the homology of X.

Dyer and Lashof [DL62] studied homology operations for n-fold loop spaces at odd primes p, mimicking Steenrod's definition of Steenrod operations in terms of the homology of the symmetric group Σ_p. These operations are now generally called Dyer–Lashof operations. This method of construction does not give enough operations to compute $H_*(\Omega^n \Sigma^n X; \mathbb{Z}_p)$. Dyer and Lashof define $QX = \bigcup \Omega^n \Sigma^n X$, where the union is taken over the inclusions $\Omega^n \Sigma^n X \to \Omega^{n+1} \Sigma^{n+1} X$ obtained by suspending a map $S^n \to X \wedge S^n$ to a map $S^{n+1} \to X \wedge S^{n+1}$. They then prove that their operations, plus the Bockstein, are sufficient to compute $H_*(QX; \mathbb{Z}_p)$ as a functor of $H_*(X; \mathbb{Z}_p)$.

In 1966, Milgram [Mil66] generalized the James construction to obtain a combinatorial model $J_n X$ for $\Omega^n \Sigma^n X$, where X is a connected CW complex. The spaces $J_n X$ are themselves CW complexes with cellular chain complexes identified in terms of the cellular chains of X. This allows a computation of $H_*(\Omega^n \Sigma^n X)$, but Milgram's work was not connected up with homology operations until much later. This is analogous to Cartan's original computation of the homology of Eilenberg–MacLane spaces without use of Steenrod operations. The later theory of operads led to a simpler, but equivalent, model for $\Omega^n \Sigma^n X$ and allowed the specification of sufficiently many homology operations to compute $H_*(\Omega^n\Sigma^n X;\mathbb{Z}_p)$ as a functor of $H_*(X;\mathbb{Z}_p)$. It also made it clear that Milgram's work and Stasheff's work on A_∞-spaces are closely related, something that was not apparent at the time.

20. Algebraic K-theory and homotopical algebra

The cohomology of groups, homological algebra, algebraic K-theory, and category theory are algebraic areas of mathematics that developed simultaneously with stable algebraic topology and gradually evolved into separate subjects. All remain closely connected to stable algebraic topology. I shall mention some directions that seem to me to be of particular interest or to have been important forerunners of later developments.

I will point to just a few relevant papers concerning the cohomology of groups. Of course, with trivial coefficients, the homology of a discrete group G agrees with the homology of its classifying space BG. At about the same time that Dyer and Lashof were computing $H_*(QX)$, Nakaoka [Na60, Na61] was computing the homologies of the symmetric groups and in particular the homology of the infinite symmetric group. With $X = S^0$, it would later turn out that these were essentially the same computation.

The equivariant homology and cohomology groups of spaces that were studied by Borel and others in [Bo60] are $H_*^G(X) = H_*(EG\times_G X)$ and $H_G^*(X) = H^*(EG\times_G X)$. Swan [Sw60a] in 1960 introduced the Tate cohomology of spaces $\widehat{H}_G^*(X)$. Just as in group theory, he gave a long exact sequence relating $H_*^G(X)$, $H_G^*(X)$ and $\widehat{H}_G^*(X)$. It has been shown recently that one can replace ordinary homology and cohomology by the theories represented by any spectrum and still get such a long exact sequence.

Although a little off the subject, the early applications of algebraic K-theory to algebraic topology deserve brief mention. Swan [Sw60b] used his study [Sw60c] of projective modules over group rings to show that any finite group with periodic cohomology acts freely on a homotopy sphere. This led to Wall's K-theoretic finiteness obstruction [Wall65] that determines whether or not a finitely dominated CW complex is homotopy equivalent to a finite CW complex. Applications of algebraic K-theory in surgery theory also began in the 1960's, but are beyond our scope.

There are several papers in algebraic K-theory and what later became known as homotopical algebra that may be viewed as harbingers of things to come in stable algebraic topology. The feature to emphasize is the evolution from analogies between similarly defined objects in different subjects to direct mathematical connections and fruitful common generalizations. These topics were to have much direct contact with iterated loop space theory, but that could not have been visible in 1966. Their connections with the mainstream of stable algebraic topology were visible from the beginning, although the forms these connections would eventually take could not have been anticipated.

Topological K-theory grew directly out of Grothendieck's work, and the analogy with algebraic K-theory was thus visible from the outset. Swan [Sw62] gave the analogy mathematical content by proving that, for a compact space X, $K(X)$ is naturally isomorphic to the Grothendieck group of finitely generated projective modules over the ring $C(X)$ of continuous real-valued functions on X. The isomorphism sends a vector space ξ to the $C(X)$-module $\Gamma(\xi)$ of sections of ξ; $\Gamma(\xi)$ is a finitely generated projective $C(X)$-module since ξ is a summand of a trivial bundle.

As Adams wrote in his review of a paper of Bass [Bass64]:

> This leads to the following programme: take definitions, constructions and theorems from bundle-theory; express them as particular cases of definitions, constructions, and statements about finitely-generated projective modules over a general ring; and finally, try to prove the statements under suitable assumptions.

With this analogy clearly in mind, Bass defines and studies K^0 (following Grothendieck) and K^1 of rings in the cited paper. Higher algebraic K-groups came later, and their study would lead to substantial developments in stable algebraic topology that would be closely related to both high dimensional geometric topology and infinite loop space theory.

There are many other areas where analogies between algebra and topology have been explored. For example, starting with Eckmann–Hilton duality [Eck57, Hil58], there was considerable work in the late 1950's and early 1960's exploring the idea of a homotopy theory of modules, or, more generally, of objects in Abelian categories, by analogy with the homotopy theory of spaces. I shall say nothing about that work.

Rather, I shall say a little about the analogy between stable homotopy theory and differential homological algebra. Differential homological algebra studies such objects as differential graded modules over differential graded algebras and is a natural tool in both algebraic topology and algebraic geometry. The analogy between homotopies in topology and chain homotopies in homological algebra was already clear by 1945. However, the structural analogy between stable homotopy theory and differential homological algebra goes much deeper. It later led both to an axiomatic understanding of homotopy theory in general categories and to concrete mathematical comparisons between such categories in topology and algebra, beginning with the fundamental work of Quillen [Qu67].

Dold and Puppe gave important precursors of this in the early 1960's. The first systematic exploration of the analogy was given by Dold [Dold60], in 1960. He develops cofiber sequences of chain complexes of modules over a ring, a Whitehead type theorem for such chain complexes, Postnikov systems of chain complexes, and so forth. The next year [DP61], Dold and Puppe gave a remarkable and original use of simplicial methods in algebra by defining and studying derived functors of non-additive functors between Abelian categories. Unlike the additive case, these functors do not commute with suspension. This fact is analyzed by use of a bar construction defined in terms of cross-effect functors that measure the deviation from additivity.

In 1962, Puppe [Pu62], motivated by the need for a good stable homotopy category, gave an axiomatic treatment of exact triangles. That paper precedes the introduction of the derived category of chain complexes over a ring in Verdier's 1963 thesis (which was published much later [Ver71]). Verdier's axioms for exact triangles give the notion of a "triangulated category". Algebraic topologists and algebraic geometers have developed several areas of differential homological algebra independently, with different details, nomenclature, and, of course, assignment of credit. The definition of triangulated categories is a case in point.

About the same time as Stasheff's work on A_∞ spaces, and with mutual influence, MacLane [Mac65] in 1963 introduced coherence theory in categorical algebra. This explains what it means for a category to have a product that is associative, commutative, and unital "up to coherent natural isomorphism". The coherence isomorphisms are analogues of higher homotopies in topology. In familiar examples, like Cartesian products and tensor products, the isomorphisms are so obvious that they hardly seem worth mentioning. In less obvious situations, they require serious attention. The analogy between coherence isomorphisms and higher homotopies was later to be given mathematical content via infinite loop space theory, with extensive applications to algebraic K-theory.

Also around the same time, Adams and MacLane collaborated in the development and study of certain algebraic categories, the "PROPs" and "PACTs" discussed briefly in [Mac65]. Their goal was to understand coherence homotopies in differential homological algebra. I have gone through a box full of correspondence between Adams and MacLane and can attest that this was one of the largest scale collaborations never to have reached print. When later translated into topological terms, their work was to be very influential in infinite loop space theory; the original algebraic motivation reached fruition much more recently.

21. The stable homotopy category

In our discussion of the Adams spectral sequence and of cobordism, we have indicated the need for a good stable homotopy category of spectra, and we have discussed the S-category of Spanier and J.H.C. Whitehead [SW57] and the category of G.W. Whitehead [Wh62a] as important precursors. We begin this section by discussing a very important 1966 paper for which such foundations are needed.

We have seen that the quotient $B = A/(\beta)$ of the Steenrod algebra appears naturally in the study of cobordism. For all classical groups G and for $G = PL$, $H^*(MG; \mathbb{Z}_p)$ is a free B-module for each odd prime p, where we think of B as a sub-Hopf algebra of A. In the classical group case, but not in the case of PL, $H^*(MG; \mathbb{Z})$ is torsion free. For each prime p, Brown and Peterson [BP66] construct a spectrum, now called BP, such that $H^*(BP; \mathbb{Z}_p) \cong B$ as an A-module. They then prove that any spectrum X whose mod p cohomology is a free B-module and whose integral cohomology is torsion free admits a map f into a product of suspensions of BP that induces an isomorphism on mod p cohomology. A posteriori, f is a p-local equivalence. Since Brown and Peterson compute the homotopy groups of BP, one can read off the homotopy groups of X, modulo torsion prime to p. The method of proof is to use Milnor's results on the structure of A to write down a free resolution of $A/(\beta)$ as an A-module and then to realize the resolution by an inductive construction of a generalized Postnikov system whose inverse limit is BP.

This was the first time that a spectrum with desirable properties was tailor made. The spectra studied earlier had been ones that occurred "in nature" as sequences of spaces. For the foundations of their work, Brown and Peterson write

> We will make various constructions on spectra, for example, forming fibrations and Postnikov systems, just as one does with topological spaces. For the details of this see [–].

The reference they give in [–] is Whitehead [Wh62a]. However, the Whitehead category is not designed for this purpose and is not triangulated. Intuitively, one needs a category that is equivalent to the category of cohomology theories on spectra, not just spaces.

Moreover, it would later be seen that BP, like S, K, KO, and the MG is a "commutative and associative ring spectrum". To attach a satisfactory meaning to this notion, one needs a smash product in the stable homotopy category of spectra that is associative, commutative, and unital up to coherent natural isomorphism. A ring spectrum R is then a spectrum with a product $\wedge : R \wedge R \to R$ and unit $S \to R$ such that the appropriate diagrams commute in the stable homotopy category.

The minimal requirements of a satisfactory stable homotopy category $\mathcal{S}_h$ include the following very partial list.

(1) It must have a suspension spectrum functor $\Sigma^\infty : \mathcal{C}_h \to \mathcal{S}_h$, where $\mathcal{C}_h$ is the homotopy category of based CW complexes.

(2) It must have a suspension functor $\Sigma : \mathcal{S}_h \to \mathcal{S}_h$ such that $\Sigma\Sigma^\infty \cong \Sigma^\infty\Sigma$.

(3) Σ^∞ must induce a full embedding of the S-category of finite CW complexes, so that Spanier–Whitehead duality makes sense.

(4) $\mathcal{S}_h$ must represent cohomology theories: isomorphism classes of objects E of $\mathcal{S}_h$ must correspond bijectively to isomorphism classes of cohomology theories $\widetilde{E}^*$ in such a way that, $\mathcal{S}_h(\Sigma^\infty X, E) \cong \widetilde{E}^0(X)$ for based CW complexes X.

(5) $\mathcal{S}_h$ must be triangulated; in particular, $\mathcal{S}_h$ must be an additive category and $\Sigma : \mathcal{S}_h \to \mathcal{S}_h$ must be an equivalence of categories.

(6) $\mathcal{S}_h$ must be symmetric monoidal under a suitably defined smash product.

It is not an easy matter to construct such a category, and a rigorous development of modern stable algebraic topology would not have been possible without one.

Adams made several attempts to construct such a category, first in a very brief account in 1959 [Ad59] and then in more detail in his 1961 Berkeley notes [Ad61c]. There he gave an amusing discussion of the approaches a hare and a tortoise might take. In retrospect, his decision to come down on the side of the tortoise was misguided: a more inclusive and categorically sophisticated approach was needed. In [Ad66b], Adams assumed the existence of a good stable category and sketched the development of an Adams spectral sequence based on connective K-theory. This was the first attempt at setting up an Adams spectral sequence based on a generalized cohomology theory. However, convergence was not proven and the approach was still based on cohomology rather than homology.

More fruitful approaches would come a little later, with the development of the Adams–Novikov spectral sequence [Nov67], namely the Adams spectral sequence based on MU or, more usefully, BP. This, together with Quillen's observation of the relationship between complex cobordism and formal groups [Qu69], would lead later to the realization that MU and spectra constructed from it are central to the structural analysis of the stable homotopy category.

Adams' version [Ad61c] of the stable homotopy category and the slightly later version of Puppe [Pu67], following up [Pu62], were based on the use of spectra T such that T_n is a CW complex and ΣT_n is a subcomplex of T_{n+1}. Connectivity and convergence conditions were imposed. In Adams, these had the effect that all spectra were (-1)-connected. In Puppe, they had the effect that all spectra were bounded below. The specification of maps was a little complicated. Roughly, the basic diagrams (13.1) were required to commute on the point-set level rather than only up to homotopy, as in Whitehead's category, but maps were not required to be defined on the whole spectrum, only on some cofinal part of it. Puppe's category was triangulated, and his discussion of exact triangles has been quite influential.

Kan [Kan63a, Kan63b] introduced simplicial spectra in 1963 and began the development of the stable homotopy category in terms of them. Simplicial spectra are not defined as sequences of simplicial sets and maps, but rather as generalized analogues of simplicial sets that admit infinitely many face operators in each simplicial degree.

Neither Adams nor Puppe addressed the crucial problem of constructing a smash product. Kan's original papers did not address that problem either, but Kan and Whitehead [KW65a] constructed a smash product of simplicial spectra not much later. They proved that their smash product is commutative, but they did not address its associativity. In [KW65b], they used this product to discuss ring and module spectra and to study degrees of orientability, defined in terms of higher order cohomology operations, but still without addressing the question of associativity. In particular, they defined the notion of a commutative ring spectrum without defining the notion of an associative ring spectrum. Further study of simplicial spectra was made in a series of papers by Burghelea and Deleanu [BD67, BD68, BD69]. While they proved some additional properties of the smash product, they too failed to address the question of its associativity. In fact, as far as I know, that question has never been addressed in the literature.

Although simplicial spectra have not been studied much in recent years, the simplicial approach does lend itself naturally to the study of algebraically defined functors. This was exploited in the papers [KW65a, KW65b] of Kan and Whitehead and in the paper [BCKQRS66] of Bousfield, Curtis, Kan, Quillen, Rector, and Schlesinger. That paper gave a new construction of the Adams spectral sequence in terms of the mod-p lower central series of free simplicial group spectra. For the sphere spectrum, the E_1-term given by their construction is the "Λ-algebra", which is a particularly nice differential graded algebra whose homology is the cohomology of the Steenrod algebra. It would become apparent later that the Λ-algebra is closely related to the Dyer–Lashof algebra of homology operations on infinite loop spaces.

The first satisfactory construction of the stable homotopy category was given by Boardman in 1964 [Bo64]. Although mimeographed notes were made available [Bo65, Bo69], Boardman never published his construction. An exposition was given by Vogt [Vogt70]. Boardman begins with the category $\mathcal{F}$ of based finite CW complexes. He constructs from it the category $\mathcal{F}_s$ of finite CW spectra by a categorical stabilization construction. Its homotopy category $\mathcal{F}_{sh}$ is equivalent to the category obtained from the S-category by adjoining formal desuspensions. As Boardman notes, this is the right category in which to study Spanier–Whitehead duality since here the pesky dimension n in Spanier's definition can be eliminated: a duality between finite CW spectra X and Y is specified by a suitably behaved map $\varepsilon : Y \wedge X \to S$.

Freyd [Fre66] studied the category $\mathcal{F}_{sh}$ categorically. He observed that any additive category $\mathcal{C}$ with cofiber sequences, such as $\mathcal{F}_{sh}$, embeds as a full subcategory of an Abelian category $\mathcal{A}$, namely the evident category whose objects are the morphisms of $\mathcal{C}$. Moreover, $\mathcal{A}$ has enough injective and projective objects, its injective and projective objects coincide, and the objects of $\mathcal{C}$ map to projective objects in $\mathcal{A}$. He observed further that idempotents induce splittings into wedge summands in $\mathcal{C}$ for suitable $\mathcal{C}$, such as $\mathcal{F}_{sh}$, and deduced that $\mathcal{C}$ is then the full subcategory of projective objects of $\mathcal{A}$. Although he was not in possession of $\mathcal{S}_h$, it satisfies the hypotheses he makes on $\mathcal{C}$. Focusing on $\mathcal{F}_{sh}$, he posed a provocative question, "the generating hypothesis", which asserts that a map between finite CW spectra is null homotopic if it induces the zero homomorphism of homotopy groups. Despite much work, it is still unknown whether or not this is true.

Boardman next constructs a category $\mathcal{S} = \mathcal{F}_{sw}$ of CW spectra by a categorical adjunction of colimits construction. Thus his spectra are the colimits of directed systems of inclusions of finite CW spectra. The homotopy category $\mathcal{S}_h$ is the desired stable homotopy category. The most interesting feature of his work is his construction of smash products. He constructs a category $\mathcal{S}(U)$ similarly for each countably infinite dimensional real inner product space, and he constructs an external smash product $\bar{\wedge} : \mathcal{S}(U) \times \mathcal{S}(V) \to \mathcal{S}(U \oplus V)$. He shows that any linear isometry $f : U \to U'$ induces a functor $f_* : \mathcal{S}(U) \to \mathcal{S}(U')$, and he proves that, up to canonical isomorphism, the induced functor $\mathcal{S}_h(U) \to \mathcal{S}_h(U')$ on homotopy categories is independent of the choice of f. An internal smash product on $\mathcal{S} = \mathcal{S}(\mathbb{R}^\infty)$ is a composite $f_* \vee \bar{\wedge} : \mathcal{S} \times \mathcal{S} \to \mathcal{S}$ for any linear isometry $f : \mathbb{R}^\infty \times \mathbb{R}^\infty \to \mathbb{R}^\infty$. Any two such internal smash products become canonically equivalent after passage to homotopy, and this allows the proof that $\mathcal{S}_h$ is symmetric monoidal.

This was very much the hare's approach and it has greatly influenced later hares (such as myself), who have needed vastly more precise properties of a good category of spectra than would have seemed possible in 1966. In particular, for much current work of interest, it is essential to have an underlying symmetric monoidal category of spectra, before passage to homotopy categories. However, perhaps for the benefit of the tortoises, Boardman [Bo69] gave a precise comparison between his construction of $\mathcal{S}_h$ and earlier approaches, and he explained how to modify the approaches of Adams and Puppe to obtain a category equivalent to $\mathcal{S}_h$. He wrote "the complication will show why we do not adopt this as definition". Nevertheless, Adams soon after gave an exposition along these lines [Ad71b] which, in the absence of a published version of Boardman's category, has served until recently as a stopgap reference.

In other parts of our story, definitive foundations were in place by 1966. The axioms for generalized homology and cohomology theories and the understanding of the representation of homology and cohomology theories were firmly established. So were the basics of K-theory and cobordism and much of the basic machinery of computation. Of course, the calculations themselves, once in place, are fixed forever: the answers will not change. The development of the stable category seems now also to have reached such a level of full understanding, and I ask the reader's indulgence in offering the monograph [EKMM97] as evidence.

My arbitrary stopping point of 1966 has the effect both of allowing me to document the invention of a marvelous new area of mathematics and of throwing into high relief how very much has been done since. There are truly vast areas of stable algebraic topology that were barely visible over the horizon or well beneath it in 1966. But that is a story for another occasion.

Bibliography

[Ad58a] J.F. Adams, *The structure and applications of the Steenrod algebra*, Comment. Math. Helv. **32** (1958), 180–214.

[Ad58b] J.F. Adams, *On the non-existence of elements of Hopf invariant one*, Bull. Amer. Math. Soc. **64** (1958), 279–282.

[Ad59] J.F. Adams, *Sur la théoric de l'homotopie stable*, Bull. Soc. Math. France **87** (1959), 277–280.

[Ad60] J.F. Adams, *On the nonexistence of elements of Hopf invariant one*, Ann. of Math. (2) **72** (1960), 20–153.

[Ad61a] J.F. Adams, *On Chern characters and the structure of the unitary group*, Proc. Cambridge Phil. Soc. **57** (1961), 189–199.

[Ad61b] J.F. Adams, *On formulae of Thom and Wu*, Proc. London Math. Soc. (3) **11** (1961), 741–752.

[Ad61c] J.F. Adams, *Stable Homotopy Theory*, Lecture Notes, Berkeley (1961); and Lecture Notes in Math. vol. 3, Springer (1964), (2nd ed. 1966; 3rd ed. 1969).

[Ad62a] J.F. Adams, *Vector fields on spheres*, Topology **1** (1962), 63–65.

[Ad62b] J.F. Adams, *Vector fields on spheres*, Bull. Amer. Math. Soc. **68** (1962), 39–41.

[Ad62c] J.F. Adams, *Vector fields on spheres*, Ann. of Math. (2) **75** (1962), 603–632.

[Ad62d] J.F. Adams, *Applications of the Grothendieck–Atiyah–Hirzebruch functor* $K(X)$, Proc. Internat. Congress of Mathematicians, Stockholm (1962), 435–441; and Inst. Mittag-Leffler, Djursholm (1963); see also *Colloquium on Algebraic Topology*, Aarhus Universitet (1962), 104–113.

[Ad63] J.F. Adams, *On the groups* $J(X)$ *I*, Topology **2** (1963), 181–193.

[Ad65a] J.F. Adams, *On the groups* $J(X)$ *II*, Topology **3** (1965), 137–171.

[Ad65b] J.F. Adams, *On the groups* $J(X)$ *III*, Topology **3** (1965), 193–222.

[Ad66a] J.F. Adams, *On the groups* $J(X)$ *IV*, Topology **5** (1966), 21–77; Correction, Topology **7** (1968), 331.

[Ad66b] J.F. Adams, *A spectral sequence defined using* K*-theory*, Proc. Colloq. de Topologie, Bruxelles (1964), Gauthier-Villars, Paris (1966), 149–166.

[Ad71a] J.F. Adams, *A variant of E.H. Brown's representability theorem*, Topology **10** (1971), 185–198.

[Ad71b] J.F. Adams, *Stable Homotopy and Generalized Cohomology*, Univ. of Chicago Press (1974). (Reprinted 1995.)

[AA66] J.F. Adams and M.F. Atiyah, K*-theory and the Hopf invariant*, Quart. J. Math. Oxford Ser. (2) **17** (1966), 31–38.

[AL72] J.F. Adams and A. Liulevicius, *The Hurewicz homomorphism for MU and BP*, J. London Math. Soc. (2) **5** (1972), 539–545.

[AW64] J.F. Adams and G. Walker, *On complex Stiefel manifolds*, Proc. Cambridge Phil. Soc. **60** (1964), 81–103.

[Adem52] J. Adem, *The iteration of the Steenrod squares in algebraic topology*, Proc. Nat. Acad. Sci. USA **38** (1952), 720–726.

[Adem53] J. Adem, *Relations on iterated reduced powers*, Proc. Nat. Acad. Sci. USA **39** (1953), 636–638.

[Adem56] J. Adem, *A cohomology criterion for determining essential compositions of mappings*, Bol. Soc. Mat. Mexicana (2) **1** (1956), 38–48 (in Spanish).

[Adem57] J. Adem, *The relations on Steenrod powers of cohomology classes. Algebraic geometry and topology*, A symposium in Honor of S. Lefschetz, Princeton Univ. Press (1957), 191–238.

[Adem58] J. Adem, *Second order cohomology operations associated with Steenrod squares*, Symp. Internacional de Topologica Algebraica, Universidad Nacional Autónoma de México and UNESCO, Mexico City (1958), 186–221 (in Spanish).

[Adem59] J. Adem, *On the product formula for cohomology operations of second order*, Bol. Soc. Mat. Mexicana (3) **4** (1959), 42–65.

[An64] D.W. Anderson, Ph.D. thesis, Berkeley (1964).

[ABP66a] D.W. Anderson, E.H. Brown, Jr. and F.P. Peterson, *SU-cobordism, KO-characteristic numbers, and the Kervaire invariant*, Ann. of Math. (2) **83** (1966), 54–57.

[ABP66b] D.W. Anderson, E.H. Brown, Jr. and F.P. Peterson, *Spin cobordism*, Bull. Amer. Math. Soc. **72** (1966), 256–260.

[And66] P.G. Anderson, *Cobordism classes of squares of orientable manifolds*, Ann. of Math. (2) **83** (1966), 47–53.

[AK56] S. Araki and T. Kudo, *Topology of* H_n*-spaces and* H*-squaring operations*, Mem. Fac. Sci. Kyusyu Univ. Ser. A **10** (1956), 85–120.

[AT65] S. Araki and H. Toda, *Multiplicative structures in* mod q *cohomology theories*, Osaka J. Math. **2** (1965), 71–115; and **3** (1966), 81–120.

[At61a] M.F. Atiyah, *Bordism and cobordism*, Proc. Cambridge Phil. Soc. **57** (1961), 200–208.

[At61b] M.F. Atiyah, *Characters and cohomology of finite groups*, Inst. Hautes Études Sci. Publ. Math. **9** (1961), 23–64.

[At61c] M.F. Atiyah, *Thom complexes*, Proc. London Math. Soc. (3) **11** (1961), 291–310.

[At62] M.F. Atiyah, *Vector bundles and the Künneth formula*, Topology **1** (1962), 245–248.

[At64] M.F. Atiyah, K*-Theory*, Lecture Notes, Harvard (1964); and Benjamin, New York (1967).

[At66a] M.F. Atiyah, *Power operations in* K*-theory*, Quart. J. Math. Oxford Ser. (2) **17** (1966), 165–193.

[At66b] M.F. Atiyah, K*-theory and reality*, Quart. J. Math. Oxford Ser. (2) **17** (1966), 367–386.

[AB64] M.F. Atiyah and R. Bott, *On the periodicity theorem for complex vector bundles*, Acta Math. **112** (1964), 229–247.

[ABS64] M.F. Atiyah, R. Bott and A. Shapiro, *Clifford algebras*, Topology **3**, Suppl. 1 (1964), 3–38.

[AH59] M.F. Atiyah and F. Hirzebruch, *Riemann–Roch theorems for differentiable manifolds*, Bull. Amer. Math. Soc. **65** (1959), 276–281.

[AH61a] M.F. Atiyah and F. Hirzebruch, *Vector bundles and homogeneous spaces*, Proc. Sympos. Pure Math. vol. 3, Amer. Math. Soc., Providence, RI (1961), 7–38.

[AH61b] M.F. Atiyah and F. Hirzebruch, *Charakteristische Klassen und Anwendungen*, Enseign. Math. (2) **7** (1961), 188–213.

[AH61c] M.F. Atiyah and F. Hirzebruch, *Cohomologie-Operationen und charakteristische Klassen*, Math. Z. **77** (1961), 149–187.

[AS69] M.F. Atiyah and G.B. Segal, *Equivariant K-theory and completion*, J. Differential Geom. **3** (1969), 1–18.

[AS63] M.F. Atiyah and I.M. Singer, *The index of elliptic operators on compact manifolds*, Bull. Amer. Math. Soc. **69** (1963), 422–433.

[AT60] M.F. Atiyah and J.A. Todd, *On complex Stiefel manifolds*, Proc. Cambridge Phil. Soc. **56** (1960), 343–353.

[Av59] R.G. Averbuh, *Algebraic structure of cobordism groups*, Dokl. Akad. Nauk SSSR **125** (1959), 11–14.

[Ba55] M.G. Barratt, *Track groups*, Proc. London Math. Soc. (3) **5** (1955), 71–106 and 285–329.

[Bass64] H. Bass, *K-theory and stable algebra*, Inst. Hautes Études Sci. Publ. Math. **22** (1964), 5–60.

[BG] J.C. Becker and D.H. Gottlieb, *A history of duality in algebraic topology*, History of Topology, I.M. James, ed., Elsevier, Amsterdam (1999), 725–745.

[BM51] A.L. Blakers and W.S. Massey, *The homotopy groups of a triad I*, Ann. of Math. (2) **53** (1951), 161–205.

[BM52] A.L. Blakers and W.S. Massey, *The homotopy groups of a triad II*, Ann. of Math. (2) **55** (1952), 192–201.

[BM53] A.L. Blakers and W.S. Massey, *The homotopy groups of a triad III*, Ann. of Math. (2) **58** (1953), 409–417.

[Bo64] J.M. Boardman, Ph.D. thesis, Cambridge (1964).

[Bo65] J.M. Boardman, *Stable Homotopy Theory (Summary)*, Mimeographed notes, Warwick University (1965–1966).

[Bo69] J.M. Boardman, *Stable Homotopy Theory*, Mimeographed notes, Johns Hopkins University, Baltimore, MD (1969–1970).

[BP60] V.G. Boltyanskii and M.M. Postnikov, *On the fundamental concepts of algebraic topology. Axiomatic definition of cohomology groups*, Dokl. Akad. Nauk SSSR **133** (1960), 745–747; English translation: Soviet Math. Dokl. **1**, 900–902.

[Bo53a] A. Borel, *Sur la cohomologie des espaces fibrés principaux et des espaces homogènes de groupes de Lie compacts*, Ann. of Math. (2) **57** (1953), 105–207.

[Bo53b] A. Borel, *La cohomologie mod 2 de certains espaces homogènes*, Comment. Math. Helv. **27** (1953), 165–197.

[Bo60] A. Borel, *Seminar on Tranformation Groups*, With contributions by G. Bredon, E.E. Floyd, D. Mongomery and R. Palais, Ann. of Math. Stud. vol. 46, Princeton Univ. Press, Princeton, NJ (1960).

[BH58] A. Borel and F. Hirzebruch, *Characteristic classes and homogeneous spaces I*, Amer. J. Math. **80** (1958), 458–538.

[BH59] A. Borel and F. Hirzebruch, *Characteristic classes and homogeneous spaces II*, Amer. J. Math. **81** (1959), 315–382.

[BH60] A. Borel and F. Hirzebruch, *Characteristic classes and homogeneous spaces III*, Amer. J. Math. **82** (1960), 491–504.

[BS51] A. Borel and J.-P. Serre, *Détermination des p-puissances réduites de Steenrod dans la cohomologie des groupes classiques. Applications*, C. R. Acad. Sci. Paris **233** (1951), 680–682.

[BS53] A. Borel and J.-P. Serre, *Groupes de Lie et puissances réduites de Steenrod*, Amer. J. Math. **75** (1953), 409–448.

[BS58] A. Borel and J.-P. Serre, *Le théorème de Riemann–Roch (d'après Grothendieck)*, Bull. Soc. Math. France **86** (1958), 97–136.

[Bor36] K. Borsuk, *Sur les groupes des classes de transformations continues*, C. R. Acad. Sci. Paris **202** (1936), 1400–1403.

[Bott56] R. Bott, *An application of the Morse theory to the topology of Lie groups*, Bull. Soc. Math. France **84** (1956), 251–281.

[Bott57] R. Bott, *The stable homotopy of the classical groups*, Proc. Nat. Acad. Sci. USA **43** (1957), 933–935.

[Bott58] R. Bott, *The space of loops on a Lie group*, Michigan Math. J. **5** (1958), 35–61.

[Bott59a] R. Bott, *The stable homotopy of the classical groups*, Ann. of Math. (2) **70** (1959), 313–337.

[Bott59b] R. Bott, *Quelques remarques sur les théorèmes de périodicité*, Bull. Soc. Math. France **87** (1959), 293–310.

[Bott61] R. Bott, *Review of Borel and Hirzebruch* ([BH59] above), Math. Reviews **22** (1961), 171–174.

[Bott62] R. Bott, *A note on the KO-theory of sphere-bundles*, Bull. Amer. Math. Soc. **68** (1962), 395–400.

[Bott63] R. Bott, *Lectures on $K(X)$*, Lecture Notes, Harvard (1963); Benjamin, New York (1969).

[BM58] R. Bott and J.W. Milnor, *On the parallelizability of the spheres*, Bull. Amer. Math. Soc. **64** (1958), 87–89.

[BCKQRS66] A. Bousfield, E.B. Curtis, D.M. Kan, D.G. Quillen, D.L. Rector and J.W. Schlesinger, *The mod p-lower central series and the Adams spectral sequence*, Topology **5** (1966), 331–342.

[Br60] W. Browder, *Homology operations and loop spaces*, Illinois J. Math. **4** (1960), 347–357.

[BLP66] W. Browder, A. Liulevicius and F.P. Peterson, *Cobordism theories*, Ann. of Math. (2) **84** (1966), 91–101.

[Br63] E.H. Brown, Jr., *Cohomology theories*, Ann. of Math. (2) **75** (1962), 467–484; Correction: Ann. of Math. (2) **78** (1963), 201.

[Br65] E.H. Brown, Jr., *Abstract homotopy theory*, Trans. Amer. Math. Soc. **119** (1965), 79–85.

[BP66] E.H. Brown, Jr. and F.P. Peterson, *A spectrum whose $\mathbb{Z}_p$ cohomology is the algebra of reduced p-th powers*, Topology **5** (1966), 149–154.

[BD67] D. Burghelea and A. Deleanu, *The homotopy category of spectra I*, Illinois J. Math. **11** (1967), 454–473.

[BD68] D. Burghelea and A. Deleanu, *The homotopy category of spectra II*, Math. Ann. **178** (1968), 131–144.

[BD69] D. Burghelea and A. Deleanu, *The homotopy category of spectra III*, Math. Z. **108** (1969), 154–170.

[Ca55] H. Cartan, *Sur l'itération des opérations de Steenrod*, Comment. Math. Helv. **29** (1955), 40–58.

[CE56] H. Cartan and S. Eilenberg, *Homological Algebra*, Princeton Univ. Press, Princeton, NJ (1956).

[Ca54-55] H. Cartan et al., *Séminaire H. Cartan de l'ENS, 1954–1955. Algèbres d'Eilenberg–MacLane et Homotopie*, Secr. Math. vol. 11, R.P. Curie, Paris (1956).

[Ca59-60] H. Cartan et al., *Séminaire H. Cartan de l'ENS, 1959–1960. Périodicité des Groupes d'Homotopie Stables des Groupes Classiques, d'après Bott*, Secr. Math. vol. 11, R.P. Curie, Paris (1961).

[CS52a] H. Cartan and J.-P. Serre, *Espaces fibrés et groupes d'homotopie I. Constructions générales*, C. R. Acad. Sci. Paris **234** (1952), 288–290.

[CS52b] H. Cartan and J.-P. Serre, *Espaces fibrés et groupes d'homotopie II. Applications*, C. R. Acad. Sci. Paris **234** (1952), 393–395.

[CF64a] P.E. Conner and E.E. Floyd, *Differentiable Periodic Maps*, Ergebnisse der Mathematik und ihrer Grenzgebiete (N.F.), Band 3, Academic Press, New York; Springer, Berlin (1964).

[CF64b] P.E. Conner and E.E. Floyd, *Periodic maps which preserve a complex structure*, Bull. Amer. Math. Soc. **70** (1964), 574–579.

[CF66a] P.E. Conner and E.E. Floyd, *Torsion in SU-bordism*, Mem. Amer. Math. Soc. **60** (1966).

[CF66b] P.E. Conner and E.E. Floyd, *The Relation of Cobordism to K-Theories*, Lecture Notes in Math. vol. 28, Springer, Berlin (1966).

[Dieu] J. Dieudonné, *A History of Algebraic and Differential Topology 1900–1960*, Birkhäuser, Basel (1989).

[Dold56a] A. Dold, *Erzeugende der Thomschen Algebra $\mathcal{N}$*, Math. Z. **65** (1956), 25–35.

[Dold56b] A. Dold, *Vollständingkeit der Wu-schen Relationen zwischen den Stiefel–Whitneyschen Zahlen differenzierbarer Mannigfaltigkeiten*, Math. Z. **65** (1956), 200–206.

[Dold58a] A. Dold, *Démonstration elémentaire de deux résultats du cobordisme*, Ehresmann Seminar Notes, Paris (1958–1959).

[Dold58b] A. Dold, *Homology of symmetry products and other functors of complexes*, Ann. of Math. (2) **68** (1958), 54–80.

[Dold60] A. Dold, *Zur Homotopietheorie der Kettenkomplexe*, Math. Ann. **140** (1960), 278–298.

[Dold62] A. Dold, *Relations between ordinary and extraordinary cohomology*, Colloquium on Algebraic Topology, Aarhus Universitet (1962), 2–9.

[Dold63] A. Dold, *Partitions of unity in the theory of fibrations*, Ann. of Math. **78** (1963), 223–255.

[Dold66] A. Dold, *Halbexakte Homotopiefunktoren*, Lecture Notes in Math. vol. 12, Springer, Berlin (1966).

[DP61] A. Dold and D. Puppe, *Homologie nicht-additiver Funktoren. Anwendungen*, Ann. Inst. Fourier (Grenoble) **11** (1961), 201–312.

[DT58] A. Dold and R. Thom, *Quasifaserungen und unendlische symmetrische Produkte*, Ann. of Math. (2) **67** (1958), 239–281.

[Dyer62] E. Dyer, *Relations between cohomology theories*, Colloquium on Algebraic Topology, Aarhus Universitet (1962), 89–93.

[Dyer63] E. Dyer, *Chern characters of certain complexes*, Math. Z. **80** (1963), 363–373.

[Dyer69] E. Dyer, *Cohomology Theories*, Benjamin, New York (1969).

[DL61] E. Dyer and R.K. Lashof, *A topological proof of the Bott periodicity theorems*, Annali di Math. **54** (1961), 231–254.

[DL62] E. Dyer and R.K. Lashof, *Homology of iterated loop spaces*, Amer. J. Math. **84** (1962), 35–88.

[Eck42] B. Eckmann, *Gruppentheoretischer Beweis des Satzes von Hurwitz–Radon über die Komposition quadratischer Formen*, Comment. Math. Helv. **15** (1942), 358–366.

[Eck57] B. Eckmann, *Homotopie et dualité*, Colloque de Topologie Algèbrique, Louvain (1956), 41–53; George Thone, Liège; Masson, Paris (1957).

[EH58] B. Eckmann and P.J. Hilton, *Groupes d'homotopie et dualité*, C. R. Acad. Sci. Paris **246** (1958), 2444–2447, 2555–2558, 2991–2993.

[Eil40] S. Eilenberg, *Cohomology and continuous maps*, Ann. of Math. (2) (1940), 231–251.

[EM43] S. Eilenberg and S. MacLane, *Relations between homology and homotopy groups*, Proc. Nat. Acad. Sci. USA **29** (1943), 155–158.

[EM45a] S. Eilenberg and S. MacLane, *Relations between homology and homotopy groups of spaces*, Ann. of Math. (2) **46** (1945), 480–509.

[EM45b] S. Eilenberg and S. MacLane, *General theory of natural equivalences*, Trans. Amer. Math. Soc. **58** (1945), 231–294.

[EM50] S. Eilenberg and S. MacLane, *Relations between homology and homotopy groups of spaces II*, Ann. of Math. (2) **51** (1950), 514–533.

[ES45] S. Eilenberg and N.E. Steenrod, *Axiomatic approach to homology theory*, Proc. Nat. Acad. of Sci. USA **31** (1945), 117–120.

[ES52] S. Eilenberg and N.E. Steenrod, *The Foundations of Algebraic Topology*, Princeton Univ. Press, Princeton, NJ (1952).

[EKMM97] A.D. Elmendorf, I. Kriz, M.A. Mandell and J.P. May, *Rings, Modules, and Algebras in Stable Homotopy Theory*, Surveys and Monographs in Math. vol. 47, Amer. Math. Soc., RI (1997).

[Ep66] D.B.A. Epstein, *Steenrod operations in homological algebra*, Invent. Math. **1** (1966), 152–208.

[Est] W.T. van Est, *Hans Freudenthal*, History of Topology, I.M. James, ed., Elsevier, Amsterdam (1999), 1009–1019.

[FS] G. Frei and U. Stammbach, *Heinz Hopf (1894–1971)*, History of Topology, I.M. James, ed., Elsevier, Amsterdam (1999), 991–1008.

[Fr37] H. Freudenthal, *Über die Klassen der Sphärenabildungen I*, Comp. Math. **5** (1937), 299–314.

[Fre66] P. Freyd, *Stable homotopy*, Proc. Conf. Categorical Algebra, La Jolla (1965), Springer, Berlin (1966), 121–172.

[Gr64] P.S. Green, *A cohomology theory based upon self-conjugacies of complex vector bundles*, Bull. Amer. Math. Soc. **70** (1964), 522–524.

[Gro57] A. Grothendieck, *Sur quelques points d'algèbra homologique*, Tôhoku Math. J. (2) **9** (1957), 119–221.

[HW65] A. Haefliger and C.T.C. Wall, *Piecewise linear bundles in the stable range*, Topology **4** (1965), 209–214.

[Ha66] A. Hattori, *Integral characteristic numbers for weakly almost complex manifolds*, Topology **5** (1966), 259–280.

[He55] A. Heller, *Homotopy resolutions of semi-simplicial complexes*, Trans Amer. Math. Soc. **80** (1955), 299–344.

[Hil51] P.J. Hilton, *Suspension theorems and the generalized Hopf invariant*, Proc. London Math. Soc. (3) **1** (1951), 462–493.

[Hil58] P.J. Hilton, *Homotopy theory of modules and duality*, Symp. Internacional de Topologica Algebraica, Universidad Nacional Autónoma de México and UNESCO, Mexico City (1958), 273–281.

[Hil65] P.J. Hilton, *Homotopy Theory and Duality*, Gordon and Breach, New York (1965).

[Hir61] M.W. Hirsch, *On combinatorial submanifolds of differentiable manifolds*, Comment. Math. Helv. **36** (1961), 108–111.

[Hirz53] F. Hirzebruch, *On Steenrod's reduced powers, the index of inertia, and the Todd genus*, Proc. Nat. Acad. Sci. USA **39** (1953), 951–956.

[Hirz56] F. Hirzebruch, *Neue Topologische Methoden in der Algebraischen Geometrie*, Ergebnisse der Mathematik und ihrer Grenzgebiete (N.F.), Heft 9, Springer, Berlin (1956).

[Hirz59] F. Hirzebruch, *A Riemann–Roch theorem for differentiable manifolds*, Séminaire Bourbaki, Exp. 177 (Février 1959).

[Hopf35] H. Hopf, *Über die Abbildungen von Sphären auf Sphären niedriger Dimension*, Fund. Math. **25** (1935), 427–440.

[HsW65] W.C. Hsiang and C.T.C. Wall, *Orientability of manifolds for generalized homology theories*, Trans. Amer. Math. Soc. **118** (1965) 352–359.

[Hu60] S.-T. Hu, *On axiomatic approach to homology theory without using the relative groups*, Portugal. Math. **19** (1960), 211–225.

[Hu61] P.J. Huber, *Homotopical cohomology and Čech cohomology*, Math. Ann. **144** (1961), 73–76.

[Hur35] W. Hurewicz, *Beiträge zur Topologie der Deformationen*, Nederl. Akad. Wetensch. Proc. Ser. A **38** (1935), 112–119, 521–528; **39** (1936), 117–126, 213–224.

[Ja55] I.M. James, *Reduced product spaces*, Ann. of Math. (2) **62** (1955), 170–197.

[Ja56a] I.M. James, *On the suspension triad*, Ann. of Math. (2) **63** (1956), 191–247.

[Ja56b] I.M. James, *On the suspension triad of a sphere*, Ann. of Math. (2) **63** (1956), 407–429.

[Ja57] I.M. James, *On the suspension sequence*, Ann. of Math. (2) **65** (1957), 74–107.

[Ja58a] I.M. James, *The intrinsic join: a study of the homotopy groups of Stiefel manifolds*, Proc. London Math. Soc. (3) **8** (1958), 507–535.

[Ja58b] I.M. James, *Cross-sections of Stiefel manifolds*, Proc. London Math. Soc. (3) **8** (1958), 536–547.

[Ja59] I.M. James, *Spaces associated with Stiefel manifolds*, Proc. London Math. Soc. (3) **9** (1959), 115–140.

[JW58] I.M. James and J.H.C. Whitehead, *Homology with zero coefficients*, Quart. J. Math. Oxford Ser. (2) **9** (1958), 317–320.

[Ka66] T. Kambe, *The structure of K_Λ-rings of the lens space and their applications*, J. Math. Soc. Japan **18** (1966), 135–146.

[KMT66] T. Kambe, H. Matsunaga and H. Toda, *A note on stunted lens spaces*, J. Math. Kyoto Univ. **5** (1966), 143–149.

[Kan55] D.M. Kan, *Abstract homotopy I–IV*, Proc. Nat. Acad. Sci. USA **41** (1955), 1092–1096; **42** (1956), 255–258, 419–421, 542–544.

[Kan58] D.M. Kan, *Adjoint functors*, Trans. Amer. Math. Soc. **87** (1958), 294–329.

[Kan63a] D.M. Kan, *Semisimplicial spectra*, Illinois J. Math. **7** (1963), 463–478.

[Kan63b] D.M. Kan, *On the k-cochains of a spectrum*, Illinois J. Math. **7** (1963), 479–491.

[KW65a] D.M. Kan and G.W. Whitehead, *The reduced join of two spectra*, Topology **3**, Suppl. 2 (1965), 239–261.

[KW65b] D.M. Kan and G.W. Whitehead, *Orientability and Poincaré duality in general homology theories*, Topology **3** (1965), 231–270.

[Kar66] M. Karoubi, *Cohomologie des catégories de Banach*, C. R. Acad. Sci. Paris Sér A/B **263** (1966), A275–A278, A341–A344, and A357–A360.

[Kar68] M. Karoubi, *Algèbres de Clifford et K-théorie*, Ann. Sci. École Norm. Sup. **1** (2) (1968), 1–270.

[Kee51] J.W. Keesee, *On the homotopy axiom*, Ann. of Math. (2) **54** (1951), 247–249.

[Ke59] G.M. Kelly, *Single-space axioms for homology theory*, Proc. Cambridge Phil. Soc. **55** (1959), 10–22.

[Ker58] M.A. Kervaire, *Non-parallelizability of the sphere for $n > 7$*, Proc. Nat. Acad. Sci. USA **44** (1958), 280–283.

[KM60] M.A. Kervaire and J. Milnor, *Bernoulli numbers, homotopy groups, and a theorem of Rohlin*, Proc. Internat. Congress Math. (1958), Cambridge Univ. Press, Cambridge (1960), 454–458.

[KM63] M.A. Kervaire and J. Milnor, *Groups of homotopy spheres I*, Ann. of Math. (2) **77** (1963), 504–537.

[Ki64] J.M. Kister, *Microbundles are fiber bundles*, Ann. of Math. (2) **80** (1964), 190–199.

[KA56] T. Kudo and S. Araki, *Topology of H_n-spaces and H-squaring operations*, Mem. Fac. Sci. Kyusyu Univ. Ser. A **10** (1956), 85–120.

[Lan66] P.S. Landweber, *Künneth formulas for bordism theories*, Trans. Amer. Math. Soc. **121** (1966), 242–256.

[Las63] R.K. Lashof, *Poincaré duality and cobordism*, Trans. Amer. Math. Soc. **109** (1963), 257–277.

[LR65] R.K. Lashof and M. Rothenberg, *Microbundles and smoothing*, Topology **3** (1965), 357–388.

[Le49] J. Leray, *L'homologie filtrée. Topologie algébrique*, Colloques Internat. du Centre National de la Recherche Scientifique vol. 12, Paris (1949), 61–82.

[Lima59] E.I. Lima, *The Spanier–Whitehead duality in new homotopy categories*, Summa Brasil Math. **4** (1959), 91–148.

[Lima60] E.I. Lima, *Stable Postnikov invariants and their duals*, Summa Brasil Math. **4** (1960), 193–251.

[Liu62a] A. Liulevicius, *The factorization of cyclic reduced powers by secondary cohomology operations*, Mem. Amer. Math. Soc. **42** (1962).

[Liu62b] A. Liulevicius, *A proof of Thom's theorem*, Comment. Math. Helv. **37** (1962/1963), 121–131.

[Liu64] A. Liulevicius, *Notes on homotopy of Thom spectra*, Amer. J. Math. **86** (1964), 1–16.

[Mac63] S. MacLane, *Natural associativity and commutativity*, Rice Univ. Studies **49** (4) (1963), 28–46.

[Mac65] S. MacLane, *Categorical algebra*, Bull. Amer. Math. Soc. **71** (1965), 40–106.

[Ma] W.S. Massey, *A history of cohomology theory*, History of Topology, I.M. James, ed., Elsevier, Amsterdam (1999), 579–603.

[May65a] J.P. May, *The cohomology of restricted Lie algebras and of Hopf algebras*, Bull. Amer. Math. Soc. **71** (1965), 372–377.

[May65b] J.P. May, *The cohomology of the Steenrod algebra; stable homotopy groups of spheres*, Bull. Amer. Math. Soc. **71** (1965), 377–380.

[May66] J.P. May, *The cohomology of restricted Lie algebras and of Hopf algebras*, J. Algebra (1966), 123–146.

[May67] J.P. May, *Simplicial Objects in Algebraic Topology*, Van Nostrand (1967); reprinted Univ. of Chicago Press (1982 and 1992).

[May70] J.P. May, *A general Algebraic Approach to Steenrod Operations*, Lecture Notes in Math. vol. 168, Springer, Berlin (1970).

[May1] J.P. May, *Memorial address for J. Frank Adams. Reminiscences on the life and mathematics of J. Frank Adams*, Math. Intelligencer **12** (1990), 40–48.

[May2] J.P. May, *The work of J.F. Adams*, Adams Memorial Sympos. on Algebraic Topology, Vol. 1, London Math. Soc. Lecture Notes vol. 175 (1992), 1–21.

[Mc] J. McCleary, *Spectral sequences*, History of Topology, I.M. James, ed., Elsevier, Amsterdam (1999), 631–663.

[Mil66] R.J. Milgram, *Iterated loop spaces*, Ann. of Math. (2) **84** (1966), 386–403.

[Mil56a] J.W. Milnor, *On manifolds homeomorphic to the 7-sphere*, Ann. of Math. (2) **64** (1956), 399–405.

[Mil56b] J.W. Milnor, *The construction FK*, Mimeographed notes, Princeton (1956).

[Mil58a] J.W. Milnor, *The Steenrod algebra and its dual*, Ann. of Math. (2) **67** (1958), 150–171.

[Mil58b] J.W. Milnor, *On the Whitehead homomorphism J*, Bull. Amer. Math. Soc. **64** (1958), 79–82.

[Mil58c] J.W. Milnor, *Some sequences of a theorem of Bott*, Ann. of Math. **68** (1958), 444–449.

[Mil59] J.W. Milnor, *On spaces having the homotopy type of a CW-complex*, Trans. Amer. Math. Soc. **90** (1959), 272–280.

[Mil60] J.W. Milnor, *On the cobordism ring Ω^* and a complex analogue I*, Amer. J. Math. **82** (1960), 505–521.

[Mil62a] J.W. Milnor, *On axiomatic homology theory*, Pacific J. Math. **12** (1962), 337–341.

[Mil62b] J.W. Milnor, *A survey of cobordism theory*, Enseign. Math. (2) **8** (1962), 16–23.

[Mil64] J.W. Milnor, *Microbundles I*, Topology **3**, Suppl. 1 (1964), 53–80.

[Mil65] J.W. Milnor, *On the Stiefel–Whitney numbers of complex manifolds and of Spin manifolds*, Topology **3** (1965), 223–230.

[MM65] J.W. Milnor and J.C. Moore, *On the structure of Hopf algebras*, Ann. of Math. (2) (1965), 211–264.

[MS60] J.W. Milnor and E. Spanier, *Two remarks on fiber homotopy type*, Pacific J. Math. **10** (1960), 585–590.

[Mo54] J.C. Moore, *On homotopy groups of spaces with a single nonvanishing homology group*, Ann. of Math. (2) **59** (1954), 549–557.

[Mo58] J.C. Moore, *Semi-simplicial complexes and Postnikov systems*, Sympos. Internacional de Topologia Algebraica, Universidad Nacional Autónoma de México and UNESCO, Mexico City (1958), 232–247.

[Na60] M. Nakaoka, *Decomposition theorem for homology groups of symmetric groups*, Ann. of Math. (2) **71** (1960), 16–42.

[Na61] M. Nakaoka, *Homology of the infinite symmetric group*, Ann. of Math. (2) **73** (1961), 229–257.

[Nom60] Y. Nomura, *On mapping sequences*, Nagoya Math. J. **17** (1960), 111–145.

[Nov60] S.P. Novikov, *Some problems in the topology of manifolds connected with the theory of Thom spaces*, Dokl. Akad. Nauk SSSR **132** (1960), 1031–1034; English translation: Soviet Math. Dokl. **1**, 717–720.

[Nov62] S.P. Novikov, *Homotopy properties of Thom complexes*, Mat. Sb. (N.S.) **57**(99) (1962), 407–442 (in Russian).

[Nov65] S.P. Novikov, *New ideas in algebraic topology. K-theory and its applications*, Uspekhi Mat. Nauk **20** no. 3(123) (1965), 41–66; English translation: Russian Math. Surveys **20**(3) (1965), 37–62.

[Nov67] S.P. Novikov, *The methods of algebraic topology from the viewpoint of cobordism theories*, Izv. Akad. Nauk SSSR Ser. Mat. **31** (1967), 855–951 (in Russian); English translation: Math. USSR - Izv (1967), 827–913.

[Pa65] R.S. Palais (with contributions by M.F. Atiyah, A. Borel, E.E. Floyd, R.T. Seeley, W. Shih and R. Solovay), *Seminar on the Atiyah–Singer Index Theorem*, Ann. of Math. Stud. vol. 57, Princeton Univ. Press, Princeton, NJ (1965).

[Pe56a] F.P. Peterson, *Some results on cohomotopy groups*, Amer. J. Math. **78** (1956), 243–258.

[Pe56b] F.P. Peterson, *Generalized cohomotopy groups*, Amer. J. Math. **78** (1956), 259–281.

[Pe57] F.P. Peterson, *Functional cohomology operations*, Trans. Amer. Math. Soc. **86** (1957), 197–211.

[PS59] F.P. Peterson and N. Stein, *Secondary cohomology operations: two formulas*, Amer. J. Math. **81** (1959), 281–305.

[Pon42] L.S. Pontryagin, *Characteristic cycles on manifolds*, C. R. (Doklady) Acad. Sci. URSS (N.S.) **35** (1942), 34–47.

[Pon47] L.S. Pontryagin, *Characteristic cycles on differentiable manifolds*, Mat. Sb. N.S. **21** (63) (1947), 233–284; English translation: Amer. Math. Soc. Translation **32** (1950).

[Pon50] L.S. Pontryagin, *Homotopy classification of the mappings of an $(n+2)$-dimensional sphere on an n-dimensional one*, Dokl. Akad. Nauk SSSR (N.S.) **70** (1950), 957–959.

[Pos51a] M.M. Postnikov, *Determination of the homology groups of a space by means of the homotopy invariants*, Dokl. Akad. Nauk SSSR (N.S.) **76** (1951), 359–362.

[Pos51b] M.M. Postnikov, *On the homotopy type of polyhedra*, Dokl. Akad. Nauk SSSR (N.S.) **76** (1951), 789–791.

[Pos51c] M.M. Postnikov, *On the classification of continuous mappings*, Dokl. Akad. Nauk SSSR (N.S.) **79** (1951), 573–576.

[Pu58] D. Puppe, *Homotopiemengen und ihre induzierten Abbildungen I*, Math. Z. **69** (1958), 299–344.

[Pu62] D. Puppe, *On the formal structure of stable homotopy theory*, Colloquium on Algebraic Topology, Aarhus Universitet (1962), 65–71.

[Pu67] D. Puppe, *Stabile Homotopietheorie I*, Math. Ann. **169** (1967), 243–274.

[Qu67] D.G. Quillen, *Homotopical Algebra*, Lecture Notes in Math. vol. 43, Springer, Berlin (1967).

[Qu69] D.G. Quillen, *On the formal group laws of unoriented and complex cobordism theory*, Bull. Amer. Math. Soc. **75** (1969), 1293–1298.

[Ro51] V.A. Rohlin, *Classification of mappings of S^{n+3} onto S^n*, Dokl. Akad. Nauk SSSR **81** (1951), 19–22 (in Russian).

[Ro52] V.A. Rohlin, *New results in the theory of 4-manifolds*, Dokl. Akad. Nauk SSSR **89** (1952), 221–224 (in Russian).

[Ro53] V.A. Rohlin, *Intrinsic homologies*, Dokl. Akad. Nauk SSSR **89** (1953), 789–792 (in Russian).

[Ro58] V.A. Rohlin, *Internal homologies*, Dokl. Akad. Nauk SSSR **119** (1958), 876–879 (in Russian).

[Seg68] G.B. Segal, *Equivariant K-theory*, Pub. IHES **34** (1968), 129–151.

[Se51] J.-P. Serre, *Homologie singulière des espaces fibrés. Applications*, Ann. of Math. (2) **54** (1951), 425–505.

[Se53a] J.-P. Serre, *Groupes d'homotopie et classes de groupes Abéliens*, Ann. of Math. (2) (1953), 258–294.

[Se53b] J.-P. Serre, *Cohomologie modulo 2 des complexes d'Eilenberg–MacLane*, Comment. Math. Helv. **27** (1953), 198–232.

[Sm62] S. Smale, *On the structure of manifolds*, Amer. J. Math. **84** (1962), 387–399.

[SY61] N. Shimada and T. Yamanoshita, *On triviality of the* mod *p Hopf invariant*, Japan J. Math. **31** (1961), 1–25.

[Sm38] P.A. Smith, *Transformations of finite period*, Ann. of Math. **39** (1938), 127–164.

[Sp50] E. Spanier, *Borsuk's cohomotopy groups*, Ann. of Math. (2) **51** (1950), 203–245.

[Sp56] E. Spanier, *Duality and S-theory*, Bull. Amer. Math. Soc. **62** (1956), 194–203.

[Sp58] E. Spanier, *Duality and the suspension category*, Sympos. Internacional de Topologica Algebraica, Universidad Nacional Autónoma de México and UNESCO, Mexico City (1958), 259–272.

[Sp59a] E. Spanier, *Infinite symmetric products, function spaces, and duality*, Ann. of Math. (2) **69** (1959), 142–198; erratum, 733.

[Sp59b] E. Spanier, *Function spaces and duality*, Ann. of Math. (2) **70** (1959), 338–378.

[SW53] E. Spanier and J.H.C. Whitehead, *A first approximation to homotopy theory*, Proc. Nat. Acad. Sci. USA **39** (1953), 655–660.

[SW55] E. Spanier and J.H.C. Whitehead, *Duality in homotopy theory*, Mathematika **2** (1955), 56–80.

[SW57] E. Spanier and J.H.C. Whitehead, *The theory of carriers and S-theory. Algebraic geometry and topology*, A Sympos. in Honor of S. Lefschetz, Princeton Univ. Press, Princeton, NJ (1957), 330–360.

[SW58] E. Spanier and J.H.C. Whitehead, *Duality in relative homotopy theory*, Ann. of Math. (2) **67** (1958), 203–238.

[Sta63a] J.D. Stasheff, *Homotopy associativity of H-spaces I*, Trans. Amer. Math. Soc. **108** (1963), 275–292.

[Sta63b] J.D. Stasheff, *Homotopy associativity of H-spaces II*, Trans. Amer. Math. Soc. **108** (1963), 293–312.

[Sta63c] J.D. Stasheff, *A classification theorem for fibre spaces*, Topology **2** (1963), 239–246.

[St47] N.E. Steenrod, *Products of cycles and extensions of mappings*, Ann. of Math. **48** (1947), 290–320.

[St49] N.E. Steenrod, *Cohomology invariants of mappings*, Ann. of Math. (2) **50** (1949), 954–988.

[St51] N.E. Steenrod, *The Topology of Fibre Bundles*, Princeton Univ. Press, Princeton, NJ (1951).

[St52] N.E. Steenrod, *Reduced powers of cohomology classes*, Ann. of Math. **56** (1952), 47–67.

[St53a] N.E. Steenrod, *Homology groups of symmetric groups and reduced power operations*, Proc. Nat. Acad. Sci. **39** (1953), 213–217.

[St53b] N.E. Steenrod, *Cyclic reduced powers of cohomology classes*, Proc. Nat. Acad. Sci. **39** (1953), 217–223.

[St57] N.E. Steenrod, *Cohomology operations derived from the symmetric group*, Comment. Math. Helv. **31** (1957), 195–218.

[SE62] N.E. Steenrod, *Cohomology Operations*, Lectures by N.E. Steenrod written and revised by D.B.A. Epstein, Ann. of Math. Stud. vol. 50, Princeton Univ. Press, Princeton, NJ (1962).

[StMR] N.E. Steenrod, *Reviews of Papers in Algebraic and Differential Topology, Topological Groups, and Homological Algebra*, Amer. Math. Soc., Providence, RI (1968).

[ST57] N.E. Steenrod and E. Thomas, *Cohomology operations derived from cyclic groups*, Comment. Math. Helv. **32** (1957), 129–152.

[Sto63] R.E. Stong, *Determination of* $H^*(BO(k,\dots,\infty),\mathbb{Z}_2)$ *and* $H^*(BU(k,\dots,\infty),\mathbb{Z}_2)$, Trans. Amer. Math. Soc. **107** (1963), 526–544.

[Sto65] R.E. Stong, *Relations among characteristic numbers I*, Topology **4** (1965), 267–281.

[Sto66a] R.E. Stong, *Relations among characteristic numbers II*, Topology **5** (1966), 133–148.

[Sto66b] R.E. Stong, *On the squares of oriented manifolds*, Proc. Amer. Math. Soc. **17** (1966), 706–708.

[Sto68] R.E. Stong, *Notes on Cobordism Theory*, Princeton Univ. Press, Princeton, NJ (1968).

[Su57a] M. Sugawara, *On a condition that a space is an H-space*, Math. J. Okayama Univ. **6** (1957), 109–129.

[Su57b] M. Sugawara, *A condition that a space is group-like*, Math. J. Okayama Univ. **7** (1957), 123–149.

[Su60] M. Sugawara, *On the homotopy-commutativity of groups and loop spaces*, Mem. Colloq. Sci. Univ. Kyoto Ser. A Math. **33** (1960/1961), 257–269.

[Sull70] D. Sullivan, *Geometric topology, Part I. Localization, periodicity, and Galois symmetry*, Mimeographed notes (1970).

[Sw60a] R.G. Swan, *A new method in fixed point theory*, Comment. Math. Helv. **34** (1960), 1–16.

[Sw60b] R.G. Swan, *Induced representations and projective modules*, Ann. of Math. (2) **71** (1960), 552–578.

[Sw60c] R.G. Swan, *Periodic resolutions for finite groups*, Ann. of Math. (2) **72** (1960), 267–291.

[Sw62] R.G. Swan, *Vector bundles and projective modules*, Trans. Amer. Math. Soc. **105** (1962), 264–277.

[Swi75] R.M. Switzer, *Algebraic Topology – Homotopy and Homology*, Springer, Berlin (1975).

[Thom50a] R. Thom, *Classes caractéristiques et i-carrés*, C. R. Acad. Sci. Paris **230** (1950), 427–429.

[Thom50b] R. Thom, *Variétés plongées et i-carrés*, C. R. Acad. Sci. Paris **230** (1950), 507–508.

[Thom52] R. Thom, *Espaces fibrés en sphères et carrés de Steenrod*, Ann. Sci. Ecole Norm. Sup. (3) **69** (1952), 109–182.

[Thom53a] R. Thom, *Sous-variétés et classes d'homologie des variétés différentiables. Le théorème général*, C. R. Acad. Sci. Paris **236** (1953), 453–454 and 573–575.

[Thom53b] R. Thom, *Sur un problème de Steenrod. Résultats et applications*, C. R. Acad. Sci. Paris **236** (1953), 1128–1130.

[Thom53c] R. Thom, *Variétés différentiable cobordantes*, C. R. Acad. Sci. Paris **236** (1953), 1733–1735.

[Thom54] R. Thom, *Quelques propriétés globales des variétés différentiables*, Comment. Math. Helv. **28** (1954), 17–86.

[To56] H. Toda, *On the double suspension E^2*, J. Inst. Polytech. Osaka City Univ. Ser. A. **7** (1956), 103–145.

[To58a] H. Toda, *p-primary components of homotopy groups I*, Mem. Colloq. Sci. Univ. Kyoto Ser. A Math. **31** (1958), 129–142.

[To58b] H. Toda, *p-primary components of homotopy groups II*, Mem. Colloq. Sci. Univ. Kyoto Ser. A Math. **31** (1958), 143–160.

[To58c] H. Toda, *p-primary components of homotopy groups III*, Mem. Colloq. Sci. Univ. Kyoto Ser. A Math. **31** (1958), 191–210.

[To59] H. Toda, *p-primary components of homotopy groups IV*, Mem. Colloq. Sci. Univ. Kyoto Ser. A Math. **32** (1959), 297–332.

[To62a] H. Toda, *A topological proof of theorems of Bott and Borel–Hirzebruch for homotopy groups of unitary groups*, Mem. Colloq. Sci. Univ. Kyoto Ser. A Math. **32** (1962), 103–119.

[To62b] H. Toda, *Composition Methods in Homotopy Groups of Spheres*, Ann. of Math. Stud. vol. 49, Princeton Univ. Press, Princeton, NJ (1962).

[To63] H. Toda, *A survey of homotopy theory*, Sûgaku **15** (1963/1964), 141–155 (in Japanese); English translation: Adv. in Math. **10** (1973), 417–455.

[Ver71] J.L. Verdier, *Catégories Dérivées*, Lecture Notes in Math. vol. 569, Springer, Berlin (1971).

[Vogt70] R. Vogt, *Boardman's Stable Homotopy Category*, Lecture Notes Series vol. 71, Aarhus Universitet (1970).

[Wall60] C.T.C. Wall, *Determination of the cobordism ring*, Ann. of Math. (2) **72** (1960), 292–311.

[Wall62] C.T.C. Wall, *A characterization of simple modules over the Steenrod algebra* mod 2, Topology **1** (1962), 249–254.

[Wall65] C.T.C. Wall, *Finiteness conditions for CW-complexes*, Ann. of Math. (2) **81** (1965), 56–69.

[Wall66] C.T.C. Wall, *Addendum to a paper of Conner and Floyd*, Proc. Cambridge Phil. Soc. **62** (1966), 171–175.

[We] C.A. Weibel, *History of homological algebra*, History of Topology, I.M. James, ed., Elsevier, Amsterdam (1999), 797–835.

[Wh42] G.W. Whitehead, *On the homotopy groups of spheres and rotation groups*, Ann. of Math. (2) **43** (1942), 634–640.

[Wh50] G.W. Whitehead, *A generalization of the Hopf invariant*, Ann. of Math. (2) **51** (1950), 192–237.

[Wh53] G.W. Whitehead, *On the Freudenthal theorems*, Ann. of Math. (2) **57** (1953), 209–228.

[Wh56] G.W. Whitehead, *Homotopy groups of joins and unions*, Trans. Amer. Math. Soc. **83** (1956), 55–69.

[Wh60] G.W. Whitehead, *Homology theories and duality*, Proc. Nat. Acad. Sci. USA **46** (1960), 554–556.

[Wh62a] G.W. Whitehead, *Generalized homology theories*, Trans. Amer. Math. Soc. **102** (1962), 227–283.

[Wh62b] G.W. Whitehead, *Some aspects of stable homotopy theory*, Proc. Internat. Congress of Mathematicians, Stockholm (1962), Inst. Mittag-Leffler, Djursholm (1963), 502–506; see also Colloquium on Algebraic Topology, Aarhus Universitet (1962), 94–101.

[Wh1] G.W. Whitehead, *The work of Norman E. Steenrod in algebraic topology: An appreciation*, Lecture Notes in Math. vol. 168, Springer, Berlin (1970), 1–10.

[Wh2] G.W. Whitehead, *Fifty years of homotopy theory*, Bull. Amer. Math. Soc. **8** (1983), 1–29.

[Whi48] J.H.C. Whitehead, *Combinatorial homotopy*, Bull. Amer. Math. Soc. **55** (1948), 213–245, 453–496.

[Whi56] J.H.C. Whitehead, *Duality in topology*, J. London Math. Soc. **31** (1956), 134–148.

[Whit41] H. Whitney, *On the topology of differentiable manifolds*, Lectures in Topology, Univ. of Michigan Press (1941), 101–141.

[Wi66] R.E. Williamson, Jr., *Cobordism of combinatorial manifolds*, Ann. of Math. (2) (1966), 1–33.

[Wood65] R. Wood, *Banach algebras and Bott periodicity*, Topology **4** (1965/1966), 371–389.

[Wu50a] W.-T. Wu, *Classes caractéristiques et i-carrés d'une variété*, C. R. Acad. Sci. Paris **230** (1950), 508–511.

[Wu50b] W.-T. Wu, *Les i-carrés dans une variété Grassmannienne*, C. R. Acad. Sci. Paris **230** (1950), 918–920.

[Wu53] W.-T. Wu, *On squares in Grassmannian manifolds*, Acta Sci. Sinica **2** (1953), 91–115.

This Page Intentionally Left Blank

CHAPTER 25

A History of Duality in Algebraic Topology

James C. Becker and Daniel Henry Gottlieb

Department of Mathematics, Purdue University, West Lafayette, IN 47907-1968, USA
E-mails: becker@math.purdue.edu, gottlieb@math.purdue.edu

1. Introduction

Duality in the general course of human affairs seems to be a juxtaposition of complementary or opposite concepts. This frequently leads to poetical sounding uses of language, both in the common language and in the precision of mathematical theorems. Thus the duality of Projective Geometry: Two points determine a line; two lines determine a point. Gergonne first introduced the word duality in mathematics in 1826. He defined it for Projective Geometry. By the time of Poincaré's note in the Comptes Rendus of 1893, duality was very much in vogue.

There are many dualities in algebraic topology. An informal survey of some topologists has revealed the following names of duality in current use. There are Poincaré, Alexander, Lefschetz, Pontrjagin, Spanier–Whitehead, Hodge, Vogell, Ranici, Whitney, Serre, Eckmann–Hilton, Atiyah, Brown–Comenetz: These are a few whose names reflect those of their discoverers. There are the dual categories, the duality between homology and cohomology and that between homotopy and cohomotopy. There is a duality between cup products and cap products, and between suspension and looping.

The remarkable things are: First, a great many of these seemingly separated dualities are intimately related; and second, those workers who tried to extend or generalize various of these dualities were led to invent widely important notions such as infinite simplicial complexes or spectra; or they discovered remarkable new relationships among important classical concepts. Such a story demands a point of view. Fortunately, one has been provided by A. Dold and D. Puppe with their concept of *strong duality*. So we will jump forward to 1980 and explain their work. Then the remainder of this work will describe in chronographical order the theorems and discoveries which led from Poincaré's first mention of Poincaré duality up to 1980 and the unifying concept of strong duality.

In Section 2, we cover very briefly Eckmann–Hilton duality and then describe Dold and Puppe's approach, which illuminates the main subject of our paper, the development and unification of Poincaré, Alexander, Lefschetz, Spanier–Whitehead, homology–cohomology duality. This material is not as widely known as it should be, so we must give

HISTORY OF TOPOLOGY
Edited by I.M. James

a somewhat technical sketch; but one indicator of the success of their point of view is that we now have a name for all those interrelated dualities: namely, strong duality.

In Section 3 we take up the early days up to 1952. We give a modern statement of the Poincaré–Alexander–Lefschetz duality theorem; and then describe the origins of all the elements of the theorem. Our main sources are J.P. Dieudonné [1989] and William Massey's article in this volume.

Section 4 deals with Spanier–Whitehead duality; Section 5 with Atiyah duality, which clarifies S-duality for manifolds; and Section 6 outlines the story of how Poincaré–Alexander duality was extended to generalized cohomology and homology theories by means of S-duality.

In Section 7 we take up Umkehr maps. Here the Eckmann–Hilton type duality of reversing the direction of maps interfaces with the strong dualities arising from compactness. These Umkehr homomorphisms played a decisive role in the generalization of the Riemann–Roch theorem and the invention of K-theory. A special class of Umkehr maps, the Transfers of Section 8, apparently at first involving only reversing direction in the Eckmann–Hilton manner (and actually first discovered by Eckmann), enjoyed increasing generalization and unifications by means of various duality concepts until at last they inspired Dold and Puppe's categorical picture of strong duality.

With Dold and Puppe we must break off our narrative, as we are too close to recent times. We deeply regret we had neither the time, energy, knowledge, or space to do justice to the many results inspired by or reflecting upon duality. We especially regret the omissions of Eckmann–Hilton duality and Poincaré Duality spaces and surgery. See [Hilton, 1980], [Wall, 1970], respectively.

2. Categorical points of view

There are two major groupings of dualities in algebraic topology: *Strong duality* and *Eckmann–Hilton* duality. Strong duality was first employed by Poincaré [1893] in a note in which "Poincaré duality" was used without proof or formal statement. The various instances of strong duality (Poincaré, Lefschetz, Alexander, Spanier–Whitehead, Pontrjagin, cohomology–homology), seemingly quite different at first, are intimately related in a categorical way which was finally made clear only in 1980. Strong duality depends on finiteness and compactness. On the other hand, Eckmann–Hilton duality is a loose collection of useful dualities which arose from categorical points of view first put forward by Beno Eckmann and P.J. Hilton in [Eckmann, 1956, 1958].

A very good description of how this duality works, and some eyewitness history is given in [Hilton, 1980]. Instead of being a collection of theorems, Eckmann–Hilton duality is a principle for discovering interesting concepts, theorems, and questions. It is based on the dual category, that is, on the duality between the target and source of a morphism; and also on the duality between functors and their adjoints.

In fact it is a method wherein interesting definitions or theorems are given a description in terms of a diagram of maps, or in terms of functors. Then there is a dual way to express the diagram, or perhaps several different dual ways. These lead to new definitions or conjectures. Some, not all, of these definitions turn out to be very fruitful and some of the conjectures turn out to be important theorems.

A very important example of this duality is the notion of cofibration with its cofiber, C_f, written

$$A \xrightarrow{f} B \xrightarrow{j} C_f$$

which is dual to to the notion of fibration with its fibre F, written

$$F \xrightarrow{i} E \xrightarrow{f} B.$$

The word cofibration, and thus the point of view, probably first appeared in [Eckmann, 1958]. (But see [Hilton, 1980], p. 163 for a rival candidate.) Puppe [1958], obtained a sequence of cofibrations in which the third space is the cofiber of the previous map

$$A \xrightarrow{f} B \xrightarrow{j} C_f \to SA \xrightarrow{Sf} SB \to SC_f \to \cdots .$$

This is known as the Puppe sequence.

The dual sequence for fibrations, in which every third space is the fiber of the previous fibration,

$$\cdots \xrightarrow{\Omega i} \Omega E \xrightarrow{\Omega f} \Omega B \xrightarrow{\omega} F \xrightarrow{i} E \xrightarrow{f} B$$

is also frequently called the Puppe sequence, although it was first published by Yasutoshi Nomura [1960].

Eckmann–Hilton duality was conceived as a method based on a categorical point of view in the early 1950's. The challenge was to use the point of view to generate interesting results.

The interrelated dualities of Poincaré, Alexander, Spanier–Whitehead, homology–cohomology, developed haphazardly from 1893 into the 1960's. They formed a collection of quite interesting results which were somehow intimately related. The challenge was to find a coherent categorical description of this phenomenon. It took a long time coming, perhaps because the challenge was not articulated. Then in a paper published in an obscure place with sketchy proofs, Albrecht Dold and Dieter Puppe [1980] laid down a framework for viewing duality. Even now, 17 years later, the name *strong duality* given to this type of duality, is not as current as it should be.

Dold and Puppe conceived their categorical description of strong duality by organizing a seminar to study a much more concrete problem: How exactly are the Becker–Gottlieb transfer and the Dold fixed point transfer related to each other? Using notation which invokes the duality of a number and its reciprocal, they define duality in a symmetric monoidal category $\mathcal{C}$ with multiplication $\otimes$ and a neutral object I. One has, therefore, natural equivalences

$$A \otimes (B \otimes C) \cong (A \otimes B) \otimes C,$$
$$I \otimes A \cong A \cong A \otimes I,$$
$$\gamma : A \otimes B \cong B \otimes A.$$

(Here $\gamma : A \otimes B \to B \otimes A$ is a natural equivalence, whose choice is really part of the definition of the monoidal category C. The explicit noting of γ is an important novel feature of Dold and Puppe's point of view.)

An object DA is the *weak dual* of A if there is a natural bijection with respect to objects X of the sets of morphisms

$$C(X \otimes A, I) \cong C(X, DA).$$

We get a definite morphism $\varepsilon : DA \otimes A \to I$ called the *evaluation*. Now from this another morphism $\delta : A \to DDA$ can be constructed. If δ is an isomorphism (so the dual of the dual is the original object), then DA is called *reflexive*. If in addition $DA \otimes A$ is canonically self dual to itself, then DA is called a *strong dual* of A. Strong duals come equipped with *coevaluations* $\eta : I \to A \otimes DA$.

Now one important example is the category of R-modules for a commutative ring R. The unit I is R, and $DM = \mathrm{Hom}_R(M, R)$ is the weak dual. It is clear that every finitely generated projective module is strongly dualizable.

For stable homotopy, strong duality is Spanier–Whitehead duality, and for the stable category of spaces over a fixed base B, strong duals can be introduced in a fibrewise way.

Now the other half of Dold and Puppe's framework is to inquire how duality is preserved under functors. A *monoidal* functor $T : C \to C'$ between two monoidal categories is a functor together with transformations $TA \otimes TB \to T(A \otimes B)$. The dual D is an example of a contravariant monoidal functor. The homology functor from chain complexes to graded groups gives rise to a monoidal functor.

Unfortunately, $T(A) \otimes T(B) \to T(A \otimes B)$ is not always an isomorphism, as the Künneth theorem shows. So we must say that A is T*-flat* if A is *strongly dualizable* and $T(A)$ is strongly dualizable to $T(DA)$. Subtleties arise here because sometimes we have to examine whether A is T-flat on an object-by-object basis.

Now the point of view given by these considerations is that strong duality takes place in various categories, and the duality theorems are expressed in terms of functors which carry strong duality from the objects of the source category to the objects of the target category. Thus the classical way to think about Poincaré duality, for example, is that it is an isomorphism between homology and cohomology of a closed manifold M, whereas the Dold–Puppe point of view is that a closed manifold and its Spanier–Whitehead dual are carried from the stable homotopy category by the homology functors to self dual graded rings, which means in this case the isomorphism between homology and cohomology comes from the Thom isomorphism and Atiyah duality.

Now strong duality seems to depend upon finiteness or compactness properties. In finite situations, one can define a notion of rank or trace. A particular triumph of the Dold–Puppe point of view is the notion of the *trace* of an endomorphism $f : A \to A$ of a strongly dualizable object.

$$\sigma f : I \xrightarrow{\eta} A \otimes DA \underset{\cong}{\xrightarrow{\gamma}} DA \otimes A \xrightarrow{id \otimes f} DA \otimes A \xrightarrow{\varepsilon} I.$$

In the category of R-modules $\sigma f(1)$ is the usual *trace*. For chain complexes, $\sigma f(1)$ is the Lefschetz number, and for graded abelian R-modules, it is the Lefschetz number. Now if T is a monoidal functor and A is a T-flat object, then $T(\sigma f) = \sigma(Tf)$. For T the homology function, we get Hopf's result that the Lefschetz number of the chain complex is the Lefschetz number for homology. Some formal properties of trace are

$$\sigma(Df) = \sigma f \quad \text{(trace of } f = \text{ trace of } f \text{ transpose)},$$
$$\sigma(f \circ g) = \sigma(g \circ f) \quad \text{(trace preserved by commutation)}.$$

Using these results, Dold and Puppe can prove the Lefschetz fixed point theorem.

3. Poincaré–Alexander–Lefschetz duality

In the text books of Spanier [1966], Dold [1972], Massey [1980], and Bredon [1993] we see, essentially, the final forms of the Poincaré–Alexander–Lefschetz duality theorem.

We choose the description of G. Bredon as our text, and then we shall detail the process by which the ideas necessary for the statement occurred, and the many important byproducts of the effort to understand duality.

THEOREM 3.1. *Let M^n be an n-manifold oriented by ϑ, and let $K \supset L$ be compact subsets of M. Then the cap product*

$$\cap\vartheta : H^p(K, L; G) \to H_{n-p}(M - L, M - K; G)$$

is an isomorphism, and it gives rise to an exact ladder relating the cohomology sequence of the pair (K, L) to the homology sequence of the pair $(M - L, M - K)$.

Henri Poincaré [1893] first mentions Poincaré duality in a note in Comptes Rendus. The goal was to "prove" that an odd-dimensional closed oriented manifold has zero Euler–Poincaré number. Poincaré duality, expressed in terms of Betti numbers, is mentioned as if everyone should know it. The earlier note, [Poincaré, 1892], which is considered the first of Poincaré's papers on topology does not seem to mention duality. See [Bollinger, 1972, Dieudonné, 1989, Henn and Puppe, 1990].

In this note, Poincaré first mentions what we now call Poincaré duality. In the note, Poincaré "shows" that a closed odd dimensional oriented manifold has Euler–Poincaré number equal to zero. He says, "It is known" that the alternating sum of the Betti numbers is the same as the generalized Euler characteristic. On the other hand, he merely states the fact that the Betti numbers in complementary dimensions are equal, which is his version of Poincaré duality, with the same fanfare which one uses to state $1 + 1 = 2$, as if it were widely known. But Poincaré's argument is exactly what one would give today to a student knowledgeable about the Euler characteristic and Poincaré duality.

When Poincaré next refers to his duality, it is in his famous paper Analysis Situs, [Poincaré, 1895]. There he states: "ce thèoréme n'a, je crois, jamais été énoncé; il était cependant connu de Plusieurs Personnes qui en ont même fait des applications", [Bollinger, 1972], see p. 124.

Poincaré elected not to point out any of the several people who knew of it and have even made applications with it, including himself and his note [Poincaré, 1893]. So it is not surprising that the Editors of Poincaré's Oeuvres misplaced the note in Volume 11, instead of placing it in Volume 6 right before Analysis Situs; and that various scholars dealing with Poincaré duality may have, for this reason, omitted mentioning it in their works. According to [Bollinger, 1972], Poincaré concludes Analysis Situs with the very theorem he announced in his 1893 note: odd-dimensional manifolds have zero Euler characteristics. In fact he gives two different proofs of it.

In Analysis Situs, [Poincaré, 1895], "Poincaré endeavored to prove his central theorem on homology, the famous duality theorem...", Dieudonné [1989], formulated in terms of Betti numbers. To do this, he invented the concept of intersection numbers, which were finally made precise in the work of Lefschetz [1926] and of De Rham [1931].

Now P. Heergard, in his dissertation in Danish in 1898, came up with a counterexample to Poincaré duality: A three-dimensional manifold whose Betti numbers were $b_1 = 2$ and $b_2 = 1$!

When Poincaré examined Heergard's paper, he found that his notion of Betti numbers was not the same as that of Betti's, contrary to his belief. And worse, his "proof" worked just as well with this contradictory version. So, Poincaré came up with a new proof in two complements to Analysis Situs in 1899 and 1900 (see [Poincaré, 1895], pp. 290–390).

In this proof, Poincaré assumes his manifold is triangulated, finds an algorithm for calculating the Betti numbers, defines barycentric subdivision of the triangulation, shows that the Betti numbers do not change under subdivision, and defines the dual triangulation. He shows the dual triangulations have complementary Betti numbers, since the incidence matrices he uses to compute the Betti numbers are transposes of the incidence matrices in the complementary dimensions.

The incidence matrix algorithm also gave rise to torsion coefficients, which Poincaré introduced. Poincaré extended the duality theorem to the torsion coefficients. He did not use the fact that invariant factors of a transposed matrix are the same as those of the original matrix. Instead he showed the torsion coefficients were related by inventing the construction of the join.

Thus, at 1900 we have Theorem 3.1 for the case $K = M$ and $L = \emptyset$ where M is a closed manifold (but with a finite triangulation structure attached), no homology groups and no cohomology. The concept of compactness was not made explicit until 1906 by Frechet. Frechet also introduced the abstract notion of metric space. In 1914, Hausdorff gave four axioms for neighborhoods, so topological spaces became a well-defined concept, and the standards of rigor became much higher in topology.

Another thread of duality begins with the duality of inside vs. outside. The intuitively obvious but difficult Jordan curve theorem, that a closed curve in a plane divides the plane into two parts was generalized by L.E.J. Brouwer [1976] (pp. 489–494) for subsets of $\mathbb{R}^n$ homeomorphic to S^{n-1}.

J.W. Alexander [1922] attacked the Jordan–Brouwer separation theorem. He gave a new proof and generalized the result to what we call Alexander duality: $\dim \widetilde{H}_p(X; \mathbb{Z}_2) = \dim \widetilde{H}_{n-p-1}(S^n - X; \mathbb{Z}_2)$ where X is a subcomplex of S^n. In this work, Alexander introduced coefficient groups $\mathbb{Z}_2$, and considered the homology of nonfinite complexes. This paper became the starting point of investigations of homology for more general spaces than merely finite complexes or open subsets of $\mathbb{R}^n$. On the other hand, it led Lefschetz [1926] to introduce the idea of relative homology, that is the homology of K mod L, which he wrote as $H_p(K, L)$ in his book [Lefschetz, 1930]. If K and $K - L$ are orientable combinatorial manifolds, Lefschetz proved that $b_p(K^*) = b_{n-p}(K, L)$ where K^* is the "complement" of L in K defined by dual cells which do not intersect L. Also he had the torsion coefficient relations, for indeed he showed the relevant incidence matrices were transposes of one another. By an argument which essentially is part of the exact sequence of the pair (S^n, L), Lefschetz proved Alexander duality.

Two other events occurred during the 1920's which improved the duality theorem. The first event was the proof of the invariance of homology for different triangulations. J.W. Alexander [1915] and [1926] showed that homology was independent of the triangulation. But his first proof had difficulties, so a satisfactory proof must be dated around 1926. Thus Poincaré duality held for combinatorial manifolds, instead of merely manifolds with a specific triangulation. The question of which manifolds can be triangulated then becomes important in order to assess the domain of applicability of Poincaré duality. S. Cairns [1930] showed that C^1 manifolds were triangulable. Much later, it was discovered that there are some topological manifolds which cannot be triangulated.

The second event was the description of homology in terms of group theory. The homology group $H_*(X)$ was the cycles modulo the boundaries. This was factored into the free part, called the Betti group, and the torsion group. Thus $H_i \cong F_i \oplus T_i$, and Poincaré duality was expressed as $F_i \cong F_{n-i}$ and $T_i \cong T_{n-i-1}$. The Betti group persisted into the 1950's as a commonly used concept. We now know that the Betti groups were not really natural, that is the splitting of homology into free and torsion parts is not functorial, and the duality with cohomology demands homology groups instead of Betti and torsion groups. But in the absence of the concept of cohomology, there is no better way to express Poincaré duality than by means of isomorphisms of suitable Betti groups and torsion groups.

One can see this change clearly by comparing the two books [Lefschetz, 1930] and [Lefschetz, 1942]. In "Topology", [Lefschetz, 1930], the absolute Poincaré duality theorem appears on p. 140 stated solely in terms of Betti numbers and torsion coefficients. On p. 203 of "Algebraic Topology", [Lefschetz, 1942], the theorem is headlined as the "Duality Theorem of Poincaré" and is stated in terms of Betti groups and torsion groups. Above it, on the same page is the untitled result that cohomology is isomorphic to homology in complementary dimensions.

In the 1930's, seemingly different kinds of dualities were studied which led to cohomology. In the first section of his history of cohomology in this volume, William Massey discusses "The struggle to find more general and natural statements of the duality theorems of Poincaré and Alexander" [Massey, 1999].

We will report very briefly on this story and, since Massey has told it so well, we will only mention the main points. We particularly want to note that Massey argues that Pontrjagin developed his duality between a discrete abelian group G and its compact group of continuous characters $\widehat{G}$ to study Poincaré and Alexander duality, Pontrjagin [1934]. Here Poincaré duality can be stated: $\widetilde{H}_k(M; G)$ is Pontrjagin dual to $\widetilde{H}_{n-k}(M; \widehat{G})$. And Alexander duality can be stated: $\widetilde{H}_k(X; G)$ is Pontrjagin dual to $\widetilde{H}_{n-k-1}(S^n - X; \widehat{G})$.

At the Moscow conference of 1935 both Kolmogoroff and Alexander announced the definition of cohomology, which they had discovered independently of one another. Both authors quickly published papers in which they both point out that for any finite complex K and any compact abelian group G, the homology group $H_r(K; G)$ and the cohomology group $H^r(K; \widehat{G})$ are Pontrjagin dual to one another. Kolmogoroff explicitly notes the duality theorems in terms of cohomology being isomorphic to homology.

Alexander and Kolmogoroff suggested the possibility of a product structure in cohomology. E. Čech [1936] and H. Whitney [1937] and in [1938] provided the correct details. Čech made precise the cup product and proved its basic properties. He also defined the cap product. Using the cap product, Čech proves the Poincaré duality theorem for closed, oriented, combinatorial manifolds.

The fact that the cap product with the fundamental class of a manifold explicitly gives the duality isomorphism adds concreteness to Poincaré duality and has played a very important role in applications. For example, on the chain level, the cap product with the fundamental cycle gives a chain homotopy equivalence. This plays an important role in the study of Poincaré spaces. Or again, W.V.D. Hodge [1941] defined the Hodge duality between a p form α and an $n - p$ form $*\alpha$ on $\mathbb{R}^n$. In modern terms, de Rham cohomology, defined by the differential forms on a smooth oriented closed manifold M is isomorphic to real cohomology, and the Hodge dual $*$ can be regarded as the Poincaré duality isomorphism. Hodge duality shows up in the modern formulation of Maxwell's equations:

$$dF = 0, \qquad *d*F = j.$$

H. Whitney covered even more ground than Čech. He defined the induced homomorphism f^* of a map f and gave its relations with $\cup$ and $\cap$, introducing these symbols and the names cup and cap product.

The Čech cohomology which appears on the left in our Theorem 3.1 was first defined by Steenrod [1936] in his thesis. It was, of course, the dual of Čech homology, and it was defined only for compact spaces. Dowker [1937] published a brief announcement for Čech cohomology defined for arbitrary spaces. Meanwhile, Alexander–Spanier homology was being developed by Alexander and Kolmogoroff and was modified and perfected by Spanier [1948]. Then Hurewicz, Dugundji and Dowker [1948] showed that Čech cohomology and Alexander–Spanier cohomology were equivalent for a general class of spaces.

The history of singular homology theory actually begins way back in 1915 with Alexander [1915] in his first attempt to prove the topological invariance of topology. There were several attempts to implement the idea. But it was not until Eilenberg [1944] gave the correct definition that singular homology theory was satisfactorily defined.

Both singular homology and Čech–Alexander–Spanier cohomology are valid for pairs of topological spaces. So by 1948 we have the homology and cohomology groups used in Theorem 3.1. Now the orientation class depends on the fact that $H_*(M, M - x) \simeq \widetilde{H}_*(S^n)$. In 1947, Henri Cartan realized that Sheaf theory provided a mechanism to localize orientation. In the Séminaires Henri Cartan of 1950–1951, he defined the orientation sheaf in the context of a generalized cochain complex of sheaves, and with this he could prove Poincaré and Alexander duality for C^0-manifolds (see [Dieudonné, 1989], p. 211).

The sheaf theory language of the proof was eliminated in mimeographed notes of Milnor in 1964. But there are versions of Poincaré and Alexander duality for homology and cohomology of sheaves, for local coefficients, for cohomology with compact supports; all of which play very important roles in various branches of mathematics besides topology.

Finally we discuss the book of Eilenberg–Steenrod [1952], which does not mention Poincaré duality, and yet it sets in motion several ideas which improve the theorem. In [Eilenberg and Steenrod, 1945], the axioms for homology and cohomology theory were published. The proofs were deferred to the book. With the axioms, we see that for (finite) CW complexes at least, the particular versions of homology and cohomology do not really matter in the statement of Theorem 3.1. Also, the axioms gave rise in the 1950's to the concept of generalized homology and cohomology, in which duality plays an even more important role.

It is clear that in their note of [1945], Eilenberg and Steenrod did not have all the details written down, because they stated that Čech homology satisfied the axioms. This is false and was rectified in their book. But the fact that Čech homology did not satisfy the axioms led to the devaluation of Čech homology, and one does not see it used today in algebraic topology. It is ironic that Čech's name is given to a cohomology theory he did not define, yet the fact that Čech first realized that the Poincaré duality isomorphism could be expressed by the cap product has all but been forgotten.

Eilenberg and Steenrod's book [1952] effected a revolution in mathematical notation. Perhaps not since Descartes' La géométrie has a book influenced how we write Mathematics. One knew they were looking at mathematics before 1600 because of the geometric diagrams with vertices and sides labeled by alphabetic letters. La géométrie in 1637 gave us nearly modern forms of equations, especially the notation of the exponent, i.e. a^3. The diagrams of Eilenberg–Steenrod not only made algebraic topology intelligible, but eventually swept out to other parts of mathematics, providing an efficient way to express complex, functorial relationships and giving us powerful methods of proofs by means of diagram chasing.

So, at last, we can talk about that quintessential diagram, the exact ladder, in the last part of Theorem 3.1, its proof and utility depending upon the five-lemma of Eilenberg–Steenrod.

4. Spanier–Whitehead duality

In 1936, K. Borsuk [1936] showed that under certain conditions the set of homotopy classes of maps from a space X to a space Y, denoted by $[X; Y]$, could be given a natural abelian group structure. About a dozen years later E. Spanier [1949] returned to this idea and made a thorough investigation of these groups when Y is a sphere. He denoted $[X; S^n]$ by $\pi^n(X)$ and called it the n-th cohomotopy group of X. The group is defined when the dimension of X is less than $2n - 1$. Spanier showed that the cohomotopy groups satisfied the Eilenberg–Steenrod axioms for cohomology to the extent that they could be formulated given that the groups are not defined for all n. He then went on to give a group-theoretic formulation of Hopf's classification of maps of an n-complex into S^n and Steenrod's classification of maps of an $(n + 1)$-complex into S^n, noting in the introduction an apparent "duality" between Hopf's theorem for cohomotopy groups and Hurewicz's theorem for homotopy groups.

In order to bypass the difficulty that $[X; Y]$ is not always an abelian group, Spanier and J.H.C. Whitehead [1953, 1957] defined what they called the suspension category, giving birth to what is now called stable homotopy theory. They defined by means of suspension the S-group

$$\{X; Y\} = \lim \left[S^k X; S^k Y\right].$$

They generalized Freudenthal's suspension theorem as follows: If Y is $(n-1)$-connected, the suspension map from $[X; Y]$ to $[SX; SY]$ is bijective if $\dim(X) < 2n-1$, and surjective if $\dim(X) = 2n - 1$. Consequently, when the abelian group structure on $[X; Y]$ is defined, the natural inclusion from $[X; Y]$ to $\{X; Y\}$ is an isomorphism.

They also established exactness and excision properties for the S-groups. Defining $\{X; Y\}_q$ to be $\{S^q X; Y\}$ if $q > 0$, and $\{X; S^{-q} Y\}$ if $q < 0$, they pointed out that for fixed

X the Eilenberg–Steenrod homology axioms are satisfied whereas for fixed Y the cohomology axioms are satisfied (except of course for the dimension axiom). As with Spanier's paper on cohomotopy, their interest in these formal properties reflected the profound influence that the Eilenberg–Steenrod axiomatic approach was having upon the subject.

Shortly after introducing the S-category, Spanier and Whitehead [1955] developed their duality theory. Given a polyhedron X in S^n, an *n-dual* $D_n X$ of X is a polyhedron in $S^n - X$ which is an S-deformation retract of $S^n - X$ (i.e. some suspension of $D_n X$ is a deformation retract of the corresponding suspension of $S^n - X$). They defined for polyhedra X and Y in S^n a duality map

$$D_n : \{X; Y\} \to \{D_n Y; D_n X\}.$$

To do this they first consider the case where there are inclusions $i : X \to Y$ and $i' : D_n Y \to D_n X$. Then $D_n(\{i\})$ is defined to be $\{i'\}$. For a general S-map from X to Y they reduce to the case of an inclusion by means of the mapping cylinder construction. Eventually they show that this leads to a well defined isomorphism. They establish a number of basic properties of the duality map D_n including its relation with Alexander duality. In particular, they make precise the duality between the Hopf and Hurewicz maps which Spanier had noted earlier.

In 1959, Spanier [1959] gave a new treatment of Spanier–Whitehead duality in which he shifted attention from the concept of a dual space to that of a duality map. The way in which this approach came about appears to be as follows: Let $F(X; S^n)$ denote the space of maps from X to S^n, $\omega : X \wedge F(X; S^n) \to S^n$ the evaluation map, and $\gamma \in H^n(S^n)$ a generator. John Moore [1956] had shown that slant product defines an isomorphism $\omega^*(\gamma)/\_ : H_q(F(X; S^n) \to H^{n-q}(X)$, $q < 2(n - \dim(X))$. So the function space $F(X; S^n)$ appears to be $(n+1)$-dual to X at least through a range of dimensions. In order to remove the dimensional restriction, Spanier formed the spectrum $\mathbf{F}(X)$ whose n-th space is $F(X; S^n)$. The connecting maps $h : SF(X; S^n) \to F(X; S^{n+1})$ are given by $h(t, f)(x) = (t, f(x))$. He called $\mathbf{F}(X)$ the *functional dual* of X. Spectra had been introduced earlier by E. Lima [1959], a student of Spanier, in order to study duality for infinite complexes. A spectrum $\mathbf{Y} = \{Y_n; \varepsilon_n\}$ is simply a sequence of spaces Y_n, $n \in \mathbb{Z}$, together with maps $\varepsilon_n : SY_n \to Y_{n+1}$. If X is a finite complex and $\mathbf{Y}$ is a spectrum,

$$\{X; \mathbf{Y}\} = \lim \left[S^n X; Y_n\right].$$

Let X^* be a deformation retract of $S^{n+1} - X$, hence an $(n+1)$-dual of X. To make precise the duality between X and $\mathbf{F}(X)$, Spanier wished to construct a weak equivalence of spectra $X^* \to S^n\mathbf{F}(X)$, where $S^n\mathbf{F}(X)$ is the n-fold suspension of $\mathbf{F}(X)$. It was apparently well known by this time that Alexander duality could be described by means of a slant product. Specifically, by removing a point of S^{n+1} which is not in X or X^*, one can regard X and X^* as subspaces of R^{n+1}. Define $\mu : X \times X^* \to S^n$ by $\mu(x, x^*) = (x - x^*)/|x - x^*|$. Its restriction to $X \vee X^*$ is null homotopic so it induces a map $\mu : X \wedge X^* \to S^n$. Then $\mu^*(\gamma)/\_ : H_q(X) \to H^{n-q}(X^*)$ realizes the Alexander duality isomorphism. Spanier had used this description in an earlier paper [1959a] in which he studied the relation between infinite symmetric products and duality. Now the map $\mu : X \wedge X^* \to S^n$ defines a map $X^* \to F(X; S^n)$, and by virtue of the naturality of the slant product, the induced map of spectra $X^* \to S^n\mathbf{F}(X)$ is a weak equivalence.

Combining this weak equivalence with the exponential correspondence

$$[Z; F(X^*; S^n)] = [Z \wedge X^*; S^n],$$

he showed that the map

$$R_\mu : \{Z; X\} \to \{Z \wedge X^*; S^n\}, \quad f \to \mu(f \wedge 1), \tag{1}$$

is an isomorphism. By symmetry,

$$L_\mu : \{Z; X^*\} \to \{X \wedge Z; S^n\}, \quad f \to \mu(1 \wedge f), \tag{2}$$

is an isomorphism.

Now, if Y is also a polyhedron in S^{n+1}, Y^* is a deformation retract of $S^{n+1} - Y$, and $\nu : Y^* \wedge Y \to S^n$ is the associated map, a duality isomorphism

$$D(\mu, \nu) : \{X; Y\} \to \{Y^*; X^*\} \tag{3}$$

is defined in terms of the fundamental isomorphisms (1) and (2), by

$$\{X; Y\} \xrightarrow{R_\nu} \{X \wedge Y^*; S^n\} \xrightarrow{L_\mu^{-1}} \{Y^*; X^*\}.$$

The existence of the isomorphism R_μ of course does not depend on the geometric origins of X^* but only on the existence of the map μ. Thus, Spanier was led to define a *duality map* to be a map $\mu : X \wedge X^* \to S^n$ such that the slant product $\mu * (\gamma)/\_\_ : H_q(X) \to H^{n-q}(X^*)$ is an isomorphism. He then showed that the map $\tilde{\mu} : X^* \wedge X \to S^n$ obtained by composing μ with the interchange map $X^* \wedge X \to X \wedge X^*$ is also a duality map, from which it follows that L_μ is also an isomorphism. Given a second duality map $\nu : Y \wedge Y^* \to S^n$, he derived the duality isomorphism $D(\mu, \nu)$ as we have described above. In addition to being more general, the formal properties of the duality are much more readily established. Moreover, the theory gives a simple criterion for S-maps $f : X \to Y$ and $g : Y^* \to X^*$ to be dual. They are dual if and only if the diagram

$$\begin{array}{ccc} X \wedge Y^* & \xrightarrow{f \wedge 1} & Y \wedge Y^* \\ \downarrow{\scriptstyle 1 \wedge g} & & \downarrow{\scriptstyle \nu} \\ X \wedge X^* & \xrightarrow{\mu} & S^n \end{array}$$

is homotopy commutative. (A comparison of the two approaches to duality revealed a minor notational problem: A geometric $(n + 1)$-dual X^* gives rise to an n-duality map $X \wedge X^* \to S^n$. Spanier suggested that it would be more natural to call X^* an n-dual of X, which is the terminology that is now used.)

A few years later, Wall [1967] added an additional refinement which would prove useful in applications; particularly to surgery theory. He noted that the whole theory could be given a "dual" formulation. In this description, an *n-duality* is a map $\mu : S^n \to X \wedge X^*$

such that $\mu_*(\gamma)\backslash\_ : H^q(X) \to H_{n-q}(X^*)$ is an isomorphism, where γ is a generator of $H_n(S^n)$.

When Spanier's textbook on algebraic topology appeared in 1966, it contained an exercise outlining a categorical formulation of Spanier–Whitehead duality which he attributed to P. Freyd and D. Husemoller. The fundamental isomorphisms (1) and (2) are easily generalized to isomorphisms

$$R_\mu : \{Z; E \wedge X\} \to \{Z \wedge X^*; E \wedge S^n\}, \tag{4}$$
$$L_\mu : \{Z; X^* \wedge E\} \to \{X \wedge Z; S^n \wedge E\}, \tag{5}$$

where E is an arbitrary CW-complex. In the categorical formulation one now defines μ to be a duality map if R_μ and L_μ are isomorphisms. (By a standard argument it is only necessary to assume that R_μ and L_μ are isomorphisms when Z and E are spheres.) One then shows by an induction over cells argument that for every finite CW-complex X there is an integer n and a finite CW-complex X^* for which there is a duality map $\mu : X \wedge X^* \to S^n$.

This formulation exhibits Spanier–Whitehead duality as an intrinsic property of the S-category quite independent of Alexander's duality theorem. The latter now emerges as the fundamental connection between this duality and geometry: If X is contained in S^{n+1} and X^* is a deformation retract of $S^{n+1} - X$ then X^* is a Spanier–Whitehead n-dual of X.

As generalizations of the S-category arose, the appropriate formulation of Spanier–Whitehead duality soon followed. In 1970 at the International Congress in Nice, G. Segal [1970] introduced for a finite group G, the equivariant S-category whose objects are finite CW-complexes and whose morphisms are

$$\{X; Y\}_G = \lim\left[S^V X; S^V Y\right]_G,$$

the limit taken over all representations V of G. This marked the beginning of equivariant stable homotopy theory – a theory which has undergone rapid development in recent years. Duality was extended to this category by Wirthmüller [1970] and Dold and Puppe [1980].

A second generalization of S-theory involves the consideration of families of pointed spaces parametrized by a fixed space B. The homotopy theory of such spaces was developed by I.M. James [1971], T. tom Dieck, K. Kamps and D. Puppe [1970], among others. There is the corresponding S-category based on fiberwise suspension, and duality in this category was derived in [Becker and Gottlieb, 1976].

5. Atiyah's duality theorem

Milnor and Spanier [1960] clarified the relation between Spanier–Whitehead duality and Poincaré duality on a closed differentiable manifold. Thom [1952] had introduced what is now called the Thom space M^α of a vector bundle α over M. It is defined to be $D(\alpha)/S(\alpha)$ where $D(\alpha)$, $S(\alpha)$ are, respectively, the unit disk and sphere bundles of α. The Milnor–Spanier theorem states that if M is a closed manifold embedded into Euclidean space R^s with normal bundle ν and M^+ is M disjoint union with a point then M^ν is s-dual to M^+. Their proof was geometric; exhibiting M^ν as a deformation retract of the complement of

M^+ in S^{s+1}. With the dual formulation of Spanier–Whitehead duality, which came later, a duality map

$$\mu : S^s \to M^+ \wedge M^\nu$$

is easily constructed. There is the Pontrjagin–Thom map $c : S^s \to M^\nu$ defined by embedding the normal disk bundle $D(\nu)$ into R^s as a tubular neighborhood and letting c collapse the complement of the interior of $D(\nu)$ to a point. Then μ is the composition of c with the "diagonal" map $M^\nu \to M^+ \wedge M^\nu$, $\vec{v}_x \mapsto x \wedge \vec{v}_x$. Lefschetz duality for $(S^s, D(\nu))$ implies that μ is a duality map. There is the commutativity relation

$$\begin{array}{ccc} H^{k+s-n}(M^n) & \xrightarrow{D} & H_{n-k}(M^+) \\ {\scriptstyle \phi_\mu} \uparrow & & \uparrow {\scriptstyle v \cap \_} \\ H^k(M^+) & = & H^k(M^+) \end{array}$$

where $D = \mu_*(\gamma)/\_$, $u \in H^{s-n}(M^n)$, ϕ_u is the Thom homomorphism, and $v = D(u)$. Since D is an isomorphism, $v \cap \_$ is an isomorphism precisely when ϕ_u is an isomorphism. This basic relation, which goes back to Thom, carries over to generalized homology–cohomology theories, and the Milnor–Spanier theorem implies that D remains an isomorphism. Thus, the question of the orientability of a manifold M is equivalent to that of the orientability of its normal bundle in the sense of Thom. The latter is usually studied as part of the general question of orientability of a vector bundle or spherical fibration with respect to a cohomology theory.

Atiyah [1961a] generalized the Milnor–Spanier theorem to manifolds with boundary, and derived from it the following relation among Thom spaces: Let α and β be vector bundles over a closed manifold M such that $\alpha \oplus \beta \simeq M \times R^t$. Then if M is embedded in R^s with normal bundle ν,

$$M^\alpha \text{ is } (s+t)\text{-dual to } M^{\beta \oplus \nu}.$$

This relationship, which was also obtained by R. Bott and A. Shapiro (unpublished), is now known as Atiyah duality. It provides a fundamental connection between duality theory and the theory of differentiable manifolds. Atiyah gave two applications of this relation. The first was to extend the work of I.M. James reducing the question of the existence of vector fields on spheres to a homotopy question about what he called stunted projective spaces. This reduction was later used by Adams in his celebrated paper [1962], in which he obtained a complete solution of the problem.

6. Generalized homology and cohomology theories

The concept of a generalized homology or cohomology theory emerged over a period of roughly seven years from 1955 to 1962. Along with the examples – stable homotopy–cohomotopy [Spanier and Whitehead, 1957], K-theory [Atiyah and Hirzebruch, 1959], bordism and cobordism [Atiyah, 1961b], [Conner and Floyd, 1964] – the search for a satisfactory duality theory guided the development of the subject.

If $\mathbf{E} = \{E_n; e_n\}$ is a spectrum, a generalized cohomology theory $H^*(\ ;\mathbf{E})$ is defined on the category of (pointed) finite CW-complexes by

$$H^q(X; \mathbf{E}) = \lim \left[S^{n-q}X; E_n\right].$$

E.H. Brown [1962] showed that every generalized cohomology theory defined on the category of finite CW-complexes, and having countable coefficient group, arises from a spectrum in this way. It was well known that the cohomology theories which existed up to this time had such a description, the first of which (singular cohomology) goes back to Eilenberg and MacLane [1943].

A year later G.W. Whitehead [1962] undertook a comprehensive study of generalized homology–cohomology theories from a homotopy point of view. He defined the generalized homology groups of X with coefficients in the spectrum $\mathbf{E}$ by

$$H_q(X; \mathbf{E}) = \overrightarrow{\lim} \left[S^{n+q}; E_n \wedge X\right].$$

As motivation for this definition, he cited D.M. Kan's [1958] theory of adjoint functors. He had also shown in an earlier paper [1956] that his definition gave the correct answer when $\mathbf{E}$ is an Eilenberg–MacLane spectrum.

After laying out the theory of products, Whitehead proved general Poincaré and Alexander duality theorems. From the fact that $H^*(\ ;\mathbf{E})$ and $H_*(\ ;\mathbf{E})$ are related by Alexander duality, he established conclusively that his definition of homology was the correct one. It is a consequence of Spanier–Whitehead duality that every cohomology theory H^* determines a "formal dual" homology theory H_*: Given a space X, choose a duality map $X \wedge X^* \to S^n$ and define $H_q(X) = H^{n-q}(X^*)$. This eventually leads to a homology theory H_*. Now, the fact that $H^*(\ ;\mathbf{E})$ and $H_*(\ ;\mathbf{E})$ are related by Alexander duality implies that $H_*(\ ;\mathbf{E})$ is the formal dual of $H^*(\ ;\mathbf{E})$ as desired.

Whitehead, G.W.'s fundamental paper gave a complete and satisfactory generalization of Poincaré and Alexander duality to arbitrary homology–cohomology theories.

7. Umkehr maps

An Umkehr map is a map related to an original map which reverses the arrow, that is the source of the original map becomes the target of the Umkehr map. The name has not solidified yet; sometimes Umkehr maps are called wrong way maps, or Gysin maps, or even transfer maps.

The first appearance of Umkehr maps occurred in [Hopf, 1930]. For a map $f: M \to N$ between two combinatorial manifolds of the same dimension, f induces a homomorphism $f_*: H_*(M; Q) \to H_*(N; Q)$. Now intersection theory gave rise to a ring structure, called the intersection ring, due to Lefschetz. The map f_* is not a ring homomorphism, however, Hopf managed to define a "wrong way homomorphism" which did preserve the ring structure when the manifolds had the same dimension. He called it the "Umkehr homomorphismus" from $H_*(N; Q) \to H_*(M; Q)$.

With the invention of cohomology, Hans Freudenthal [1937] could explain the Umkehr homomorphism in terms of Poincaré duality. In modern notation the Umkehr homomorphism is what we call the Poincaré duality map

$$f_! = D_M \circ f^* \circ D_N^{-1},$$

where D_M denotes the Poincaré isomorphism from cohomology to homology. This idea begs to be generalized to manifolds of different dimensions. It was finally done by Hopf's student Gysin [1941] in his dissertation. Because of this, Umkehr maps are frequently called Gysin maps.

Integration along the fibre was introduced by A. Lichnerowicz [1948]. Suppose $F \to E \xrightarrow{\pi} B$ is a fibration. If it is a fibre bundle with F and E and B all C^∞ manifolds, then using de Rham cohomology, an i-form on E can be integrated over each fibre to give an $(i-n)$-form on B, where n is the dimension of F. This gives a cochain map on the de Rham cochain complex of forms, and thus we get an Umkehr map $\pi^! : H^i(E; \mathbb{Q}) \to H^{i-n}(B; \mathbb{Q})$ called *integration along the fibre*.

Chern and Spanier [1950] extended integration along the fibre to more general fibre bundles where $H_n(F) \cong \mathbb{Z}$ for the top dimension n. With the Serre spectral sequence in [Serre, 1951], integration along the fibre could be defined for any oriented fibration whose fibre has a top nonzero homology group $H_n(F; V) \cong V$ and V is any field of coefficients. Then reading along the top line of the E^∞ and E^2 terms gives integration along the fibre both for homology and cohomology.

In the late 1950's and early 1960's, Umkehr maps played an important role in the generalization of the Riemann–Roch theorem and in the Atiyah–Singer index theorem.

Dieudonné [1989], gives a very lively account of the Riemann–Roch theorem. "In the late 1950's the growing usefulness of categorical notions gradually convinced mathematicians that morphisms rather than objects had to be emphasized in many situations. It was that trend that led Grothendieck to believe that the Riemann–Roch–Hirzebruch formula ... is only a special case of a relative Riemann–Roch relation dealing with a morphism $f : X \to Y$ of smooth projective varieties; the relation ... would then be the case in which Y is reduced to a single point. The problem was thus to replace both sides [of the relation] by meaningful generalizations when X and Y are arbitrary." On one side of the Riemann–Roch–Hirzebruch equation he needs to replace "integration" by an Umkehr map. So Grothendieck introduced the Poincaré duality map $f^! : H^*(X) \to H^*(Y)$, which is the dual to the map $f_!$ on homology defined by Gysin [1941]. For the other side of the equation, he had to invent K-theory.

Atiyah and Hirzebruch [1959] extended K-theory, and hence the Riemann–Roch theorem and the K-theory Umkehr map.

An important set of notes was produced by J.M. Boardman [1966]. In it he develops his very influential ideas on spectra. Chapter V was devoted to "Duality and Thom Spectra". In Section 6, entitled Transfer Homomorphisms, Boardman collected together eight constructions and called them transfers. (By now, the word "transfer" is usually taken to mean an Umkehr map which does not shift dimension.) These Umkehr homomorphisms all satisfied seven equations and turned out to agree on generalized homology and cohomology in the situations in which their definitions were valid. The seven relations forms a very tricky but useful calculus. In particular the Umkehr is functorial. A particularly useful relation gives $f^!(f^*(\alpha)) = \alpha \cup f^!(1)$. That is: Composition of the Umkehr homomorphism with the induced homomorphism results in multiplication by a fixed element $f^!(1)$. This turns out to be very important in the development of the transfer.

Boardman, as a graduate student, had discovered a simple Umkehr map for bordism theory, and he asked several people for more examples and got quite a few leads. Although the constructions were quite-different looking, the homomorphisms agreed on the intersections of their domains of validity.

Boardman's thesis (Cambridge 1964) problem was a computational problem and he was collecting as many tools as he could. What he published of his thesis used a slick proof, and his Umkehr maps were not mentioned. Fortunately, he collected his material in Chapter V, Section 6, of his notes, where they played a seminal role in our next topic – Transfers.

8. Transfers

Duality has influenced and been influenced by almost every subarea of algebraic topology. Even if we concentrate only on the key theorems of strong duality, we find the number of topics too vast to describe. We choose the subject of Transfers to illustrate the action and reaction of strong duality on a particular subject. Therefore we limit our discussion of the transfer to its origins and its complex relation to strong duality. We choose Transfers, even though there are more important topics we could have considered and even though the subject is still developing, because we know that area well and have worked in it and so we can provide information which is not readily available in the published record. In addition to our own recollections, we benefited from conversations and correspondence with J.M. Boardman, A. Dold, D.S. Kahn and S.B. Priddy.

The transfer began as a group-theory construction, which produced a homomorphism from a group G made abelian to a subgroup H of G of finite index made abelian. Thus $G^{ab} \to H^{ab}$ (see [Hall, 1959], p. 201).

It was Beno Eckmann who realized that this was a special case of a construction made for covering spaces and when applied to $K(\pi, 1)$'s gave the group-theory result. He gave the name *transfer* to the homomorphism in cohomology for a covering space with finite fibre, $\tau : H^i(\widetilde{X}) \to H^i(X)$, whose composition with the projection homomorphism p^* is multiplication by the number of elements in the fibre.

The relationship to group theory came about as follows. W. Hurewicz [1936] recognized that aspherical spaces were classified up to homotopy by their fundamental group. Now we call these spaces $K(\pi, 1)$, where π is the group. If H is a subgroup of π, then $K(H, 1)$ is a covering space of $K(\pi, 1)$. Then the group cohomology of π is the cohomology of the $K(\pi, 1)$. Eckmann [1953] noted that the group theory transfer was in fact the dual of his transfer homomorphism for $H_1(K(\pi, 1))$, which is the group π made abelian. One can see part of the idea for transfer in [Eckmann, 1945–1946].

The emphasis on describing the transfer as a homomorphism was probably part of the general movement inspired by the Eilenberg–Steenrod axioms which indicated that the morphisms were as important as the objects. The (covering space) transfer was also used by Conner and Floyd and S.D. Liao in the study of finite transformation groups in the early 1950's.

By the early 1970's, covering transfers were in the air. An interesting account of the transfer up to this time is given in J.F. Adams [1978] book in §4.1. Mainly, the thrust was the ad hoc construction of covering space transfers for different cohomology theories. These could all be united by a construction of the transfer as an S-map. Various S-map constructions were made independently by 1971 by Dan Kahn, Jim Becker, and F.W. Roush [1971]. None of these were published since the discoverers did not believe they were important.

J.F. Adams [1978] wrote in his book on pp. 104–105: "Transfer came to the attention of the general topological public when Kahn and Priddy [1972] published their well-known paper in 1972 ... Kahn and Priddy wrote: 'the existence of the transfer seems to be well-known, but we know of no published account'. The topological world thus learned that all well-informed persons were supposed to know about transfer, although hardly anyone did unless they were lucky ... the rapid spread of a general conviction that the transfer was very good business owed much to the fact they solved a problem of some standing in homotopy theory".

The transfer for finite covering fibrations has a spectacular generalization to fibrations. If $F \xrightarrow{i} E \xrightarrow{p} B$ is a Hurewicz fibration with compact fibre F (and very mild conditions on B), there is an S-map $\tau : B^+ \to E^+$ which induces homomorphisms τ_* and τ^* on ordinary homology and cohomology, respectively, such that $p_* \circ \tau_*$ and $\tau^* \circ p^*$ are both multiplication by the Euler–Poincaré characteristic $\chi(F)$.

The only hints that transfers could exist for fibrations came from an early consequence of the Leray–Serre spectral sequence: For a field of coefficients, if i^* maps onto the top dimensional cohomology group of F, then p^* must be injective in cohomology with the same coefficients. Now Borel [1956] observed that if F were a closed smooth oriented manifold and $F \xrightarrow{i} E \xrightarrow{p} B$ were a smooth oriented fibre bundle, then i^* mapped onto the top cohomology for $\mathbb{Z}_p$ coefficients whenever p did not divide $\chi(F)$. Thus for such p, the projection p^* is injective. This result would follow immediately if there were a transfer for fibrations.

Also, dual to the projection p is the transgression $\omega : \Omega B \to F$ from the Nomura–Puppe sequence mentioned in Section 2. It had recently been discovered, [Gottlieb, 1972], that $\chi(M)\omega^* = 0$ for M a smooth manifold. This result was true for cohomology with *any* coefficients. A proof depended on the same fact about i^* which was central to Borel's result.

These considerations led to the question of the existence of a transfer for fibrations in the Fall of 1972. Boardman's Umkehr map calculus immediately gave a transfer in Borel's special situation for singular cohomology:

$$\tau(\cdot) := p^!(\cdot \cup \chi),$$

where $p^!$ is the spectral sequence version of integration along the fibre and χ is the Euler class of the bundle of tangents along the fibre. By the Spring of 1973 the transfer theorem had been extended to fibre bundles whose fibre was a manifold with boundary, [Gottlieb, 1975], and the transfer existed as an S-map for fibre bundles whose structure group was a compact Lie group, [Becker and Gottlieb, 1975].

In the announcement [Becker et al., 1975] all the conditions on the fibration were essentially removed (as long as F was homotopic to a finite complex). The transfer was an S-map which satisfied in singular homology and cohomology the following relations.

$$p_* \circ \tau_* = \text{multiplication by } \chi(F),$$
$$\tau^* \circ p^* = \text{multiplication by } \chi(F).$$

In fact, if the fibration is equipped with a fibre-preserving map, then there is a transfer which satisfies the above conditions with a suitable Lefschetz number replacing the Euler characteristic.

A consequence of the generalization of the transfer was the generalization of the *transgression theorem* $\chi(F)\omega^* = 0$, which now holds in all cohomology theories and in all homology theories, for essentially any fibration whose fibre is homotopy equivalent to a finite complex. In addition, for fibre-preserving maps, the Euler characteristic can be replaced by the same Lefschetz number as in the transfer theorem (see [Becker and Gottlieb, 1976]).

To extend the transfer theorem to Hurewicz fibrations and to construct the transfer as an S-map, the use of strong duality became vital. The two methods indicated in the announcement were explained in detail in [Casson and Gottlieb, 1977] and in [Becker and Gottlieb, 1976]. In the first method, the role of integration along the fibre was played by the Poincaré duality map. Topological maps on Thom complexes induced the Poincaré duality map on the homology level of the underlying manifolds, and the cup product could be induced by a map between Thom complexes as well. These considerations yielded the transfer as an S-map in the smooth fibre bundle case where everything was a smooth oriented manifold. The fact that every Hurewicz fibration was a fibrewise retract of these smooth oriented fibre bundles was proved by a series of tricks, and the retraction of the smooth transfer resulted in a transfer for the general case.

On the other hand, the second method depended upon the existence of S-duality in the category of ex-spaces over a space B. This allowed the construction of the transfer via a chain of duality maps and a diagonal map.

Meanwhile Albrecht Dold [1974a] was conducting a deep study of fixed point theory. He studied the index of parameterized families of maps. He was influenced by R.J. Knill [1971]. As a by-product to defining the index for a parameterized family of maps, Dold discovered a transfer. He did not regard it as important until Puppe told him about the transfer for fibrations. He sent a note to the Comptes Rendus, [Dold, 1974b], and expanded the paper in [Dold, 1976], calling the transfer the fixed point transfer. Nowadays it is also called the Dold transfer or the Becker–Gottlieb–Dold transfer as well.

Dold and Puppe organized a seminar on the transfer for fibrations. As a result of this they realized, [Dold and Puppe, 1980], that the observation that the degree of the map $S^n \xrightarrow{\mu} DX \wedge X \to X \wedge DX \xrightarrow{\mu} S^n$ equals $\chi(X)$ leads to the categorical definition of trace given here in §2 of this paper. The categorical definition of transfer is given by the map

$$\tau_f : I \xrightarrow{\eta} A \otimes DA \xrightarrow{\gamma} DA \otimes A \xrightarrow{Df\otimes\Delta} DA \otimes A \otimes A \xrightarrow{\varepsilon\otimes id} I \otimes A = A.$$

Thus in the category of ex-spaces over B, which they call Stab_B, they note that if p is a well-sectioned Hurewicz fibration whose fibre has the stable homotopy type of a finite CW complex, then the fibrewise dual in [Becker and Gottlieb, 1976] results in the transfer for fibrations for Stab_B, while the fact that fibrewise *ENR* also have strong duals, leads to the Dold transfers, [Dold and Puppe, 1980].

In addition, they can prove the Lefschetz fixed point theorem as a consequence of their point of view. So, many important theorems which seemed to be independent, and which seemed to have little to do with duality, can be shown to be consequences of Dold and Puppe's concept of strong duality for monoidal categories.

Postscript: Dold and Puppe [1980] explicitly remarked that the condition on the fibre in the transfer theorem could be relaxed from being homotopically equivalent to a finite complex to merely being S-equivalent to a finite complex, and that this was implicitly proved in [Becker and Gottlieb, 1976]. This advance permits the observation that the transfer exists in a purely group-theoretic setting where fibration is replaced by surjective homomorphism and the condition on the fibre is replaced by the condition that the kernel has finitely generated homology, [Gottlieb, 1983]. Thus the transfer is returned back to group theory with all the topological conditions removed in a vastly more general situation. But its construction is not at all group-theoretic.

We also note that there is quite recent work by Dwyer [1996] which relaxes the hypothesis on the fibre to the case where the fiber of the map satisfies only a homological finiteness condition relative to some spectrum.

Bibliography

Adams, J.F. (1962), *Vector fields on spheres*, Ann. of Math. **75** (2), 603–632.

Adams, J. (1978), *Infinite Loop Spaces*, Princeton Univ. Press, Princeton, NJ.

Alexander, J.W. (1915), *A proof of the invariance of certain constants of analysis situs*, Trans. Amer. Math. Soc. **16**, 148–154.

Alexander, J.W. (1922), *A proof and extension of the Jordan–Brouwer separation theorem*, Trans. Amer. Math. Soc. **23**, 333–349.

Alexander, J.W. (1926), *Combinatorial analysis situs*, Trans. Amer. Math. Soc. **28**, 301–329.

Atiyah, M.F. (1961a), *Thom complexes*, Proc. London Math. Soc. **11** (3), 291–310.

Atiyah, M.F. (1961b), *Bordism and cobordism*, Proc. Cambridge Philos. Soc. **57**, 200–208.

Atiyah, M. and Hirzebruch, F. (1959), *Riemann–Roch theorems for differentiable manifolds*, Bull. Amer. Math. Soc. **65**, 276–281.

Atiyah, M. and Hirzebruch, F. (1962), *Vector bundles and homogeneous spaces*, Proc. Sympos. Pure Math. **III**, American Mathematical Society, Providence, RI, 7–38.

Becker, J. and Gottlieb, D. (1974), *Applications of the evaluation map and transfer map theorems*, Math. Ann. **211**, 277–288.

Becker, J. and Gottlieb, D. (1975), *The transfer map and fiber bundles*, Topology **14**, 1–12.

Becker, J. and Gottlieb, D. (1976), *Transfer maps for fibrations and duality*, Compositio Mathematica **33**, 107–133.

Becker, J., Casson, A. and Gottlieb, D. (1975), *The Lefschetz number and fiber preserving maps*, Bull. Amer. Math. Soc. **81**, 425–427.

Bollinger, M. (1972), *Geschichtliche Entwicklung des Homologiebegriffs*, Archive for the History of the Exact Sciences, Vol. 9, 94–170.

Borel A. (1956), *Sur la torsion des groupes de Lie*, J. Math. Pures Appl. **35** (9), 127–139.

Borsuk K. (1936), *Sur les groupes des classes de transformations continues*, C. R. Acad. Sci. Paris **202**, 1400–1403.

Bredon, G. (1993), *Topology and Geometry*, Springer, New York.

Brouwer, L.E.J. (1976), *Collected Works, Vol. II*, North-Holland, Amsterdam.

Brown, E.H. (1962), *Cohomology theories*, Ann. of Math. **75**, 467–484; Correction: (1963), **78**, 201.

Cairns, S. (1930), *The cellular division and approximation of regular spreads*, Proc. Nat. Acad. Sci. USA **16**, 488–490.

Casson, A. and Gottlieb, D.H. (1977), *Fibrations with compact fibres*, Amer. J. Math. **99**, 159–189.

Čech, E. (1936), *Multiplications on a complex*, Ann. of Math. **37**, 681–697; Čech, *Topological Papers*, Prague, 1968, Akademia, 417–433.

Chern, S.S. and Spanier, E. (1950), *The homology structure of fibre bundles*, Proc. Nat. Acad. Sci. USA **36**, 248–255.

Conner, P. and Floyd, E. (1964), *Differentiable Periodic Maps*, Ergebnisse der Mathematik und ihrer Grenzgebiete, Academic Press, New York.

tom Dieck, T., Kamps, K. and Puppe, D. (1970), *Homotopietheorie*, Lecture Notes in Math. vol. 157, Springer, Berlin.

Dieudonné, J. (1989), *A History of Algebraic and Differential Topology 1900–1960*, Birkhäuser, Boston.

Dold, A. (1972), *Lectures on Algebraic Topology*, Springer, Heidelberg.

Dold, A. (1974a), *The fixed point index of fibre preserving maps*, Inventiones Math. **25**, 281–297.

Dold, A. (1974b), *Transfer des points fixes d'une famille continue d'applications*, C. R. Acad. Sci. Paris, Sér. A **278**, 1291–1293.

Dold, A. (1976), *The fixed point transfer of fibre-preserving maps*, Math. Zeitschr. **148**, 215–244.

Dold, A. and Puppe, D. (1980), *Duality, trace, and transfer*, Proceedings of the International Conference on Geometric Topology (Warsaw, 1978), PWN, Warsaw, 81–82.

Dowker, C.H. (1937), *Hopf's theorem for non-compact spaces*, Proc. Nat. Acad. Sci. USA **23**, 293–294.

Dwyer, W.G. (1996), *Transfer maps for fibrations*, Math. Proc. Cambridge Philos. Soc. **120**, 221–235.

Eckmann, B. (1945–1946), *Der cohomologie–Ring einer beliebigen Gruppe*, Comment Math. Helvetici **18**, 223–282.

Eckmann, B. (1953), *Cohomology of Groups and Transfer*, Ann. of Math. **58** (2), 481–493.

Eckmann, B. (1956), *Homotopie et dualité*, (Coll. Top. Alg. Louvain) Centre Belge des Rech. Math., 41–53.

Eckmann, B. (1958), *Groupes d'homotopie et dualité*, Bull. Soc. Math. France **86**, 271–281.

Eilenberg, S. (1944), *Singular homology*, Ann. of Math. **45**, 407–447.

Eilenberg, S. and MacLane, S. (1943), *Relations between homology and homotopy groups*, Proc. Nat. Acad. Sci. USA **29**, 155–158.

Eilenberg, S. and Steenrod, N. (1945), *Axiomatic approach to homology theory*, Proc. Nat. Acad. Sci. USA **31**, 177–180.

Eilenberg, S. and Steenrod, N. (1952), *Foundations of Algebraic Topology*, Princeton Univ. Press.

Freudenthal, H. (1937), *Zum Hopfschen Umkehrhomomorphismus*, Ann. of Math. **38**, 847–853.

Gottlieb, D. (1972), *Applications of bundle map theory*, Trans. Amer. Math. Soc. **171**, 23–50.

Gottlieb, D. (1975), *Fibre bundles and the Euler characteristic*, J. Diff. Geom. **10**, 39–48.

Gottlieb, D. (1983), *Transfers, centers, and group cohomology*, Proc. Amer. Math. Soc. **89**, 157–162.

Gysin, W. (1941), *Zur Homologietheorie der Abbildungen und Faserungen der Mannigfaltigkeiten*, Comment. Math. Helvetici **14**, 61–122.

Hall, Marshall (1959), *The Theory of Groups*, MacMillan, New York.

Henn, H.W., and Puppe, D. (1990), *Algebraische Topologie*, Ein Jahrhundert Mathematik 1890–1990, Dokumente Gesh. Math. vol. 6, F. Vieweg, Braunschweig.

Hilton, P. (1980), *Duality in homotopy theory: a retrospective essay*, J. Pure Appl. Algebra **19**, 159–169.

Hodge, W.V.D. (1941), *The Theory and Applications of Harmonic Integrals*, Cambridge Univ. Press.

Hopf, H. (1930), *Zur Algebra der Abbildungen von Mannigfaltigkeiten*, J. für reine und angew. Math. **105**, 71–88; Hopf, *Selecta*, Springer, 1964, 14–37.

Hurewicz, W. (1936), Asphärische Räume **39**, 215–224.

Hurewicz, W., Dugundji, J. and Dowker, C. (1948), *Connectivity groups in terms of limit groups*, Ann. of Math. **49**, 391–406.

James, I.M. (1971), *Ex-homotopy theory I*, Ill. J. Math. **15**, 324–327.

Kan, D. (1958), *Adjoint functors*, Trans. Amer. Math. Soc. **87**, 294–329.

Kahn, D.S. and Priddy, S.B. (1972), *Applications to stable homotopy theory*, Bull. Amer. Math. Soc. **78**, 981–987.

Knill, R. (1971), *On the homology of a fixed point set*, Bull. Amer. Math. Soc. **77**, 184–190.

Lefschetz, S. (1926), *Intersections and transformations of complexes and manifolds*, Trans. Amer. Math. Soc. **28**, 1–49; (also in *Selected Papers*, Chelsea, New York (1971), 199–247).

Lefschetz, S. (1928), *Closed point sets on a manifold*, Ann. of Math. **29**, 232–254; (also in *Selected Papers*, Chelsea, New York (1971), 545–568).

Lefschetz, S. (1930), *Topology*, Amer. Math. Soc. Coll. Publ. No. 12, Providence, RI.

Lefschetz, S. (1942), *Algebraic Topology*, Amer. Math. Soc. Coll. Publ. No. 27, Providence, RI.

Lichnerowicz, A. (1948), *Un théorème sur l'homologie dans les espaces fibrés*, C. R. Acad. Sci. Paris **227**, 711–712.

Lima, E. (1959), *The Spanier–Whitehead duality in new homotopy categories*, Summa Brasil. Math. **4**, 91–148.

Massey, W. (1980), *Singular Homology Theory*, Springer, Heidelberg.

Massey, W. (1999), *A history of cohomology theory*, History of Topology, I.M. James, ed., Elsevier, Amsterdam, 579–603.

Milnor, J. and Spanier, E. (1960), *Two remarks on fiber homotopy type*, Pacific J. Math. **10**, 585–590.

Moore, J.C. (1956), *On a theorem of Borsuk*, Fund. Math. **43**, 195–201.
Nomura, Y. (1960), *On mapping sequences*, Nagoya Math. J. **17**, 111–145.
Poincaré, H. (1892), *Sur l'Analysis Situs*, C. R. Acad. Sci. Paris **115**; (also in *Oeuvres*, Vol. VI, 186–192).
Poincaré, H. (1893), *Sur la géneralisation d'un théorème d'Euler relatif aux Polyédres*, C. R. Acad. Sci. Paris **117**, 144–145; (also in *Oeuvres*, Vol. XI, 6–7).
Poincaré, H. (1895), *Oeuvres*, Vol. VI, Gauthier-Villars, Paris 1953.
Pontrjagin, L. (1934), *The general topological theorem of duality for closed sets*, Ann. of Math. **35**, 904–914.
Puppe D. (1958), *Homotopiemengen und ihre induzierten Abbildungen I*, Math. Z. **69**, 299–344.
de Rham, G. (1931), *Sur l'Analysis Situs des variétès á n dimensions*, J. Math. Pures Appl. **10** (9), 115–200; (also in *Oeuvres Mathématiques*, L'Enseignment math., Genéve, 1981, 23–213).
Roush, J.W. (1971), *Transfer in generalized cohomology theories*, Ph.D. Thesis, Princeton.
Segal, G. (1970), *Equivariant stable homotopy theory*, Actes du Congres International des Mathematiciens (Nice, 1970), Vol. 2, 59–63.
Serre, J.-P. (1951), *Homologie singuliére des espaces fibrés*, Ann. of Math. **54**, 425–505, 24–204.
Spanier, E. (1948), *Cohomology theory for general spaces*, Ann. of Math. **49**, 407–427.
Spanier, E. (1949), *Borsuk's cohomotopy groups*, Ann. of Math. **50**, 203–245.
Spanier, E. (1959), *Function spaces and duality*, Ann. of Math. **70**, 338–378.
Spanier, E. (1959a), *Infinite symmetric products, function spaces, and duality*, Ann. of Math. **69**, 142–198; erratum, 733.
Spanier, E. (1966), *Algebraic Topology*, McGraw Hill, New York.
Spanier, E. and Whitehead, J.H.C. (1953), *A first approximation to homotopy theory*, Proc. Nat. Acad. Sci. USA **39**, 655–660.
Spanier, E. and Whitehead, J.H.C. (1955), *Duality in homotopy theory*, Mathematika **2**, 56–80.
Spanier, E. and Whitehead, J.H.C. (1957), *The theory of carriers and S-theory*, Algebraic geometry and topology. A symposium in honor of S. Lefschetz, Princeton Univ. Press, Princeton, NJ, 330–360.
Steenrod, N. (1936), *Universal homology groups*, Amer. J. Math. **58**, 661–701.
Thom, R. (1952), *Espaces fibrés en sphéres et carrés de Steenrod*, Ann. Ec. Norm. Sup. **69**, 109–181.
Thom, R. (1954), *Quelques propriétés globales des variétés différentiables*, Comment. Math. Helvetici, 17–86.
Wall, C.T.C. (1967), *Poincaré complexes I*, Ann. of Math. **86**, 213–245.
Wall, C.T.C. (1970), *Surgery on Compact Manifolds*, Academic Press, New York.
Whitehead, G.W. (1956), *Homotopy groups of joins and unions*, Trans. Amer. Math. Soc. **83**, 55–69.
Whitehead, G.W. (1962), *Generalized homology theories*, Trans. Amer. Math. Soc. **102**, 227–283.
Whitehead, G.W. (1983), *Fifty years of homotopy theory*, Bull. Amer. Math. Soc. **8**, 1–29.
Whitney, H. (1937), *On products in a complex*, Proc. Nat. Acad. Sci. USA **23**, 285–291.
Whitney, H. (1938), *On products in a complex*, Ann. of Math. **39**, 397–432.
Wirthmüller, K. (1975), *Equivariant S-duality*, Arch. Math. (Basel) **26**, 427–431.

This Page Intentionally Left Blank

CHAPTER 26

A Short History of H-spaces

J.R. Hubbuck

Department of Mathematical Sciences, University of Aberdeen, Old Aberdeen AB24 3UE UK
E-mail: j.hubbuck@maths.abdn.ac.uk

1. The basic homotopy theory

At its simplest, an H-space is a triple $(X, *, \mu)$, where X is a topological space, $\mu : X \times X \to X$ is a continuous multiplication and the base point $*$ is a two-sided multiplicative unit, $\mu(*, x) = \mu(x, *) = x$ for all $x \in X$. One writes "X is an H-space", if $*$ and μ are specified. Frequently it is convenient to replace the unit by a homotopy unit, that is, require only that $\mu(*, \ )$ and $\mu(\ , *)$ are homotopic to the identity map of X in the usual sense, where here and throughout this note, maps and homotopies preserve base points. We will assume in the first two sections that all spaces have the homotopy types of CW-complexes as that is an assumption made in much of the literature. A consequence is that the multiplication of an H-space with a homotopy unit can be deformed to one with a unit. By requiring only a homotopy unit, it follows that a space homotopic to an H-space is itself an H-space. We will assume also that H-spaces are connected, for otherwise one normally considers the set of path components separately.

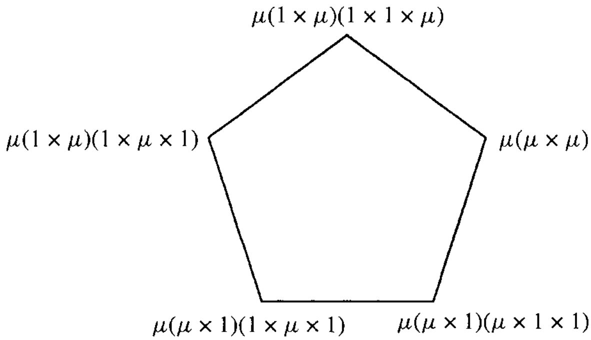
The H in H-space was suggested by J.-P. Serre in recognition of the influence exerted on the subject by Heinz Hopf. The latter considered compact, connected Lie groups, but the usual interpretation of the main theorem of [23] requires only H-space properties. We will return to Hopf's theorem in Section 2.

Contrasting the definition of an H-space with that of a connected topological group, the former lacks multiplicative inverses and no associativity assumptions are given for the multiplication.

The absence of inverses is not a major problem from a homotopy perspective. A left inverse $l : X \to X$ of an H-space is a map such that the composition $\mu(l \times 1)\Delta : X \to X$ is homotopically trivial. A right inverse is defined similarly. M. Sugawara proved in [44] that left and right inverses exist. They can, however, be homotopically, even homologically distinct. About the same time I. James in [28] established that μ induces on the set of based homotopy classes $[K, X]$ an algebraic loop structure: in more recent notation this says that for any two homotopy classes f and g, there exists a unique third class $D(f, g)$ with $f = D(f, g) + g$, where $+$ denotes the operation induced by μ.

HISTORY OF TOPOLOGY
Edited by I.M. James

The absence of any associativity assumption on μ is more problematical and for many purposes one needs to introduce some such assumption. The simplest approach, with the initial definition above, is just to assume that the multiplication μ is associative. This has the major disadvantage that a space homotopically equivalent to an associative H-space may not itself inherit an associative multiplication. A homotopy invariant definition is that μ is homotopy associative, that is $\mu(\mu \times 1) \simeq \mu(1 \times \mu) : X^3 \to X$. This ensures that an algebraic loop is a group. (One needs no assumption on an H-multiplication to ensure that the algebraic loop structure induced on $\pi_1(X)$ by the H-multiplication coincides with the normal group structure and that the group is Abelian [18].) But homotopy associativity is not a sufficiently strong assumption for many purposes. Sometimes it can be extended to an invariant form of associativity. This represents work of J. Stasheff in the early 1960's [41, 42]. Writing down accurate details is complicated, but the basic idea can be indicated. If X is an H-space, we have a map $r_2 : S^0 \to X^{X^3}$ defined by $r_2(-1) = \mu(1 \times \mu)$, $r_2(1) = \mu(\mu \times 1)$. If μ is homotopy associative, this can be extended to a map $s_3 : D^1 \to X^{X^3}$ and we say that X is an A_3-space. (An H-space is an A_2-space.) We can use μ and the associating homotopy to define $r_3' : P \to X^{X^4}$, where P is a regular pentagon:



the images of the vertices are the maps labelled and the edges are mapped using the obvious homotopies. Topologically we have a map $r_3 : S^1 \to X^{X^4}$ and we ask if it extends to $s_4 : D^2 \to X^{X^4}$. If it does, we say informally that X is an A_4-space – more formally, as one needs the multiplication to define an H-space, one need the multiplication and all the homotopies, the A_4-form, to define an A_4-space. The process can be continued in a less transparent manner to define an A_n-space for all n and Stasheff defined an ingenious polyhedron to keep track of the data. If compatible A_n-structures exist for all n, then X is an A_∞-space. One must also introduce appropriate morphisms.

An associative H-space X has a classifying space BX and there is a homotopy equivalence $X \to \Omega BX$ preserving the multiplications up to homotopy; many authors have worked in this area but in this context, the result is usually attributed to A. Dold and R. Lashof [13]. Related arguments show that X is an A_∞-space if and only if it has the homotopy type of a loop space [40]. So when X is an associative H-space, a homotopy equivalent space is an A_∞-space and, with appropriate definitions of morphisms, the A_∞-structures are preserved.

We return to the concept of an H-space $(X, *, \mu)$. In general, if a space has one H-multiplication, it has many homotopically distinct multiplications. There is a sequence of algebraic loops

$$0 \to [X \wedge X, X] \xrightarrow{j^*} [X \times X, X] \xrightarrow{i^*} [X \vee X, X] \to 0$$

which is short exact in the sense that j^* is injective, i^* is surjective and $i^*f = i^*g$ if and only if $D(f, g)$ lies in the image of j^*. It follows that the cardinality of the set of homotopically distinct multiplications on a fixed space X is the same as that of $[X \wedge X, X]$; in general this is infinite, even for compact Lie groups [4].

The interest in different multiplications can be illustrated by the simplest nontrivial example S^3 or $SU(2)$ with its Lie multiplication μ, [29, 30]. As $\pi_6(S^3) \cong Z/12Z$, there are 12 homotopically distinct multiplications $\mu + nj^*(\omega)$, where ω is the standard generator. Such a multiplication is homotopy associative if and only if $n = 0$ or 1 mod 8. In [40] it is shown that each of these latter multiplications can be extended to an A_∞-multiplication on S^3, but not in general an associative multiplication. In fact, S^3 can support uncountably many different A_∞ or loop multiplications, or equivalently, there are uncountably many homotopically distinct spaces whose loop spaces have the homotopy type of S^3, [39]. If one considers S^7 with multiplication given by the product of Cayley numbers of norm one, a similar analysis shows that there 120 distinct multiplications, none of which is homotopy associative.

As the loop space on any simply connected CW-complex is an H-space, examples of the latter abound. One needs therefore to impose additional structure to obtain a coherent theory. One direction, not appropriate here, is to consider infinite loop spaces. The theory of H-spaces was generally concerned with establishing the relationship with the homotopy of connected Lie groups. Therefore for the remainder of this article, we impose finiteness conditions on the space of an H-space.

2. Homological considerations

We assume that the space of an H-space has the homotopy type of a connected, finite complex; this is abbreviated to "finite H-space".

The proof of the main theorem of Hopf of [23] referred to in the section above implies that the rational cohomology ring $H^*(X, Q)$ is an exterior algebra on odd-dimensional generators. This is purely a Hopf algebraic result for graded, connected, finite-dimensional Hopf algebras over a field of characteristic zero, where the multiplication is associative and commutative; in this area, neither multiplications nor comultiplications are assumed automatically to be associative or commutative, [36]. So $H^*(X, Q)$ and $H_*(X, Q)$ are dual Hopf algebras in this sense. The number of generators is called the rank of X and it is shown in [23] that this is consistent with the usage in Lie theory. If in addition, the comultiplication induced by μ on $H^*(X, Q)$ is associative, for example, if μ is homotopy associative, each exterior generator can be chosen to be primitive (that is, lies in the kernel of $\mu^* - \pi_1^* - \pi_2^*$, where π_i is a projection) and $H^*(X, Q)$ is said to be primitively generated.

There are related results for $H^*(X, Z/pZ)$ for each prime p, [5]. For an odd prime p, this ring is a tensor product of an exterior algebra on odd-dimensional generators with polynomial algebras on single even-dimensional generators truncated at heights p^q with (distinct) $q \geqslant 1$; when $p = 2$ one must drop the condition that the polynomial generators have even dimensions. The Hopf algebras $H^*(X, Z/pZ)$ and $H_*(X, Z/pZ)$ are dual.

The theory of finite H-spaces received great impetus from the research of W. Browder in the early 1960's who investigated the relationship between these theorems. He established that a covering space of a finite H-space was a finite H-space, [6]. He proved that if X is simply-connected, it has the homotopy type of a topological manifold; the relationship with smooth manifolds appears to remain unresolved. It is curious that the close relationship between finite H-spaces and manifolds has had so little influence in the former area, though presumably it helped to motivate Browder's later work in surgery theory. In [10] it is shown that X^+, the union of a finite H-space with a disjoint point is self-dual in the sense of Spanier and Whitehead.

Browder's work analyses the torsion structure in the homology of finite H-spaces. He proved the key result, [7], that $H_*(X, Z)$ has no p-torsion if and only if $H^*(X, Z/pZ)$ is an exterior algebra on odd-dimensional generators. The main tool in the analysis is the standard mod p Bockstein spectral sequence in cohomology or homology; at each level these are dual Hopf algebras. An important concept is that of "infinite implication" in a Hopf algebra and its dual, generalising the concept of infinite height. In this context, an element x has height p^q if $x^{p^{q-1}} \neq 0$ and $x^{p^q} = 0$ and we can write $x_0 = x, x_1 = x_0^p$, $\ldots, x_{q-1} = x_{q-2}^p$ and $0 = x_{q-1}^p$. In the corresponding implication sequence, each x_i can lie in the Hopf algebra or the dual Hopf algebra where if $x_i^p = 0$ for $i < q - 1$, there is an element $\bar{x}_i$ in the dual Hopf algebra with $\langle x_i, \bar{x}_i \rangle = 1$ and $x_{i+1} = \bar{x}_i^p \neq 0$. If either multiplication is not associative, one must define p-th powers by multiplying in a fixed order. There can be no element of infinite implication in the Bockstein spectral sequences of a finite H-space. It is shown that only in rather restricted circumstances can there exist a class which is both primitive and a boundary in the spectral sequences, unless it has infinite implication.

An H-space is homotopy commutative if $\mu T \simeq \mu : X \times X \to X$, where T is the switching map. It is shown in [8] that when X is a homotopy commutative finite H-space, $H_*(X, Z)$ has no 2-torsion and there are related results for odd primes. In [9] it is proved that the space of a finite H-space of rank one is one of S^1, S^3, S^7, RP^3 or RP^7. The question as to whether or not there were any H-multiplications on simply-connected finite complexes except when the space had the homotopy type of a product of a Lie group and 7-spheres was raised about this time and motivated much subsequent research.

Homological evidence for the reasonableness of the question was provided by work of E. Thomas [46, 47] on the action of the Steenrod algebra on the mod 2 cohomology of H-spaces, particularly on the submodule of primitive elements. He showed that they had to follow the same general pattern as for Lie groups. His most complete results are when $H^*(X, Z/2Z)$ is an exterior algebra on primitive odd-dimensional generators. A typical result, a proto-type of many similar results proved in later years by other mathematicians, was that if x_{2s+1} is one of the primitive generators and t is a positive integer such that the binomial coefficient $\binom{2s+1}{2t}$ is nonzero mod 2, then there exists a class $y_{2s-2t+1}$ whose image under Sq^{2t} is x_{2s+1}. This relation completely determines the action of the Steenrod algebra in $H^*(U(n), Z/2Z)$ or $H^*(Sp(n), Z/2Z)$. Further if the exterior generators occur in distinct degrees and the highest of these degrees is $2^q + 1$, then $H^*(X, Z/2Z)$ is isomorphic as an algebra over the Steenrod algebra to $H^*(U(n), Z/2Z)$ or $H^*(SU(n), Z/2Z)$ where $n = 2^{q-1} + 1$.

To prove such theorems, Thomas considered the projective plane of an H-space, somewhat following the approach of Adams to the Hopf invariant one problem [1]. Built into the fabric of an A_n-space X is the existence of a projective space P_nX. For simplicity, we will again assume that $H^*(X, Z/pZ)$ is an exterior algebra on odd-dimensional generators. With mild additional assumptions, certainly weaker than assuming that X is an A_{n+1}-space, a quotient of the cohomology ring $H^*(P_nX, Z/pZ)$ is a polynomial algebra truncated at height $n+1$ over the mod p Steenrod algebra on generators which are "suspensions" of the exterior generators of $H^*(X, Z/pZ)$. When X is a loop space, $P_\infty X = BX$ and $H^*(BX, Z/pZ)$ is a polynomial algebra. For the specific result of Thomas quoted above, one considers the action of the Steenrod algebra on a polynomial algebra truncated at height 3. Properties of the Steenrod squares place strong conditions on this action, leading to the proof.

A similar approach can be used replacing cohomology by complex K-cohomology and the Steenrod algebra by the algebra of exterior powers. One again assumes that $H^*(X, Z/2Z)$ is an exterior algebra on odd-dimensional generators and a marginally stronger hypothesis than that assumed above. If one again assumes that the generators occur in distinct degrees, then provided that the highest degree is not of the form $2^q - 1$, then $H^*(X, Z/2Z)$ is isomorphic over the Steenrod algebra to $H^*(G, Z/2Z)$, where G is $U(n)$, $SU(n)$, $Sp(n)$ or $S^1 \times Sp(n)$, [25]. However, a major deficiency in these approaches to classifying the cohomology rings of families of H-spaces is that there is no natural concept analogous to that of simplicity in Lie groups.

A related K-theory approach was used by J. Hubbuck to show that a homotopy commutative finite H-space has the same homotopy type as a product of circles [24]. The proof uses the result of Browder on homology torsion mentioned above and a geometric construction of James, [27].

The question of the existence of simply-connected finite H-spaces which are not homotopy equivalent to a product of a Lie group and 7-spheres was resolved decisively in 1968 by P. Hilton and J. Roitberg. In investigating cancellation phenomena for products of simply connected finite complexes, they discovered that the S^3-bundle over S^7 classified by 7ω is an H-space [19]. It does not have the homotopy type of either of the previously known examples with the same rational cohomology. An analysis of the result completely changed perspectives in the subject. This and other related developments made it natural thereafter, not to consider finite complexes, but p-local finite complexes which were H-spaces [20, 32, 50]; these became known as "mod p finite H-spaces". A major role was played in the subsequent analysis by A. Zabrodsky [49], though several mathematicians were involved. We indicate the techniques by returning to the general setting of the example of Hilton and Roitberg. One considers the pull-back diagram of fibre bundles

$$\begin{array}{ccc} E_{k\omega} & \longrightarrow & Sp(2) \\ \downarrow & & \downarrow \\ S^7 & \xrightarrow{\;k\;} & S^7 \end{array}$$

by a map of degree k mod 12. When $k = 4$, $E_{4\omega}$ is an H-space which looks like $S^3 \times S^7$ at the prime 2 and like $Sp(2)$ at all other primes. This is an example which can be obtained from a general construction of mixing homotopy types. To avoid technical qualifications, we indicate only a particular case. Let X and Y be simply-connected finite complexes whose rationalisations X_0 and Y_0 have been identified. If one partitions the set of all primes into two disjoint subsets P_1 and P_2, one has the localisation maps, $r_1 : X_{P_1} \to X_0$ and $r_2 : Y_{P_2} \to X_0$. The homotopy pull back of these maps produces a space with the homotopy type of a finite complex, which looks like X at the primes P_1 and the space Y at the primes P_2. If X and Y are rationally compatible H-spaces, one obtains a finite H-space. Of course one needs only the localised spaces to begin the construction. Some years earlier, F. Adams [2] had shown that all odd-dimensional spheres are H-spaces at odd primes. Homotopy mixing enabled new examples of finite H-spaces to be constructed from Lie groups by changing the homotopy type at an odd prime.

There is a more subtle variant called homotopy twisting. Again we mention only the case for simply-connected finite H-spaces. The Mislin genus [37] of such a space X, denoted by $G(X)$, is the set of all distinct homotopy types Y with $X_p \simeq Y_p$ for all primes p. If X is a simply-connected finite H-space then every member of $G(X)$ is a finite H-space. But more is true; if X is a homotopy associative H-space or a finite loop space, so is every member of the genus. (It seems that the details of a proof that there is a corresponding statement for A_n-spaces has not been written down.) For example, $G(Sp(2)) = \{Sp(2), E_{7\omega}\}$, and so $E_{7\omega}$ is a finite loop space, as first established by Stasheff [43]. The technique can be applied to other Lie groups producing infinite families of nonstandard finite H-spaces and loop spaces.

A quite different type of example of a finite H-space not known previously was constructed by J. Harper in [17], using obstruction theory. It is shown that for each odd prime, there exists a simply-connected mod p finite H-space with p-torsion in its integral homology. Homotopy mixing techniques enable one to construct a simply-connected finite H-space with this p-torsion. This is in marked contrast with the position for Lie groups, where torsion can exist only for the primes 2, 3 or 5.

There was in the 1970s a topic which attracted great attention, although it was not fully resolved until the early years of the next decade. This is the "loop space conjecture", which seems to have been first raised by A. Clark and J. Moore. R. Bott had shown that the loop space on a simply-connected Lie group had no homology torsion, and the conjecture was that the analogous statement was true for simply-connected finite H-spaces. Other properties of Lie groups were thought likely to hold true for finite H-spaces and were known or suspected to be consequences of a proof of the conjecture, for example, there is no p^2-torsion for any prime in the integral homology of simply-connected Lie group, the kernel of the Hurewicz homomorphism is the torsion in the homotopy and L. Hodgkin had shown in [20] that the complex K-cohomology of a simply-connected Lie group was torsion free. The verification of the loop space conjecture confirmed that these results were equally true for finite H-spaces. J. Lin [33, 34] showed that $H_*(\Omega X, Z)$ has no odd torsion when X is a simply-connected finite H-space and later obtained a weaker statement at the prime 2; the complete result for the prime 2 was established by R. Kane in [31].

The method of proof relied on earlier insights of Zabrodsky. He realised that secondary cohomology operations could be used effectively in the cohomology of H-spaces. A common difficulty with such operations defined on a particular cohomology class is to show that the image does not contain zero. He observed that using naturality properties applied to the comultiplication, one can deduce in favourable circumstances that the image is non-trivial, particularly when the initial class is primitive. Also in a remarkable note [48], he described a method of defining secondary operations on quotients of the cohomology ring of an H-space by Hopf ideals preserved by the Steenrod operations. The initial motivation appears to have been to remove technical restrictions from the type of results obtained by

Thomas mentioned earlier, the strategy was to use partial information on the action of the cyclic reduced powers to define secondary operations and then apply them to gain more information about the powers. Zabrodsky did not develop his ideas fully and this was done later by Lin and Kane; they and others added significant technical advances needed in the proof of the conjecture. The result is so elegant and important that one hopes that a more accessible proof will be found.

3. Finite loop spaces

Since the 1980s, much of the activity in finite H-space theory has concentrated on finite loop spaces, and more particularly their classifying spaces. Of course, if X is a finite loop space, $\Omega BX \simeq X$. At about the time when Thomas was considering truncated polynomial algebras over the mod 2 Steenrod algebra, N. Steenrod raised the general question of the classification of polynomial algebras over the algebra of cyclic reduced powers; this became known as the "Steenrod Problem". (There are variants of this problem, but this is the version on which Steenrod himself worked.) We have noted above that if X is a finite loop space and $H_*(X, Z)$ has no p-torsion, then $H^*(BX, Z/pZ)$ is such a polynomial algebra. Another well known example was $(Z/pZ)[x_{2n}]$ when $n|p-1$ with an action of the reduced powers which is essentially unique. A space was constructed by R. Holszager and D. Sullivan [22, 45] whose cohomology ring realises this polynomial algebra. When $n|p-1$, Z/nZ is a subgroup of $Z/(p-1)Z$, which is the group of units in the p-adic integers $\widehat{Z}_p$. So Z/nZ acts freely on a suitable Eilenberg–Maclane complex $K(\widehat{Z}_p, 2)$. Letting X denote the quotient space under this action, one has

$$H^*(X, Z/pZ) = (Z/pZ)[t]^{Z/nZ} = (Z/pZ)[x_{2n}].$$

A. Clark and J. Ewing [11] observed in 1974 that this construction can be generalised. If G is a finite subgroup of $GL_n(\widehat{Z}_p)$, it acts freely on $K(\widehat{Z}_p^n, 2) \times EG$ in the standard manner, where EG is a contractible free G-space. Denoting the quotient space by X_G, $H^*(X_G, Z/pZ) = (Z/pZ)[t_1, t_2, \ldots, t_n]^G$. They were able to characterise all G which gave rise to a polynomial algebra, when the order of the group G was prime to p, as p-adic reflection groups. They also classified the p-adic reflection groups, using an earlier classification of complex reflection groups by G. Sheppard and J. Todd into 37 irreducible families. In the converse direction, Adams and Wilkerson developed in [3] a Galois-theoretic approach for polynomial algebras over the Steenrod algebra at odd primes. It was deduced that provided p is prime to the degree of each polynomial generator x_i in $(Z/pZ)[x_1, x_2, \ldots, x_n]$, then this latter is one of the polynomial algebras realised by the Clark–Ewing construction. Later work of W. Dwyer, H. Miller and C. Wilkerson [14] showed that there was a unique p-complete space realising these polynomial algebras, which had therefore to be the Clark–Ewing example. (The condition of p-completeness is certainly essential, [35].)

The work has been extended to the modular case, that is when the degree of some generator of the polynomial algebra is divisible by p. Dwyer, Miller and Wilkerson have shown that provided there exists a space Y with $H^*(Y, Z/pZ) = (Z/pZ)[x_1, x_2, \ldots, x_n]$, then this polynomial algebra is again the ring of invariants of a p-adic reflection group. The hypothesis here is stronger than just requiring that there is an action of the Steenrod algebra. However, the Clark–Ewing construction does not realise these examples. Other techniques have been applied successfully to realise examples. A key development was provided by S. Jackowski and J. McClure (later with B. Oliver) who showed how to construct classifying spaces for compact Lie groups from the classifying spaces of appropriate proper subgroups [26]. This construction was generalised by Dwyer and Wilkerson to find examples in the modular case. This takes us towards the topic of p-compact Lie groups [15], which without doubt has been the most exciting development in H-space theory for many years. However, it is certainly not history, and so we refer to [38] for a short survey of recent results. Here we mention just one theorem which is relevant to the contents of Section 2. There exists a 2-complete loop space X with $H^*(X, Z/2Z) = (Z/2Z)[x_7, x_{11}, x_{13}]/(x_7^4, x_{11}^2, x_{13}^2)$, [16]. One can replace the 2-complete space by a 2-local space and using homotopy mixing techniques, construct a simply-connected finite H-space with this same cohomology ring. It is certainly not an example known classically and is essentially the only new mod 2 finite H-space which has been constructed.

Bibliography

The references below are those cited in the text. In many cases these represent only samples of the work by the author or authors on the topics. A comprehensive list of references up to the mid 1980s can be found at the end of [32].

[1] F. Adams, *On the non-existence of elements of Hopf invariant one*, Ann. of Math. **72** (2) (1960), 20–104.

[2] F. Adams, *The sphere, considered as an H-space mod p*, Quart. J. Math. Oxford Ser. (2) **12** (1961), 52–60.

[3] F. Adams and C. Wilkerson, *Finite H-spaces and algebras over the Steenrod algebra*, Ann. of Math. **111** (2) (1980), 95–143.

[4] M. Arkowitz and C. Curjel, *On the number of multiplications of an H-space*, Topology **2** (1963), 205–209.

[5] A. Borel, *Sur la cohomologie des espaces fibrés principaux et des espaces homogènes des groupes de Lie compacts*, Ann. of Math. **57** (2) (1953), 115–207.

[6] W. Browder, *The cohomology of covering spaces of H-spaces*, Bull. Amer. Math. Soc. **65** (1959), 140–141.

[7] W. Browder, *Torsion in H-spaces*, Ann. of Math. **74** (2) (1961), 24–51.

[8] W. Browder, *Homotopy commutative H-spaces*, Ann. of Math. **75** (2) (1962), 283–311.

[9] W. Browder, *Higher torsion in H-spaces*, Trans. Amer. Math. Soc. **108** (1963), 353–375.

[10] W. Browder and E. Spanier, *H-spaces and duality*, Pacific J. Math. **12** (1962), 411–414.

[11] A. Clark and J. Ewing, *The realization of polynomial algebras as cohomology rings*, Pacific J. Math. **50** (1974), 425–434.

[12] A. Copeland Jr., *Binary operations on sets of mapping classes*, Michigan Math. J. **6** (1959), 7–23.

[13] A. Dold and R. Lashof, *Principal quasi-fibrations and fibre homotopy equivalence of bundles*, Illinois J. Math. **3** (1959), 285–305.

[14] W.G. Dwyer, H.R. Miller and C.W. Wilkerson, *Homotopical uniqueness of classifying spaces*, Topology **31** (1992), 29–45.

[15] W.G. Dwyer and C.W. Wilkerson, *Homotopy fixed-point methods for Lie groups and finite loop spaces*, Ann. of Math. **139** (2) (1994), 395–442.

[16] W. Dwyer and C. Wilkerson, *A new finite loop space at the prime two*, J. Amer. Math. Soc. **6** (1993), 37–64.

[17] J. Harper, *H-spaces with torsion*, Mem. Amer. Math. Soc. **22** (1979), 223.

[18] P. Hilton, *Homotopy Theory and Duality*, Gordon and Breach Science Publishers, New York (1965).

[19] P. Hilton and J. Roitberg, *On principal S^3-bundles over spheres*, Ann. of Math. **90** (2) (1969), 91–107.

[20] P. Hilton, G. Mislin and J. Roitberg, *Localization of Nilpotent Groups and Spaces*, North-Holland Mathematics Studies vol. 15. Notas de Matematica, No. 55. North-Holland, Amsterdam, Oxford; American Elsevier, New York (1975).

[21] L. Hodgkin, *On the K-theory of Lie groups*, Topology **6** (1967), 1–36.

[22] R. Holzsager, *H-spaces of category $\leqslant 2$*, Topology **9** (1970), 211–216.

[23] H. Hopf, *Über die Topologie der Gruppen-Mannigfaltigkeiten und ihre Verallgemeinerungen*, Ann. of Math. **42** (2) (1941), 22–52.
[24] J. Hubbuck, *On homotopy commutative H-spaces*, Topology **8** (1969), 119–126.
[25] J. Hubbuck, *Two examples on finite H-spaces. Geometric applications of homotopy theory*, Proc. Conf., Evanston, Ill., 1977, I, Lecture Notes in Math. vol. 657, Springer, Berlin (1978), 282–291.
[26] S. Jackowski and J. McClure, *Homotopy decomposition of classifying spaces via elementary abelian subgroups*, Topology **31** (1992), 113–132.
[27] I. James and I.M. James, *On homotopy-commutative*, Topology **6** (1967), 405–410.
[28] I. James, *On H-spaces and their homotopy groups*, Quart. J. Math. Oxford Ser. (2) **11** (1960), 161–179.
[29] I. James, *Multiplication on spheres. II*, Trans. Amer. Math. Soc. **84** (1957), 545–558.
[30] I. James, *Multiplication on spheres. I*, Proc. Amer. Math. Soc. **8** (1957), 192–196.
[31] R. Kane, *Implications in Morava K-theory*, Mem. Amer. Math. Soc. **59** (1986), 340.
[32] R. Kane, *The Homology of Hopf Spaces*, North-Holland Mathematical Library vol. 40, North-Holland, Amsterdam, New York (1988).
[33] J. Lin, *Torsion in H-spaces. I*, Ann. of Math. **103** (2) (1976), 457–487.
[34] J. Lin, *Torsion in H-spaces. II*, Ann. Math. **107** (2) (1978), 41–88.
[35] C. McGibbon and J. Møller, *On spaces with the same n-type for all n*, Topology **32** (1992), 177–201.
[36] J. Milnor and J. Moore, *On the structure of Hopf algebras*, Ann. of Math. **81** (2), 211–264.
[37] G. Mislin, *The genus of an H-space*, Symposium on Algebraic Topology, Battelle Seattle Res. Center, Seattle, Washington, 1971, Lecture Notes in Math. vol. 249, Springer, Berlin (1971), 75–83.
[38] J. Møller, *Homotopy Lie groups*, Bull. Amer. Math. Soc. **32** (1995), 413–428.
[39] D. Rector, *Loop structures on the homotopy type of S^3*, Symposium on Algebraic Topology, Battelle Seattle Res. Center, Seattle, Washington, 1971, Lecture Notes in Math. vol. 249, Springer, Berlin (1971), 99–105.
[40] J. Slifker, *Exotic multiplications on S^3*, Quart. J. Math. Oxford Ser. (2) **16** (1965), 322–359.
[41] J. Stasheff, *Homotopy associativity of H-spaces. I, II*, Trans. Amer. Math. Soc. **108** (1963), 275–292; ibid. **108** (1963), 293–312.
[42] J. Stasheff, *H-spaces from a Homotopy Point of View*, Lecture Notes in Math. vol. 161, Springer, Berlin, New York (1970).
[43] J. Stasheff, *Manifolds of the homotopy type of (non-Lie) groups*, Bull. Amer. Math. Soc. **75** (1969), 998–1000.
[44] M. Sugawara, *On a condition that a space is an H-space*, Math. J. Okayama Univ. **6** (1957), 109–129.
[45] D. Sullivan, *Geometric Topology*, Mimeographed Notes, Massachusetts Institute of Technology (1970).
[46] E. Thomas, *Steenrod squares and H-spaces*, Ann. of Math. **77** (2) (1963), 306–317.
[47] E. Thomas, *Steenrod squares and H-spaces. II*, Ann. of Math. **81** (2) (1965), 473–495.
[48] A. Zabrodsky, *Secondary cohomology operations in the module of indecomposables*, Proc. of the Advanced Study Institute on Algebraic Topology Vol. III (1970), 657–672. 13, Mat. Inst., Aarhus Univ., Aarhus (1970).
[49] A. Zabrodsky, *Homotopy associativity and finite CW complexes*, Topology **9** (1970), 121–128.
[50] A. Zabrodsky, *Hopf Spaces*, North-Holland Mathematics Studies vol. 22. Notas de Matemática, No. 59. North-Holland, Amsterdam, New York, Oxford (1976).

This Page Intentionally Left Blank

CHAPTER 27

A History of Rational Homotopy Theory

Kathryn Hess

Département de Mathématiques, Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland
E-mail: kathryn.hess@epfl.ch

Introduction

Rational homotopy theory is the study of topological spaces "modulo torsion". When rational homotopy theorists calculate algebraic invariants of a topological space, such as its homotopy or homology groups, they retain only the nontorsion information. Algebraic models of topological spaces are the most important tools of the rational homotopy theorist.

To understand more precisely the objects studied in rational homotopy theory, recall that a space E is *rational* if its homotopy groups $\pi_*(E)$ form a graded rational vector space. Any simply-connected or nilpotent space E has a *rationalization*, consisting of a rational space E_0 together with a continuous map $\varphi : E \to E_0$ inducing an isomorphism on rational homotopy groups

$$\pi_*(\varphi) \otimes \mathbb{Q} : \pi_*(E) \otimes \mathbb{Q} \to \pi_*(E_0) \otimes \mathbb{Q}.$$

Two spaces are *rationally homotopy equivalent* if their rationalizations are homotopy equivalent. Rational homotopy theory is the study of topological spaces up to rational homotopy equivalence.

Rational homotopy theory is a relatively young branch of algebraic topology, founded by D. Quillen and D. Sullivan at the end of the 1960's. To understand what inspired Quillen and Sullivan, however, it is necessary to dig deep into the past of topology, as far back as the work of H. Poincaré at the end of the last century, which is where this history begins.

In the first section of this history we trace the developments in algebraic topology which prepared the ground for the foundation of rational homotopy theory. The guiding principle behind these developments, implicit in the work of Poincaré and explicitly stated in that of E. Cartan and G. de Rham, is that one can study the algebraic topology of a differentiable manifold via its differential forms. Application of this principle led in the 1950's to the first utilisation of algebraic models for solving topological problems.

We devote the second section to the seminal work of Quillen and Sullivan that laid the foundations of rational homotopy theory, as well as to the first important applications of

HISTORY OF TOPOLOGY
Edited by I.M. James

their work. Quillen proved that the homotopy category of simply-connected rational spaces was equivalent to the homotopy category of chain Lie algebras over the field of rational numbers, $\mathbb{Q}$. Motivated by the guiding principle mentioned above, Sullivan defined a one-to-one correspondence between rational homotopy types of simply-connected spaces and certain free commutative cochain algebras. Their work established the theoretical basis for the use of algebraic models in solving problems in rational homotopy theory.

In the third section we meet the artisans who extended and refined the tools of rational homotopy theory provided by Quillen and Sullivan and who proved the fundamental structure theorems necessary to its further development. During almost a decade, rational homotopy theory expanded and evolved rapidly, though it remained in "splendid isolation" from much of the rest of mathematics.

As we explain in Section 4, rational homotopy theorists formed dynamic collaborations in the early 1980's with local algebraists, leading to remarkable results in both fields. The techniques learned through these collaborations enabled rational homotopy theorists then to expand further into mod p homotopy theory, where they have made substantial contributions to the study of the homology of loop spaces.

Those interested in learning more about the methods and scope of rational homotopy theory can refer to the books of Griffiths and Morgan [54] and of Halperin [65], as well as to the monograph in the *Astérisque* series of Lehmann [80]. Moreover, a self-contained, complete introduction to rational homotopy theory and its applications by Félix, Halperin and Thomas is available in preprint form [44].

I hope that this history of rational homotopy theory will please all of those directly concerned at least some of the time, though I dare not hope to please even some of them all of the time. I extend heartfelt thanks to all who took the time to share their knowledge of and experience with rational homotopy theory with me: Luchezar Avramov, Hans Baues, Ed Brown, Yves Félix, Pierre-Paul Grivel, André Haefliger, Steve Halperin, Peter Hilton, Daniel Lehmann, Jean-Michel Lemaire, Clas Löfwall, John McCleary, Jan-Erik Roos, Jim Stasheff, Dennis Sullivan, Daniel Tanré, Jean-Claude Thomas, and Micheline Vigué.

Terminology and notation

- $\mathbb{Q}$ denotes the field of rational numbers, while $\mathbb{R}$ denotes the field of real numbers.
- If M is a differentiable manifold, $\Omega^*_{\mathrm{DR}}(M)$ denotes the de Rham complex of M.
- A *quasi-isomorphism* of differential graded objects is a graded morphism commuting with the differential that induces an isomorphism in (co)homology.
- A *(co)chain algebra* is a (co)chain complex (A, d) endowed with an associative product $\mu : (A, d) \otimes (A, d) \to (A, d)$ that is a map of (co)chain complexes. A (co)chain algebra is *commutative* if $a \cdot b = (-1)^{\deg a \deg b} b \cdot a$ for all $a, b \in A$.
  A free commutative cochain algebra generated by a positively graded vector space $V = \bigoplus_{i>0} V^i$ is denoted $(\Lambda V, d)$, where ΛV is the tensor product of the polynomial algebra on V^{even} and the exterior algebra on V^{odd}. We denote by $\Lambda^i V$ the elements of ΛV of wordlength i and by d_i the summand of the differential d increasing wordlength by exactly $i - 1$.
- A *chain Lie algebra* over a subring of $\mathbb{Q}$ containing $\frac{1}{2}$ consists of a positively graded chain complex (L, d) together with a product denoted [,] such that

$$[a, b] = -(-1)^{\deg a \deg b}[b, a]$$

for all $a, b \in L$, $[a, [a, a]] = 0$ for all $a \in L_{\mathrm{odd}}$, and

$$(-1)^{\deg a \deg c}[a, [b, c]] + (-1)^{\deg b \deg a}[b, [c, a]] + (-1)^{\deg c \deg b}[c, [a, b]] = 0$$

for all $a, b, c \in L$.

1. The vanguard

When Quillen and Sullivan laid the foundations of rational homotopy theory at the end of the 1960's, they were inspired by the evolution in thinking about the algebraic topology of manifolds that had occurred since the end of the 19th century. In this section I describe this evolution and explain its importance in the foundation of rational homotopy theory.

In preparing this section, I benefited greatly from reading Sullivan's article [102], Haefliger's article [62], and the Notes at the end of the book of Greub, Halperin and Vanstone [52], as well as from conversations with John McCleary and a long letter from André Haefliger.

1.1. *Establishing the foundations*

Henri Poincaré made one of the first, crucial steps towards rational homotopy theory when he studied the commutative cochain algebra structure of the differential forms on a manifold in his 1895 article *Analysis situs* [89]. Therein he explained that one can obtain information about the topology of a manifold M by considering the exterior algebra generated by the 1-forms on M together with the usual exterior differential, i.e. what is now known as the de Rham complex of the manifold, denoted $\Omega^*_{\mathrm{DR}}(M)$.

Poincaré applied this idea to understanding the Betti numbers of manifolds. At that time, the kth Betti number of an m-dimensional manifold M, denoted here $\beta_k(M)$, was defined to be the greatest integer n for which there exist k-dimensional submanifolds $M_1, \ldots, M_{n-1}$ of M such that the boundary of no $(k+1)$-dimensional submanifold is composed of copies of all of the M_i's. In other words, $\beta_k(M) - 1$ is the maximal number of k-dimensional submanifolds that are "linearly independent" over the integers, i.e. no integral "linear combination" of the submanifolds is homologous to zero.

Poincaré showed that $\beta_k(M) - 1$ was an upper bound for the dimension of the image of the homomorphism induced by integration from the cohomology of $\Omega^*_{\mathrm{DR}}(M)$ to (something resembling) the real singular cohomology of M. He referred to the elements of a basis of this image as "periods". He indicated, furthermore, that the number of periods should in fact be equal to $\beta_k(M) - 1$ but did not prove this equality explicitly.

Inspired by Poincaré's work, E. Cartan conjectured in 1928 that the homomorphism induced by integration described in the previous paragraph was actually an isomorphism. Based on this conjecture, he then proved that the first two Betti numbers of a compact, semisimple Lie group G are zero, while its third Betti number is necessarily nonzero [26]. The key to his proof is the important observation that the cohomology of $\Omega^*_{\mathrm{DR}}(G)$ is the same as that of its subalgebra generated by left-invariant 1-forms.

The following year de Rham proved Cartan's conjecture, which has been known ever since as *de Rham's theorem.* He announced this result in [32], providing details of the proof in [33]. De Rham's theorem was absolutely crucial to the future development of rational homotopy theory. As we explain in Section 2, de Rham's theorem provides the link between differential forms and algebraic topology that allows the construction of commutative cochain algebra models for spaces over the rationals.

Also in 1929 Cartan used de Rham's theorem to generalize his earlier results, showing that if M is a symmetric homogeneous space acted upon by a connected, compact Lie group G, then the real cohomology of M is isomorphic as an algebra to the algebra of G-invariant differential forms on M [27]. Thus the calculation of the real cohomology of M reduces to a calculation of invariants.

In 1939 Heinz Hopf continued the investigation of the cohomology of Lie groups, showing that the real cohomology algebra of a compact, connected Lie group G is always an exterior algebra on odd generators [75]. His proof relies on the observation that $H^*(G; \mathbb{R})$ is in fact a Hopf algebra, which imposes strong restrictions on its algebra structure. Shortly thereafter, H. Samelson applied Hopf's methods to studying the real cohomology of homogeneous spaces [95].

1.2. *The first explicit algebraic models*

The second phase of the evolution that led to the inception of rational homotopy theory began at the end of the 1940's, when several prominent topologists became very interested in algebraic models for topological spaces over the reals. Given a certain class of spaces, e.g., connected, compact Lie groups or total spaces of principal bundles, their goal was to find a formula for commutative cochain algebras whose cohomology algebras were isomorphic to the real cohomology of spaces belonging to the given class.

Guy Hirsch. Hirsch was one of the first topologists to experiment with algebraic models of spaces, in a series of articles published between 1948 and 1956, of which the most important are probably [72] from 1948, [73] from 1950, and [74] from 1953. His goal was to determine the algebra structure of the cohomology of the total space of a fiber space, given the cohomology algebras of the base and the fiber.

His first step towards this goal consisted in the definition of a degree +1 "characteristic isomorphism" from a quotient of a subgroup of the cohomology of the fiber onto a quotient of a subgroup of the cohomology of the base [72]. In modern language, this is essentially the transgression map. He then claimed that this "characteristic isomorphism" together with the cohomology algebras of the base and the fiber should be enough to determine the cohomology of the total space.

In his paper in the Proceedings of the 1950 ICM [73], Hirsch showed more clearly how one could obtain the additive structure and part of the multiplicative structure of the cohomology of the total space from the "characteristic isomorphism" together with the cohomology algebras of the base and the fiber. In addition he explained what sort of extra information about the fiber bundle might allow one to compute the entire algebra structure of the cohomology of the total space, e.g., the characteristic classes of a sphere bundle over the orthogonal or unitary group as determined by the multiplicative structure of the auxiliary bundle.

It is probably in his 1953 paper [74] that Hirsch worked most explicitly with algebraic models. Therein he considered Serre fibrations with path-connected base B and fiber F of finite type such that the fundamental group of the base acts trivially on the cohomology of the fiber. He showed that over a field $\mathbb{K}$, one can define an extension of the singular cochains on the base,

$$\big(C^*(B;\mathbb{K}), d\big) \hookrightarrow \big(C^*(B;\mathbb{K}) \otimes H^*(F;\mathbb{K}), D\big),$$

together with a morphism of cochain complexes inducing an isomorphism in cohomology

$$\big(C^*(B;\mathbb{K}) \otimes H^*(F;\mathbb{K}), D\big) \to C^*(E;\mathbb{K}). \tag{1}$$

Claude Chevalley and Samuel Eilenberg. Chevalley and Eilenberg provided further impetus for the interest in algebraic modelling with their important 1948 paper [35], in which they used de Rham's theorem to show that, over a field of characteristic zero, the cohomology of a compact Lie group is isomorphic as an algebra to the cohomology of the corresponding Lie algebra. They defined the cohomology of a Lie algebra $\mathfrak{g}$ to be that of a certain cochain algebra $C^*(\mathfrak{g})$ whose underlying algebra is the exterior algebra on the dual of $\mathfrak{g}$, denoted $\hat{\mathfrak{g}}$. The formula for the differential of $C^*(\mathfrak{g})$, now well known, came from pushing forward formulas that were already known for the differential on the left-invariant forms on a compact Lie group.

As applications of the isomorphism between the cohomology of a Lie group G and that of its associated Lie algebra $\mathfrak{g}$, Chevalley and Eilenberg generalized the results of Cartan on Betti numbers and of Hopf on the algebra structure of the cohomology to semisimple Lie groups. They showed as well that the cohomology of a semisimple Lie algebra is isomorphic to the algebra of *ad*-invariant cochains in $C^*(\mathfrak{g})$.

Chevalley and Eilenberg further defined the relative complex $C^*(\mathfrak{g}, \mathfrak{h})$ of a pair of Lie algebras $\mathfrak{h} \subset \mathfrak{g}$. They showed that the cohomology of G/H, where G is compact and connected and H is a closed, connected subgroup of G, is isomorphic to the cohomology of $C^*(\mathfrak{g}, \mathfrak{h})$, where $\mathfrak{g}$ and $\mathfrak{h}$ are the Lie algebras corresponding to G and H, respectively.

Jean-Louis Koszul. Also at the end of the 1940's, Koszul was working on his thesis, in which he intended to prove the theorems of Hopf and Samelson by purely algebraic means [78]. With this goal in mind, he, too, defined a cochain algebra for calculating the cohomology of a Lie algebra, though over a field of any characteristic other than 2. Its underlying algebra was again the exterior algebra on the dual of the Lie algebra, and its differential differed from that of the cochain algebra of Chevalley and Eilenberg only up to multiplication by an integer. He showed that also with this somewhat modified definition, the real cohomology of a compact Lie group was isomorphic to that of its associated Lie algebra, which enabled him to reprove Hopf's theorem.

In writing his thesis Koszul also studied the relative complex $C^*(\mathfrak{g}, \mathfrak{h})$ of a pair of Lie algebras $\mathfrak{h} \subset \mathfrak{g}$, defined, as in the paper of Chevalley and Eilenberg, to be the subcomplex of $C^*(\mathfrak{g})$ consisting of those cochains annihilated by interior products and Lie derivatives by elements of $\mathfrak{h}$. Using this definition he reproved the results of Samelson on homogeneous spaces.

Inspired by Hirsch's work on the cohomology of fiber spaces [72], Koszul introduced the notion of *transgression* in the context of pairs of Lie algebras. Let $\mathfrak{h} \subset \mathfrak{g}$ be a pair of

Lie algebras. Then a class $\alpha \in H^k(\mathfrak{h})$ is *transgressive* if it has a cochain representative $a \in C^k(\mathfrak{h})$ that is the restriction to $\mathfrak{h}$ of a cochain $b \in C^k(\mathfrak{g})$ such that $db \in C^{k+1}(\mathfrak{g}, \mathfrak{h})$. The *transgression* of the class α is then the class $[db] \in H^{k+1}(\mathfrak{g}, \mathfrak{h})$. Koszul proved that every primitive element of $H^*(\mathfrak{h})$ is transgressive.

Henri Cartan. The next major advances in algebraic modelling appeared in the two papers by H. Cartan and in the paper by Koszul based on talks that they gave at the 1950 Colloque de Topologie in Brussels [28, 29, 79]. The two papers by Cartan, in which he presented work done both alone and in collaboration with Weil, contain an astounding wealth of information; the review in Mathematical Reviews written by Chevalley fills more than two pages! In the paragraphs below, I attempt to present a succinct outline of these two important and imposing papers, while still doing justice to their depth.

Cartan first considered smooth principal bundles with total space E and base B over a Lie group G with associated Lie algebra $\mathfrak{g}$. He defined a form on E to be *invariant* if it was annihilated by $ad(x)$, for all $x \in \mathfrak{g}$ and to be *basic* if, in addition, it was annihilated by all interior products. He then generalized the result of Chevalley and Eilenberg and Koszul on the cohomology of Lie groups, proving that the inclusion of the invariant forms into $\Omega^*_{\mathrm{DR}}(E)$ induces an isomorphism of algebras in cohomology.

Since Cartan's goal in these papers was to reduce topological problems concerning principal bundles to pure algebra, he needed an algebraic equivalent of the bundle connection, introduced by E. Cartan and studied by Weil and Ehresmann. He defined an *algebraic connection* of a principal G-bundle to be a linear map f from $\hat{\mathfrak{g}}$ into the 1-forms on E, commuting with all interior products and $ad(x)$-derivations, for all $x \in \mathfrak{g}$. Algebraic connections exist for all smooth principal bundles, according to results of Weil and Ehresmann. An algebraic connection f can always be extended to a map of algebras $\bar{f} : C^*(\mathfrak{g}) \to \Omega^*(E)$, though $\bar{f}$ will usually not commute with the differentials. The deficiency $d\bar{f} - \bar{f}d$ is the *curvature tensor.*

Given any Lie group G with corresponding Lie algebra $\mathfrak{g}$, Cartan constructed an extension of cochain algebras

$$C^*(\mathfrak{g}) = \big(\Lambda(\hat{\mathfrak{g}}), d\big) \hookrightarrow \big(\Lambda(\hat{\mathfrak{g}}) \otimes P[s\hat{\mathfrak{g}}], D\big) = W(G),$$

where Λ is the exterior algebra functor, P the polynomial algebra functor and s the suspension. The cochain algebra $W(G)$ is called the *Weil algebra* of G. He proved that for all algebraic connections $f : \hat{\mathfrak{g}} \to \Omega^1(E)$, the linear map

$$\tilde{f} = \bar{f} \otimes \tilde{f} : W(G) \to \Omega^*(E),$$

where $\tilde{f}(sx) = (d\bar{f} - \bar{f}d)(x)$ was actually a map of cochain algebras. Furthermore, $\mathfrak{g}$ acts on $W(G)$ by both interior products and ad-derivations, in a way compatible with $\tilde{f}$. Cartan observed that $W(G)$ is a sort of universal cochain algebra for principal fibrations over G.

Cartan's first application of the Weil algebra was to finding the characteristic classes of a principal bundle. He observed that the basic elements of $W(G)$ were exactly the invariant elements of $P[s\hat{\mathfrak{g}}]$, which are all cocycles. In addition, since $\tilde{f}$ is compatible with both types of derivations, it must send basic elements to basic elements, i.e. the basic elements of $W(G)$ are sent to $\Omega^*_{\mathrm{DR}}(B)$ by $\tilde{f}$. The images of the basic elements of $W(G)$ in $H^*(B;\mathbb{R})$ are the *characteristic classes* of the fiber bundle.

Cartan next applied the Weil algebra to generalizing the notion of transgression. Following Koszul's example, he defined an element of $H^*(\mathfrak{g})$ to be *transgressive* if it had an invariant representative that was the image under projection of an invariant element in $W(G)$ with basic coboundary. Since $W(G)$ is acyclic, he could then define a transgression map from the transgressive elements of $H^*(\mathfrak{g})$ to the algebra of basic elements in $W(G)$, sending a transgressive element to the coboundary of an element in the preimage of the projection from the invariants of $W(G)$ onto the invariants of $C^*(\mathfrak{g})$. Based on this definition, Cartan and Chevalley proved an extension of Koszul's theorem that had been conjectured by Weil, showing that if $\mathfrak{g}$ is reductive (i.e. the direct product of an Abelian Lie algebra and a semisimple Lie algebra), then the transgressive elements of $H^*(\mathfrak{g})$ are exactly the primitives.

The last, important results from Cartan's papers mentioned here concern explicit algebraic models that he constructed for the total space E and the base space B of a smooth principal G-bundle, when $\mathfrak{g}$ is reductive. His first such construction, of a model for E when G is compact and connected, takes as input a given transgression map and algebraic connection. For these data he produced a formula for a cochain algebra extension of $\Omega^*_{\mathrm{DR}}(B)$ by the invariants of $C^*(\mathfrak{g})$, together with a quasi-isomorphism to $\Omega^*_{\mathrm{DR}}(E)$.

Drawing on results of Hirsch, Cartan also constructed a model for B when G is connected and compact or when E is finite dimensional and $\mathfrak{g}$ is reductive. In this case he defined a cochain algebra extension of the algebra of basic elements in $W(G)$ by $H^*(E;\mathbb{R})$ that had the same cohomology algebra as $\Omega^*_{\mathrm{DR}}(B)$. As a corollary, Cartan showed that if G is compact and H a closed, connected subgroup of G, then a certain cochain algebra extension of the algebra of basic elements in $W(H)$ by the invariant elements in $C^*(\mathfrak{g})$ is a model of G/H.

Cartan only sketched the proofs of many of his results in [26, 27]. In his thorough and precise article of 1962 [6], Michel André filled in all of the missing details.

Also well worth perusing is Koszul's paper from the 1950 Colloque de Topologie, from which the following points are relevant to our desire to understand the evolution of rational homotopy theory. The first is Koszul's explanation of an algebraic theorem, due to Chevalley, which implies that the cohomology algebra of the total space of a principal fibration is entirely determined by the transgression and the structure of the base.

Second, it is important to note that Koszul defined his famous resolution in this paper. Given a smooth principal G-bundle, where G is compact and connected, let $x_1,\ldots,x_l$ form a basis of the primitives of $H^*(G;\mathbb{R})$. The de Rham complex of B is then a differential module over the polynomial ring on generators of degree 2, $S = P[c_1,\ldots,c_l]$, which is exactly the real cohomology of the classifying space BG. The Koszul resolution of $H^*(B;\mathbb{R})$ is then the cochain algebra extension $(H^*(B;\mathbb{R})\otimes\Lambda(x_1,\ldots,x_l), D)$, where $D(b\otimes x_i) = bc_i\otimes 1$ for all i. The cohomology of this cochain algebra is isomorphic to $H^*(E;\mathbb{R})$.

René Thom. According to Haefliger, one could argue that Thom's article in the 1954/1955 Cartan seminar [108] represents the "first conscious step towards rational homotopy theory". In that article Thom considered pull-back diagrams

$$\begin{array}{ccc} Q_f & \longrightarrow & P_*K(\mathbb{R}, n) \\ \downarrow & & \downarrow p \\ E & \xrightarrow{f} & K(\mathbb{R}, n) \end{array}$$

in which $K(\mathbb{R}, n)$ is a real Eilenberg–Mac Lane space, and p is the (based) path-space fibration. One important difference between the problem Thom considered and those studied by Chevalley, Eilenberg, Cartan, Koszul, and Weil is that Thom no longer required spaces to be smooth or even differentiable.

Thom's goal was to compute the real cohomology algebra of Q_f in terms of the real cohomology algebra of E and the real homotopy type of f. He explained that one approach to solving this problem would be to apply Hirsch's theory and outlined how to proceed. He noted that there remained one major obstacle to carrying out his plan: Hirsch's morphism (1) is not guaranteed to induce an isomorphism of *algebras* unless there are *commutative* cochain algebras that can be used to compute the cohomologies of all the spaces involved. Thom conjectured that, at least over $\mathbb{R}$ and for a fairly large class of spaces, such as CW-complexes, it should be possible to find such cochain algebras, noting that the result was known for finite polyhedra, as they are retracts of differentiable manifolds, for which one can take the de Rham complex. On the other hand, even then not all the difficulties would be resolved, since Hirsch relied on the Leray spectral sequence of a fibration to prove that his morphism induced an isomorphism, and the Leray spectral sequence applied only to singular cochains.

Thom had put his finger on the central, foundational problem of rational homotopy theory: how to associate canonically to each "nice" topological space a commutative cochain algebra whose cohomology algebra is isomorphic to that of the space, so that the association allows one to establish an equivalent of Hirsch's theory. This is what Quillen later referred to as "Thom's problem".

Another paper by Thom along similar lines appeared in the proceedings of the 1956 Colloque de Topologie in Louvain [109]. In that paper he investigated the homotopy groups of mapping spaces, proposing again to use Hirsch's theory extended to multiplicative structure. In particular, he proposed an algorithm for construction of a cochain algebra describing the rational homotopy type of the space of maps with fixed source and target, homotopic to a given map. Haefliger later implemented this algorithm in order to verify a claim of Sullivan; see Section 3.2.

In the summer of 1957, Thom taught a course at the University of Chicago in which he solved the problem of his 1954 paper over $\mathbb{R}$. He established the existence for any space E of a commutative cochain algebra over $\mathbb{R}$, the cohomology algebra of which is isomorphic to $H^*(E; \mathbb{R})$. His proof was based on verifying the Eilenberg–Steenrod axioms on the cohomology of the model. Almost 20 years later Swan generalized Thom's construction so that it applied also over $\mathbb{Q}$; see Section 2.3.

Frank Adams and Peter Hilton. Adams and Hilton also made remarkable contributions to algebraic modelling in the mid 1950's, though of a different nature than the work cited above. Given a CW-complex E with one 0-cell and no 1-cells, they showed how to construct a free, associative (tensor) chain algebra (A, d), together with a quasi-isomorphism of chain algebras $(A, d) \to CU_*(\Omega E)$, where CU_* denotes the 1-reduced cubical chains [2]. The generators of the free algebra A are in the one-to-one correspondence with the cells of E, while the differential d encodes the attaching maps quite explicitly. Thus, the Adams–Hilton model has the advantage of being both small, i.e. with relatively few generators, and explicitly defined. Detailed computations of the homology of loop spaces via the Adams–Hilton model are therefore entirely possible. The Adams–Hilton model has proven to be one of the most useful tools in rational homotopy theory.

Shortly after his work with Hilton, Adams defined another important algebraic model for studying loop spaces [1]. For any simply-connected space E, he showed that there was a quasi-isomorphism of chain algebras from the cobar construction on C_*E, the singular chains on E, to $C_*(\Omega E)$. Adams's cobar model has also proven to be of great use in rational homotopy theory.

John Tate. Among the algebraic model builders of this period we should include the algebraist Tate, whose research during the 1950's led him to construct algebraic models similar to those constructed by the topologists mentioned above. In 1957 Tate published a paper in which he calculated $\mathrm{Tor}^R(R/M, R/N)$, where R is a commutative Noetherian ring and R/M and R/N are residue class rings [107]. His calculation involved constructing free R-resolutions of R/M and R/N that were actually free, commutative cochain algebras over R. The construction proceeded by inductive, degreewise adjunction of new variables. As Tate himself mentioned in his article, this method was well known to the topological model builders of the time, since it was based on the Postnikov decomposition of spaces initiated by John Moore. The connection between commutative algebra and algebraic modelling of topological spaces is a recurrent theme throughout the history of rational homotopy theory, as later sections of this history confirm; see, in particular, Section 4.2.

1.3. *Further algebraic groundwork*

Though not directly concerned with algebraic modelling of spaces, the work of Milnor and Moore, as well as that of Serre, in the mid 1950's strongly influenced the genesis and development of rational homotopy theory.

John Milnor and John Moore. The monumental paper of Milnor and Moore [87], which began circulating in preprint form in the middle of the 1950's, has played a decisive role in the development of rational homotopy theory. The existence of a precise, complete description of graded algebras, coalgebras, Lie algebras and Hopf algebras, as well as of the universal enveloping algebra functor U, most probably facilitated Quillen's work on establishing an equivalence between the homotopy category of rational spaces and that of chain Lie algebras over $\mathbb{Q}$. It has also proved of crucial importance in the last decade, since rational homotopy theorists began applying their techniques to the study of the homology Hopf algebra of loop spaces over a field of positive characteristic; see Section 4.3.

The last theorem in the paper, stating that if G is a path-connected, homotopy associative H-space then the Hurewicz homomorphism $\pi_*(G) \otimes \mathbb{Q} \to H_*(G; \mathbb{Q})$ induces an isomorphism of Hopf algebras

$$U(\pi_*(G) \otimes \mathbb{Q}) \to H_*(G; \mathbb{Q}),$$

has inspired much research in rational homotopy theory; see, for example, Section 4.3. In explaining the proof of this theorem, Milnor and Moore pointed out that the k-invariants of the Postnikov decomposition of a rational H-space were all zero, a remark that ties in with the work of Sullivan, whose algebraic models of rational spaces are based on Postnikov decompositions; see Section 2.2.

Jean-Pierre Serre. Serre's 1953 paper on homotopy groups and classes of Abelian groups [96] was crucial to solidifying the foundations of rational homotopy theory, as he provided a precise definition of what it meant to "forget torsion" and study only the nontorsion homotopy and homology of a space.

Serre defined a *class* of Abelian groups to be a nonempty collection $\mathfrak{C}$ of Abelian groups such that

(1) the trivial group is in $\mathfrak{C}$;
(2) all subgroups and quotients of a group in $\mathfrak{C}$ are also in $\mathfrak{C}$; and
(3) every extension of two groups in $\mathfrak{C}$ is in $\mathfrak{C}$.

He then define various $\mathfrak{C}$-versions of notions from the usual theory of Abelian groups. For example, a homomorphism of Abelian groups is a *$\mathfrak{C}$-isomorphism* if its kernel and cokernel are in $\mathfrak{C}$. Rational homotopy theory is thus homotopy theory modulo the class of all finite Abelian groups, $\mathfrak{C}_f$; a weak rational homotopy equivalence is a continuous map inducing a $\mathfrak{C}_f$-isomorphism on homotopy groups.

Once he had defined his terminology, Serre proved "mod $\mathfrak{C}$" versions of the Hurewicz theorem and the Whitehead theorem, both of which are important tools to rational homotopy theorists. He then applied his "mod $\mathfrak{C}$"-theory to various topological problems. Of particular relevance to rational homotopy theory is his proof, based on Hopf's theorem [75] (see Section 1.1), that for any compact, connected semisimple Lie group G, there is a weak rational homotopy equivalence from G to a wedge of odd spheres.

1.4. *The geometric groundwork*

To be complete, a description of the research leading up to the foundation of rational homotopy theory should mention the work of Hassler Whitney on geometric integration theory. Whitney's beautiful book [110], in which he presented a clear, axiomatic approach to integration of forms over chains, played an important role in the development of Sullivan's approach to rational homotopy theory, which is based on a rational, piecewise-linear version of de Rham's theorem (cf. Section 2 for further details).

In the introduction to his book Whitney wrote that he wanted to answer the question "what should r-dimensional integration on n-space 'look like'?". Starting from the perception of an integral as an integrand evaluated on a domain, he gradually introduced stronger and stronger continuity-type conditions on the behavior of the integral. Under strong enough conditions, he defined the pointwise differential of the integrand and showed that the value of the integrand evaluated on a cell is equal to the integral over the cell of its pointwise differential.

Based on the input necessary to the definition of pointwise differentials, Whitney then introduced the exterior (Grassmann) algebra of vectors on n-space and its dual, the exterior algebra of covectors, or differential forms, on the ambient space. The pointwise differential of an integrand is the canonical example of a differential form.

Whitney further showed how a smooth mapping from n-space to m-space induces an homomorphism of algebras on the algebras of vectors, leading directly to natural definitions of the pullback of covectors and of the Jacobian. He explained, moreover, how Stokes' Theorem also fits naturally into this abstract context, giving rise to the differential on the algebra of forms.

One last, important contribution Whitney made in his book was in presenting a straightforward, vivid explanation of de Rham's theorem within the context of geometric integration theory.

1.5. *Postscript*

In the research outlined above, almost all of the groundwork necessary to the inception of rational homotopy theory had been carried out. Why then was there a more than ten-year hiatus between the last results mentioned above and the foundation of rational homotopy theory by Quillen? It is probable that the spectacular and promising advances of Armand Borel in the mid 1950's, translating into purely topological terms the algebraic methods of Cartan, Chevalley, Koszul and Weil, contributed greatly to diverting the attention of the topological community from the path of algebraic modelling.

2. The architects

After a quiescent period starting in the late 1950's, interest in algebraic modelling of topological spaces started to pick up again in the late 1960's. All the pieces of rational homotopy theory suddenly fell into place. Our goal in this section is to present the major innovators and their innovations of those crucial foundational years. In addition to the literature cited, I relied on long discussions with Steve Halperin and on a helpful letter from Dennis Sullivan in preparing this section.

2.1. *The renaissance of algebraic modelling*

Shortly before the actual blossoming of rational homotopy theory, algebraic models of topological spaces came back into fashion. To cite an important example, in an article published in 1970 [24], Raoul Bott applied the models of Cartan, Chevalley, Koszul and Weil for principal fibrations to determining the obstruction to complete integrability of planar vector fields.

Also in the mid 1960's, Werner Greub, Steve Halperin and Ray Vanstone began working on a three-volume study of the de Rham cohomology of smooth fiber bundles [52]. The first two volumes of the series cover the differential geometry of smooth manifolds and vector bundles, as well as the theory of Lie groups and principal bundles. The third volume is a careful, thorough exposition of the models of Cartan, Chevalley, Koszul and Weil. The techniques and terminology that Greub, Halperin and Vanstone elaborated in writing the third volume of their study proved to be most useful in the early stages of the development of rational homotopy theory.

2.2. *The innovations of Quillen and Sullivan*

In 1967 Daniel Quillen published *Homotopical Algebra*, presenting the topological and algebraic communities with a categorical, axiomatic framework in which to do homotopy theory [90]. In the introduction to *Homotopical Algebra*, he stated that his desire to compute the cohomology of commutative rings had led him to think of certain simplicial categories as possessing natural "homotopy theories", inspired by Kan's equivalence between the homotopy categories of topological spaces and of semisimplicial complexes. The similarity of the arguments he had to make in each context to conventional topological arguments motivated him to define a category-theoretic notion of homotopy theory to cover simultaneously and uniformly the different notions of "homotopy theory" known at the time.

Two years later Quillen demonstrated the tremendous power of his categorical framework for homotopy theory when he published *Rational homotopy theory* [91], the foundational paper in the subject. The main theorem in [91] states that the rational homotopy theory of simply-connected, pointed spaces is equivalent to the homotopy theory of chain Lie algebras over $\mathbb{Q}$, as well as to the homotopy theory of 2-reduced cocommutative chain coalgebras over $\mathbb{Q}$. Moreover, the equivalence is such that the homology of the chain Lie algebra corresponding to a given space E is isomorphic to the *homotopy Lie algebra* of E, i.e. $\pi_*(\Omega E) \otimes \mathbb{Q}$, endowed with the Samelson product. As a corollary Quillen obtained that every positively graded Lie algebra over $\mathbb{Q}$ was the homotopy Lie algebra of some space. He thus provided a positive answer to a question that he believed was due to Hopf.

Quillen showed as well that his equivalence of homotopy categories enabled him to solve Thom's problem of constructing a functor from spaces to commutative cochain algebras with the right cohomology algebra and the right properties with respect to fibrations. Indeed, the dual of the cocommutative chain coalgebra corresponding to a space E is a commutative cochain algebra whose cohomology is isomorphic as an algebra to $H^*(E;\mathbb{Q})$. Furthermore, since the equivalence between homotopy categories is induced by a functor on the categories themselves that preserves the necessary structure, in particular the fibrations, Quillen's functor from topological spaces to commutative cochain algebras behaves correctly, in Thom's sense, with respect to fibrations.

Further important results from [91] include the construction of a functor from the category of chain Lie algebras over $\mathbb{Q}$ to the category of 2-reduced cocommutative chain coalgebras over $\mathbb{Q}$ that generalizes the procedure for calculating the homology of a (non-graded) Lie algebra and that is adjoint to the functor given by taking the primitives of the cobar construction. Quillen also derived several interesting spectral sequences of Lie algebras and of coalgebras from his fundamental equivalence. For example, if $X \to Y$ is a cofibration of spaces, then there is a spectral sequence converging from $E^2 = (\pi_*(\Omega X) \otimes \mathbb{Q}) \oplus (\pi_*(\Omega(Y/X)) \otimes \mathbb{Q})$ to $\pi_*(\Omega Y) \otimes \mathbb{Q}$.

Quillen's paper represents a crucial, essential step in the development of rational homotopy theory, firmly establishing the theoretical justification of algebraic modelling as a tool for the investigation of the rational homotopy of spaces. On the other hand the proof of his main theorem, though constructive, had the disadvantage of providing unwieldy algebraic models. Performing actual calculations based on Quillen's models was impossible in practice.

At the beginning of the 1970's Dennis Sullivan bridged the calculability gap. Motivated by a desire to understand the "true nature" of a diffeomorphism class of compact smooth manifolds and inspired by both the algebraic models of the 1950's and Quillen's equivalence of homotopy categories, he defined a commutative cochain algebra model over any field of characteristic 0, dual to Quillen's cocommutative chain coalgebra model.

Because of his interest in manifolds, Sullivan found it natural to base his constructions on differential forms in the sense of Whitney; see Section 1.4. In particular he wanted to understand the relationship between Quillen's cocommutative chain coalgebra model and the commutative cochain algebra of forms. In the process, he had to overcome three fundamental difficulties. The first was that the algebra of forms is not of finite type, so that its dual is not a coalgebra. He remedied this problem by finding an algorithm for associating to a given commutative cochain algebra (A, d) such that $H^*(A, d)$ is of finite type a free commutative cochain algebra that is of finite type and that has the same cohomology algebra as (A, d). Furthermore, the free commutative cochain algebra thus constructed is *minimal*, which, in the case $H^0(A, d) = \mathbb{Q}$, $H^1(A, d) = 0$, means that the differential of any generator of the free model is a sum of terms of wordlength at least 2.

Sullivan's second problem was that differential forms are smooth, while his goal was to define models for all finite-type topological spaces, not only smooth manifolds. Finally, he needed to be able to work over $\mathbb{Q}$, even though de Rham theory is defined only over $\mathbb{R}$. He resolved these two problems simultaneously by basing his models on the rational piecewise-linear (Whitney) forms on a simplicial set. More precisely, given a space E, he considered the commutative cochain algebra $\mathcal{A}_{\mathrm{PL}}(E)$ whose elements of degree p are functions commuting with face and degeneracy operations that assign to each singular n-simplex on E a polynomial p-form with rational coefficients on the standard n-simplex. Whitney-type integration of forms then induces an isomorphism from $H^*\mathcal{A}_{\mathrm{PL}}(E)$ to $H^*(E; \mathbb{Q})$, the rational, singular cohomology of E.

Combining the two strategies mentioned above, Sullivan defined the minimal model of a space E to be the minimal commutative cochain algebra obtained via his algorithm from the algebra $A_{\mathrm{PL}}(E)$ of rational piecewise-linear forms on the singular simplices on E. He then showed that this association defined a one-to-one correspondence between rational homotopy types of simply-connected spaces and isomorphism classes of minimal commutative cochain algebras over $\mathbb{Q}$.

Sullivan started developing his commutative cochain algebra models in the late 1960's. He published a preliminary summary of his results in 1973 [101], mentioning the influence of the algebraic models of the 1950's on his work. He stressed the possible geometric applications of his models, such as to the study of nonabelian periods on smooth manifolds.

Sullivan wrote a second brief introduction to his models that appeared in 1976 [102]. The first part of the article is devoted to a succinct but informative account of the prehistory of rational homotopy theory, up through the work of Cartan and Koszul. In the second part Sullivan outlined further geometric applications of his models, including those that were later published in his joint articles with Deligne, Griffiths and Morgan and with Vigué; see Section 2.3.

Sullivan's definitive, detailed exposition of commutative cochain algebras in rational homotopy theory appeared in 1977 [103]. His motivating philosophy was that "any reasonable geometric construction on spaces can be mirrored by a finite algebraic one with minimal models". As confirmation of this philosophy, he outlined algebraic constructions of minimal models corresponding to free loop spaces, path fibrations, universal fibrations, and spaces of sections of a fibration.

In [103] Sullivan explained as well the important dual Postnikov nature of the minimal model of a space E. When E is simply connected and of finite type, the generating graded vector space of its minimal model is isomorphic to $\mathrm{Hom}_{\mathbb{Q}}(\pi_* E, \mathbb{Q})$. Furthermore, the differential of the minimal model very explicitly encodes the k-invariants of the Postnikov decomposition of E. The result for a nilpotent space is somewhat more difficult to express succinctly, but of the same nature.

2.3. *First applications of Sullivan models*

Sullivan collaborated with Pierre Deligne, Phillip Griffiths, and John Morgan on the first major geometric application of rational homotopy theory [31], showing that the real homotopy type of a compact Kähler manifold is a formal consequence of its real cohomology algebra. Their proof employed Sullivan's minimal models, as well as Hirsch theory for fibrations with free, transgressive fiber. In applying Sullivan's models to Kähler manifolds, they made explicit and then exploited the notion that "the manner in which a closed form which is zero in cohomology actually becomes exact contains geometric information".

Most rational homotopy theorists consider the article of Deligne, Griffiths, Morgan and Sullivan to be one of the most significant papers in the field. Not only is the geometric result presented of great interest, but the article also contains a very nice introduction to the theory of Sullivan's minimal models. In particular, the authors provided a clear and thorough explanation of the relationship between the minimal model and the rational Postnikov tower of a space.

The second major geometric application of Sullivan's models was to establishing the existence of infinitely many closed geodesics on a closed, compact, simply-connected Riemannian manifold whose rational cohomology has at least two generators [104]. Sullivan proved this result in 1974 together with Micheline Vigué, who was at that time a doctoral student in Paris. Their proof is based on a theorem of Gromoll and Meyer that states that if the rational Betti numbers of the free loop space on a manifold are unbounded, then there are infinitely many closed geodesics on the manifold [55]. To apply the theorem of Gromoll and Meyer, Sullivan and Vigué first constructed a Sullivan model for the free loop space E^{S^1} on any simply-connected space E of finite type. They then showed that if $H^*(E; \mathbb{Q})$ requires at least two algebra generators, then the rational Betti numbers of E^{S^1} are unbounded.

2.4. *Parallel developments*

While Sullivan was working on developing and refining his models of rational spaces, Kuo-Tsai Chen was persuing a different path towards a similar objective [30]. For a given simply-connected, finite-dimensional differentiable space E, Chen applied analytical methods to constructing a chain algebra quasi-isomorphic to the real chains on ΩE, the based loop space on E.

He considered derivations ∂ of degree -1 of the free algebra A on the desuspension of the $H_+(E; \mathbb{R})$; showing that those to which one can associate a *formal power series connection* satisfying a certain relation with ∂ are in fact differentials on A, i.e. (A, ∂)

is a chain algebra. Furthermore, there exists a quasi-isomorphism of cochain algebras $CU_*(\Omega E;\mathbb{R}) \to (A,\partial)$ defined via *transport* of the formal connection.

We can give a cursory description of Chen's construction, in more modern and less analytic language than that of his original description, as follows. First, we can interpret a formal power series connection ω either as an element of the completed tensor product of the differential forms on E with A or as an $\mathbb{R}$-linear map from the coalgebra C dual to A into the algebra of differential forms on E. In the second interpretation, we consider the dual $\hat{\partial}$ of ∂, which is a coderivation on C. The relation to be satisfied then becomes

$$\mathrm{d}\omega + \omega\hat{\partial} = \mu(\omega \otimes \omega)\Delta,$$

where μ is the multiplication of forms, Δ is the coproduct on C, and d is the differential on forms. In other words if ω is a *twisting cochain* with respect to $\hat{\partial}$, then iterated path integration with respect to the forms composing ω – the transport mentioned above – defines a quasi-isomorphism of cochain algebras $CU_*(\Omega E;\mathbb{R}) \to (A,\partial)$.

It is interesting to compare Chen's loop space model to that of Adams and Hilton, as well as to the cobar construction model of Adams; see Section 1.2. Like the model of Adams and Hilton, Chen's model has the advantage of having relatively few generators. Furthermore, given a couple (∂,ω) such that ω is a twisting cochain in the sense defined above, there is an explicit formula for the quasi-isomorphism $CU_*(\Omega E;\mathbb{R}) \to (A,\partial)$.

In the introduction to [30], Chen observed that his results and those of Sullivan were closely related, since the rational homotopy of a space determines its rational loop space homology and vice-versa. This is the essence of the theorem of Milnor and Moore stated in Section 1.3.

Victor Gugenheim was the first to describe Chen's formal power series connections in terms of twisting cochains, thus establishing a clear link with many other similar constructions, such as the bar construction. The theory of twisting cochains gradually developed into the field of homological perturbation theory. In [58], Gugenheim, together with Larry Lambe and Jim Stasheff, provided a detailed account of the role of Chen's formal power series connections in homological perturbation theory. The above description is based primarily on their account. Another interpretation, based on graded Lie algebras, appears in Daniel Tanré's book [106].

Also in the mid 1970's, Richard Swan developed another approach to obtaining a commutative cochain algebra over $\mathbb{Q}$ with cohomology isomorphic to that of a given space [105]. Inspired by Thom's construction of commutative cochain algebra models over $\mathbb{R}$ (see Section 1.2), Swan defined a functor $\mathcal{C}$ from the category of simplicial sets to the category of commutative cochain algebras such that

$$H^*\big(\mathcal{C}(X)\big) \cong H^*(X;\mathbb{Q}),$$

for all simplicial sets X. The elements of degree n in $\mathcal{C}(X)$ are morphisms of simplicial sets from X into a fixed simplicial $\mathbb{Q}$-module $L(n)$, consisting essentially of differential forms with polynomial coefficients on a canonical n-dimensional affine hyperplane.

Unlike Thom, who proved that his cochain algebra had the right cohomology via the Eilenberg–Steenrod axioms, Swan used the identification of cohomology classes with homotopy classes of maps into Eilenberg–Mac Lane spaces to show that $H^*(\mathcal{C}(X)) \cong H^*(X;\mathbb{Q})$, for all simplicial sets X.

3. The artisans

Once Quillen and Sullivan had laid the foundations of rational homotopy, they turned their attention to other areas of mathematics. The departure of the architects did not mean, however, that the structure of rational homotopy theory would never rise above its foundations. Even before Sullivan left the field in the middle of the 1970's, a number of enthusiastic topologists were hard at work, developing the tools of rational homotopy theory and proving the fundamental structural theorems that would eventually form the basis of the deep results of the 1980's and 1990's; see Section 4.

In this section we present the artisans who carried out the vital work of determining how and where to apply the theory of Quillen and Sullivan, and discuss some of their most important results. Long discussions with Steve Halperin, Yves Félix, Jean-Michel Lemaire, and Jean-Claude Thomas, together with helpful letters from Hans Baues, Pierre-Paul Grivel, Daniel Lehmann, Daniel Tanré and Micheline Vigué, supplied the framework for this section.

3.1. *Building on Sullivan's foundations*

A fortuitous meeting of the minds occurred at the Cornell Topology Fest in 1970, when Dennis Sullivan discussed commutative cochain algebra models for rational spaces with Steve Halperin and Domingo Toledo. At the time Halperin and Toledo were studying problems concerning characteristic classes, but were intrigued by Sullivan's ideas. Halperin, in particular, had already studied the Cartan–Koszul model for homogeneous spaces, and had learned about commutative cochain algebras and Koszul complexes from Greub. Sullivan's ideas now led him to wonder whether similar free cochain algebra models might exist for all spaces. Prompted by this query, Halperin and Sullivan worked out the correct definition of a minimal model together on the blackboard during the conference.

Two years later, knowledge of Sullivan's models began to propagate through the topological community, when Eric Friedlander, Phillip Griffiths and John Morgan taught a summer course on rational homotopy theory at the Istituto Matematico *Ulisse Dini* in Florence. Griffiths and Morgan later published their course notes, thus producing the first real textbook in rational homotopy theory [54]. In 1973 Sullivan himself contributed to the spread of rational homotopy theory, when he taught a course on minimal models at the AMS summer meeting on differential geometry at Stanford and gave a talk on his models at International Conference on Manifolds in Tokyo.

Sullivan's ideas began to spread quickly in 1974. That year Sullivan taught a semester course at the University of Paris at Orsay that covered the full development of his models. Prouté and Marie attended Sullivan's course and took extensive notes that later circulated widely under the name *Anonymous*. Furthermore, at a conference in differential geometry held in Besse-en-Chandesse, France, Vigué gave a talk on her joint work with Sullivan on the closed geodesic problem (see Section 2.3), while D. Lehmann presented a summary of Sullivan's work on minimal models.

In the autumn of 1974 the *Troisième cycle romand*, a long-running weekly mathematics seminar in French-speaking Switzerland, was devoted to an exposition of Sullivan's theory presented by F. Da Silveira, P.-P. Grivel and A. Haefliger. Da Silveira and Grivel were both doctoral students of Haefliger, working of theses in rational homotopy theory. Haefliger himself was in the process on applying Thom's suggested method for construction of the model of a functional space (see Section 1.2) to verifying that the model Sullivan suggested in [103] was indeed correct [61].

It was also in 1974 that Halperin once again became involved in rational homotopy theory. At the invitation of Koszul, Halperin spent the year in Grenoble, on leave from the University of Toronto. During his stay he made a trip with J. Weil to Geneva, where he met Haefliger, with whom he discussed foliations and Gel'fand–Fuks cohomology. At lunch they were joined by Frank Peterson, who mentioned an interesting question that Sullivan had asked in a preprint of *Infinitesimal computations in topology* [103]: if $(\Lambda V, d)$ is the Sullivan minimal model of a simply-connected space E with finite-dimensional homotopy and cohomology, is it true that the following conditions are equivalent?

(1) $\sum_i (-1)^i V^i = 0$.
(2) $\sum_i (-1)^i \dim H^*(E; \mathbb{Q}) > 0$.
(3) $H^*(E; \mathbb{Q}) = H^{\text{even}}(E; \mathbb{Q})$.

Sullivan's question was actually a translation in terms of minimal models of a conjecture of C. Allday [3], in which

$$\sum_i (-1)^i \dim \pi_*(E) \otimes \mathbb{Q} = 0$$

replaced the condition (1) above. The interest of Allday's conjecture resides in the fact that it implies that if a compact Lie group G of rank r acts with only finite isotropy on a simply-connected finite CW-complex X with finite-dimensional homotopy, then

$$\sum_i (-1)^i \dim \pi_*(X) \otimes \mathbb{Q} \leqslant -r,$$

a conjecture due to W.-Y. Hsiang.

Intrigued by Sullivan's problem, Halperin set to work almost immediately on solving it. One year later, *Finiteness in the minimal models of Sullivan* [63], the first major paper on the structure and properties of Sullivan models after the work of Sullivan himself, was ready for submission. In his first article in rational homotopy theory, Halperin not only answered Sullivan's question affirmatively, but also showed that the rational cohomology algebra of any simply-connected space E with finite-dimensional homotopy and cohomology satisfies Poincaré duality. Furthermore, he provided a formula for the degree of the top class in cohomology in terms of the degrees of the homogeneous basis elements of $\pi_*(E) \otimes \mathbb{Q}$.

Halperin pointed out in the introduction to his article that his work was a generalization of the results published by Cartan and Koszul in the proceedings of the 1950 colloquium in Bruxelles [28, 29, 79]; see Section 1.2. In particular, Cartan had solved a special case of Sullivan's problem, to which Halperin had been able to reduce the general case.

When Allday found out about Halperin's finiteness result in 1976, he applied it to showing that if a Lie group T acts locally smoothly on a topological manifold X and if F is a component of the nonempty fixed-point set, then

$$\sum_i (-1)^i \dim \pi_*(X) \otimes \mathbb{Q} = \sum_i (-1)^i \dim \pi_*(F) \otimes \mathbb{Q} \quad [4].$$

The following year Allday and Halperin collaborated in proving using Sullivan models that if G is a compact, connected Lie group of rank r that acts almost freely on a finite CW-complex X with finite-dimensional rational homotopy, then

$$\sum_i (-1)^i \dim \pi_*(X) \otimes \mathbb{Q} \leqslant -r \quad [5].$$

Daniel Lehmann of the University of Lille, an expert in characteristic classes, was also in Grenoble for a long visit in 1974. Upon learning more about Sullivan's models, he was intrigued by the fact that, from a topological point of view, the algebra of differential forms on a complex contains much more information than its cohomology algebra. Thus it is worthwhile to work with the forms themselves as long as possible, without rushing to calculate cohomology classes. The work of Halperin, such as his contribution to the Greub–Halperin–VanStone collaboration, also impressed him, leading him to invite Halperin to spend the academic year 1975–1976 in Lille.

In the autumn of 1975 Halperin and Lehmann were reunited in Lille. Halperin was unable to spend the year in Lille as planned, however, as he soon became seriously ill and was obliged to return to Canada. His doctors succeeded rapidly in vanquishing his illness, so that he was back in Lille at the beginning of April 1976.

During Halperin's absence, Lehmann organized the first seminar in rational homotopy theory at the University of Lille, in order to study Sullivan's *Anonymous*. He also started writing a book on Sullivan's models in rational homotopy theory, later published in the Astérisque series [80].

The study of *Anonymous* led one of Lehmann's students, Jean-Claude Thomas, to speculate on the Sullivan model of a fibration, which he decided to study for his *Thèse d'État*. Unfortunately for and unknown to Thomas, Haefliger's student Grivel was just completing his thesis on the same subject. In his thesis Grivel proved that for any fibration of simply-connected spaces, "the cofiber of the Sullivan model is a Sullivan model of the fiber" [53]. Somewhat more precisely, given a model of the total space that is a free extension of a model of the base space, the quotient of the total space model by the base space model is a model of the fiber. This theorem, later generalized by Halperin and Thomas to nilpotent spaces and to fiber squares [65], has proved to be of immense importance in rational homotopy theory, both for constructing models of specific interesting spaces and for proving general results comparing the values of certain rational homotopy invariants of the fiber, total space and base of a fibration.

Grivel proved his theorem using a spectral sequence he constructed for the differential forms on a simplicial fibration that was analogous to the Serre spectral sequence. Michel Zisman contributed to the construction of Grivel's spectral sequence when he explained to Grivel about bisimplicial sets, a notion that turned out to be crucial to Grivel's proof, during the *Troisième Cycle Romand* in 1974.

Before returning to Lille in the spring of 1976, Halperin paid a visit to Jim Stasheff at the University of North Carolina. They decided to work on the problem of determining when the minimal Sullivan model of a space E is *formal*, i.e. when it is also a minimal model of the cohomology algebra $H^*(E; \mathbb{Q})$ or, equivalently when the rational cohomology of E determines its rational homotopy; see Section 2.3. Their project quickly expanded to determining when an isomorphism of rational cohomology algebras of spaces can be realized by a rational homotopy equivalence. Using Sullivan models, they translated this problem into determining when an isomorphism of rational cohomology algebras of commutative cochain algebras

$$f : H^*(A, d) \xrightarrow{\cong} H^*(B, d)$$

can be realized by a homotopy equivalence, i.e. by a sequence of quasi-isomorphisms of commutative cochain algebras

$$(A, d) \xleftarrow{\simeq} (A_1, d) \xrightarrow{\simeq} \cdots \xleftarrow{\simeq} (A_n, d) \xrightarrow{\simeq} (B, d).$$

Their solution of the realizability problem consists in the construction of a sequence of computable obstructions $\{O_n(f)\}$, one for each degree. They showed that the isomorphism f is realizable if and only $O_n(f) = 0$ for all n [70]. This theorem is remarkable and esthetically pleasing, but the methods Halperin and Stasheff employed in the construction of the obstructions have proved at least as important in the development of rational homotopy theory as the obstruction theorem itself.

The first step in the construction of the obstructions consists in the definition of a bigraded minimal model of the commutative cochain algebra $(H, 0)$, where $H = H(A, d) \cong H(B, d)$, based on Tate's models of residue class rings of commutative Noetherian rings; see Section 1.2. Halperin and Stasheff then explained how to perturb the differential of the bigraded model, to obtain a canonical filtered model of (A, d) or of (B, d). Finally, they defined the sequence of obstructions in terms of the canonical filtered models.

The filtered models of Halperin and Stasheff represented a great breakthrough in rational homotopy theory. It suddenly became much easier to carry out explicit calculations, opening the field to many more exciting possible applications.

When Halperin was once again in Lille, he taught a rigorous and thorough course on Sullivan models, the ultimate goal of which was to prove Grivel's theorem for nilpotent spaces. His course notes were published first by the University of Lille in 1977, then as a memoir of the French Mathematical Society in 1983 [65] and have served ever since as the ultimate reference for basic structural theorems concerning Sullivan models.

While in France, Halperin visited the Institut des Hautes Etudes Scientifiques (I.H.E.S.) in Paris, to discuss rational homotopy theory with Sullivan. He found, however, that Sullivan was no longer very interested in the subject. On the other hand, while still at the I.H.E.S. he made the acquaintance of two other mathematicians who were interested in rational homotopy theory: Micheline Vigué, who had already worked with Sullivan on the closed geodesic problem (see Section 2.2), and Karsten Grove, who wanted to apply Sullivan models to problems in differential geometry. Both commenced working with Halperin shortly thereafter. Vigué ended up writing her *Thèse d'État* concerning relative bigraded and filtered models and realization of morphisms in cohomology under Halperin's supervision, while Grove collaborated with Halperin on several articles in differential geometry, such as [56, 57].

During Halperin's stay at Lille in 1976, Lehmann organized a conference at Luminy, entitled *Journées Lille-Valenciennes sur les formes différentielles*, the first conference in rational homotopy theory. Among the speakers were Grivel, who talked about his thesis work; Bohumil Cenkl, who talked about Chen's iterated integrals; Jean-Paul Brasselet, who talked about the obstruction theory of Halperin and Stasheff; Thomas, who talked about finiteness in minimal models; and Jean-Michel Lemaire, who talked about Quillen's work; see Section 3.2.

As the conference title indicates, in 1976 Lehmann, and certainly others as well, still considered rational homotopy theory primarily as a useful tool in differential geometry. The conference in Luminy marks the shift in focus of rational homotopy theory from differential geometry to algebraic topology.

In 1977, though Halperin was home in Toronto, the effervescence inspired by rational homotopy theory continued in Lille. Thomas, Vigué, B. Callenaere and A. Altaï, joined in 1978 by D. Tanré, as well as by Y. Félix of the University of Louvain-la-Neuve, organized a regular rational homotopy theory seminar, which ran until 1983. The seminar was termed an *Ecole de calcul* – school for computations – as the participants were intent on learning the properties of the Sullivan models, a process that entailed the calculation of numerous examples. Félix has likened the Lille seminar to playing scales.

Halperin continued his series of important contributions to rational homotopy theory in 1977, when he pursued his study of Sullivan models of fibrations [64]. He showed that there was a close relation between properties of the connecting homomorphism in the long exact sequence of a fibration and properties of the cohomology of the fiber and the base. In particular, he proved that in a fibration of simply-connected CW-complexes of finite type, if the total dimension of the rational cohomology of the fiber is finite, then the connecting map

$$\partial : \pi_{2n+1}(B) \otimes \mathbb{Q} \to \pi_{2n}(F) \otimes \mathbb{Q}$$

is the zero map for all n. As a consequence of these results, he obtained conditions implying the nonexistence of fibrings of certain homogeneous spaces. This paper served as an interesting warm-up exercise for his monumental article with Félix on Lusternik–Schnirelmann category three years later; see below. It also served later to unite rational homotopy theory and local algebra; see Section 4.1.

In collaboration with John Friedlander in 1978, Halperin applied number theory to establishing a definitive and complete description of spaces with finite-dimensional rational homotopy and cohomology [51]. Together they gave an arithmetic characterization of the sequences of odd degrees and even degrees in which such a space can have nonzero homotopy groups. As consequences of their characterization, they obtained, for example, that the total dimension of the rational homotopy and the degree of the top even degree element in rational homotopy were bounded above by the degree of the top class in rational cohomology and that the degree of the top odd degree element in rational homotopy was bounded above by twice the degree of the top class, minus one. Such precise, easily verified conditions enabled rational homotopy theorists to determine at a glance whether a given minimal algebra was the model of a space with finite-dimensional rational homotopy and cohomology, an enormous computational bonus.

Geneva continued to be a center of activity in rational homotopy theory during the late 1970's. In 1977 Grivel defended his thesis, followed in 1979 by F. da Silveira, another student of Haefliger, who studied relative minimal models. During the same period K. Shibata visited Geneva, where he worked on applying Haefliger's commutative cochain algebra model for the space of sections of a fibration to constructing an infinite number of indecomposable elements in the Gel'fand–Fuks cohomology of a sphere [99].

During the early and mid 1970's there were several other small groups working on understanding and extending Sullivan's theory. In the spring of 1974, a seminar was organized at the University of Illinois–Chicago Circle based on the notes from the 1972 summer school of Friedlander, Griffiths and Morgan. During the course of the seminar A.K. Bousfield and V.K.A.M. Gugenheim realized that, as it stood, there were a few important gaps in Sullivan's theory. In particular, they were unable to prove to their satisfaction that there was a one-to-one correspondence between rational homotopy types of simply-connected CW-complexes and isomorphism classes of minimal, simply-connected commutative cochain algebras. The problem lay in the Sullivan's geometric realization of a minimal algebra, which was based on a straightforward, but non-natural, Postnikov construction.

Bousfield and Gugenheim decided to apply methods from Quillen's *Homotopical Algebra*, as well as the notions of Bousfield–Kan localization, to obtaining a solid proof of Sullivan's equivalence of rational homotopy categories. They first defined a pair of adjoint functors between the categories of commutative cochain algebras and of simplicial sets, which then gave rise to adjoint functors between the respective homotopy categories. They then showed that when restricted to the full subcategories of commutative cochain algebras equivalent to minimal algebras with a finite number of generators in each degree and of nilpotent, rational simplicial sets of finite type, the latter pair of adjoint functors were in fact equivalences. They published their results in 1976 as a Memoir of the AMS [25].

Another team actively collaborating in rational homotopy theory throughout the 1970's was formed of R. Body, who defended his thesis at the University of British Columbia in 1972, and R. Douglas. Body's thesis consisted of study of the number of rational homotopy types with a given integral cohomology ring, based more on standard homotopy-theoretic methods than on the techniques of either Quillen or Sullivan. He published two articles on this subject, one with Douglas [21], on which C.R. Curjel also collaborated, and one alone [20]. They showed that if A is a graded ring such that $A \otimes \mathbb{Q}$ is either a finite tensor product of truncated polynomial algebras on one generator or the quotient of a free commutative algebra by a regular sequence, then there are only a finite number of rational homotopy types of finite simply-connected polyhedra E such that $H^*(E; \mathbb{Z}) \cong A$.

In 1976 Body and Douglas studied the problem of unique factorization of rational homotopy types. It was known that integral homotopy types do not factor uniquely, i.e. there exist spaces X, Y, Y' such that $X \times Y$ and $X \times Y'$ are homotopy equivalent, but Y and Y' are not. Using Sullivan's models, Body and Douglas showed that every rational homotopy type of formal, simply-connected spaces with finite-dimensional rational homotopy has a unique factorization [22]. Both Body and Douglas published several more articles on factorization problems during the years that followed.

The 1977 paper by P. Andrews and M. Arkowitz on higher order Whitehead products and the Sullivan model also had a definite influence on the later developments of rational homotopy theory [8]. Their idea was to generalize Sullivan's results that the quadratic part of the differential in the Sullivan minimal model of a space E contains the same information as the rational Whitehead products in $\pi_*(E) \otimes \mathbb{Q}$. They showed that the rth order Whitehead products in $\pi_*(E) \otimes \mathbb{Q}$ correspond to the summand of the differential that increases wordlength by exactly $r - 1$.

3.2. *Building on Quillen's foundations*

Quillen's theory was also the object of intense study during the 1970's, though by fewer researchers than Sullivan's more computational approach to rational homotopy theory. John Moore, who was an ardent proponent of Quillen's foundational work from the time it appeared in preprint form in the late 1960's, is probably the instigator of the subsequent developments and generalizations of the chain Lie algebra version of rational homotopy theory. Moore had always believed that dualizing, i.e. passing from chains to cochains, was unnecessary, if one were willing to work with coalgebras instead of algebras, as well as risky and restrictive, since one had to be careful of finite-type conditions. He thus preferred Quillen's models to Sullivan's, as Quillen never dualized the objects he studied.

In 1976, around the time Quillen's article [91] began circulating as a preprint, a young French student form the Ecole Normale in Paris, Jean-Michel Lemaire, took advantage of the exchange program between the Ecole Normale and Princeton to spend a year at Princeton, where he met and studied with Moore. Moore gave a copy of Quillen's article to Lemaire and suggested that he work on determining whether or not the Poincaré series of the homology of a loop space is always rational, a question due to Serre. Moore also discussed the notion of minimal models with Lemaire during his stay at Princeton, though in terms of coalgebras rather than algebras.

Out of the year he spent in Princeton grew Lemaire's impressive thesis, that he defended in 1973 and published in the Springer Lecture Note Series in 1974 [81]. Lemaire's thesis was crucial not only for disseminating and explaining an approach to rational modelling of topological spaces that did not rely on dualization, but also for rehabilitating the Adams–Hilton model, which plays a key role therein; see Section 1.2.

In this thesis Lemaire extended results of Bott and Samelson showing that the homology algebra of the loop space on a suspension ΣE is a free algebra on the reduced homology of E to spaces obtained by attaching the cone on a suspension to another suspension. He found explicit examples of attachments of a wedge of spheres onto a wedge of spheres such that ensuing loop space homology has either an infinite number of generators or an infinite number of relations.

Lemaire's thesis also contains many results of a more purely theoretical nature. For example, Lemaire introduced in his thesis the notion of a minimal model of a chain algebra over a field. For any graded Lie algebra L of global dimension at most 2, he also outlined a method for constructing a CW-complex E such that the rational homotopy Lie algebra of E is isomorphic to L. More generally, he proved many important and interesting results on the structure of the categories of chain Lie algebras and Hopf algebras.

It is also in Lemaire's thesis that one finds the first mention of what was to become one of the most important numerical homotopy invariants in rational homotopy theory: the *Lusternik–Schnirelmann* (L.–S.) *category* of a space E, i.e. one less than the minimal cardinality of an open cover of E by sets contractible within E. Moore had indicated to Lemaire that L.–S. category was an invariant worth investigating, an affirmation that was at least partially confirmed by Lemaire's proof that for a given graded Lie algebra L of global dimension at most n, there exists a CW-complex whose rationalization is of L.–S. category at most n and whose homotopy Lie algebra is isomorphic to L.

In 1975, Lemaire, who was by then a professor at the University of Nice, read a preprint of an article on higher-order Whitehead products by Hans Baues of the University of Bonn [17] and was struck by Baues's use of the Adams–Hilton model. He then met Baues at

Oberwolfach, where Baues was speaking about minimal rational homotopy types, and invited him to Nice for discussion and possible collaboration. The paper that resulted from their joint work represented a major step towards making Quillen's theory as computationally accessible as Sullivan's minimal models [19].

In their article Baues and Lemaire constructed minimal models for chain algebras over any field and for chain Lie algebras over the rationals. The construction corresponds to the cellular or homology decomposition of a space, the Eckmann–Hilton dual of the Postnikov decomposition, which underlies Sullivan's minimal models. Baues and Lemaire were thus the first to explore in detail the duality underlying the relationship between the theories of Quillen and Sullivan. Providing substantial supporting evidence, they conjectured that the minimal model of the Quillen chain Lie algebra of a space E was isomorphic to the minimal model of the chain Lie algebra underlying the cobar construction on the dual of Sullivan's minimal model of E. Their conjecture was proved in 1996 by Martin Majewski of the Free University of Berlin [86].

Baues published the first application of his joint work with Lemaire in an article [18] submitted practically simultaneously with [19]. He showed that any simply-connected CW-complex is rationally homotopy equivalent to an unstable CW-complex, i.e. to a CW-complex whose suspension is a wedge of spheres.

Another important application of [19] is due to Lemaire, in collaboration with François Sigrist of the University of Neuchâtel. They showed that the set of simply-connected rational homotopy type with a given finite-dimensional cohomology algebra was in one-to-one correspondence with the quotient of a certain affine algebraic variety by the action of a certain algebraic group. This is exactly the type of result Sullivan was seeking when he developed his theory of commutative cochain algebra models. It is thus particularly interesting and satisfying to see such a result based on the Eckmann–Hilton dual methods of Quillen. Note as well that the techniques of Lemaire and Sigrist were complementary to those developed by Halperin and Stasheff in [70].

Around the time that Baues and Lemaire were initiating their collaboration, Joe Neisendorfer was also studying Quillen's theory [88]. He generalized Quillen's work to nilpotent spaces and, like Baues and Lemaire, defined minimal chain Lie algebras, in order to facilitate actual computations.

3.3. *The unified approach to rational homotopy theory*

Rational homotopy theory reached an important turning point in 1979, when the researchers studying Sullivan's theory and commutative cochain algebra models, represented by Halperin, Lehmann and their students, joined forces with those studying Quillen's theory and chain Lie algebra models, represented by Lemaire and Baues, during the workshop that Félix and Lehmann organized in Louvain-la–Neuve, Belgium and Lille at the end of May, 1979. This conference was also significant because of the permanent bond formed between the rational homotopy theorists of Lille and Louvain.

The workshop consisted of expository lectures to bring all participants up to date on the recent progress in rational homotopy theory, as well as of introductory lectures on subjects that the speakers felt could be of interest when studied from a rational point of view. For example, Lemaire talked about L.–S. category (see Section 3.2), to which the Lille contingent had already been exposed via the excellent survey article of Ioan James

[76]. Baues talked about the Gottlieb groups of a space E, i.e. the groups consisting of homotopy classes of maps $\alpha : S^n \to E$ such that $id_E \vee \alpha : E \vee S^n \to E$ extends over $E \times S^n$. Both of these notions have played a very important role in rational homotopy theory since 1980, as we describe below.

A significant proportion of the conference time was devoted to problem and discussion sessions. Félix, Halperin and Lehmann collected the most interesting problems raised in these sessions and published the collection of 17 problems in the preprint series of the University of Lille. Some of the problems formulated in 1979 are still unsolved today (1997).

Thanks to the 1979 workshop, rational homotopy theorists of both faiths in the Quillen–Sullivan schism learned to appreciate and understand the duality and complementarity of the two methods. In 1980 Félix and Thomas wrote a masterful synthesis, comparing and contrasting the techniques and applications of the two approaches to algebraic modelling of rational homotopy types [47].

Félix and Lehmann had organized the 1979 workshop to coincide with Félix's thesis defense. Though not Félix's official thesis director, Lehmann had participated actively in supervising his work. The subject of Félix's thesis was the classification of rational homotopy types with a given cohomology algebra, a subject related to the work of Halperin and Stasheff, as well as to that of Lemaire and Sigrist. Félix defined an algebraic variety representing the set of filtered models (see the discussion of the work of Halperin and Stasheff) with a given cohomology algebra. The automorphism group of the underlying bigraded model acts on the variety; its orbits are the different rational homotopy types of the filtered models. Félix obtained in this manner a result analogous to that of Lemaire and Sigrist, based on commutative cochain algebra models, rather than chain Lie algebra models [36].

A highly productive synergy resulted from the workshop of 1979. For example, in 1980 Félix, Lemaire and Sigrist collaborated in disproving a conjecture of Baues and Lemaire [45] concerning the relationship between the rational *conelength* of a space and the L.–S. category of its rationalization [19]. An *iterated cone of length* n is a space formed by n successive attachments of wedges of spheres. An iterated cone is *minimal* (or unstable) if the suspension of any attachment is homotopically trivial. The (strong) rational conelength of a space E is the smallest n such that E has the rational homotopy type of a (minimal) iterated cone of length n.

Since it is easy to see that attaching a cell to a space increases its L.–S. category by at most one, it is clear that $\mathrm{cat}_0(E) \leqslant f(E) \leqslant F(E)$, where $\mathrm{cat}_0(E)$ denotes the L.–S. category of the rationalization of E, $f(E)$ its rational conelength, and $F(E)$ its strong rational conelength. Based on their experience with Quillen models, Baues and Lemaire had conjectured that $\mathrm{cat}_0(E) = F(E)$. Félix, Lemaire and Sigrist found a family of counterexamples to this conjecture with $F(E) - \mathrm{cat}_0(E)$ arbitrarily large, generalizations of an example due to Félix with $\mathrm{cat}_0(E) = 3$ and $F(E) = 4$. They relied on computations with chain Lie algebra models to determine $F(E)$. In each of their examples, however, $\mathrm{cat}_0(E) = f(E)$, which led them to wonder whether this equality might always be satisfied. It was only in 1996 that Nicolas Dupont, a former student of Thomas, managed to find a counter-example to this conjecture [34].

Lemaire and Sigrist continued their study of L.–S. category and conelength using chain Lie algebra models in [82]. They first established that if $F(E) \leqslant 3$, then $f(E) = F(E)$. Based on a chain Lie algebra formulation of the notion of formality and of the Eckmann–

Hilton dual notion of coformality, they also showed that if E is formal or coformal, then $\mathrm{cat}_0(E) = F(E)$. They then studied in detail a particular nonformal, noncoformal space, demonstrating how well adapted chain Lie algebra models were to the calculation of various numerical homotopy invariants.

In 1979 Halperin began to invite his European colleagues to Toronto for prolonged visits. One of the first was Baues, whom Halperin invited for six weeks of lectures. Halperin had been surprised to find in Baues's paper [18] a short, simple proof of the fact that a simply-connected space with only odd-degree rational cohomology is rationally homotopy equivalent to a wedge of odd-degree spheres. He and Stasheff had proved the same result in their paper on obstruction theory [70], but their proof, based on commutative cochain algebra models was considerably more complex. The same year, Thomas also spent two months working in Toronto.

In June 1980 it was Félix's turn to visit Halperin. The three weeks that Félix and Halperin spent together in Toronto had profound consequences for the future of rational homotopy theory, as the article they wrote based on their work during that period contains several of the deepest theorems in rational homotopy theory, with innumerable important applications [37]. According to Halperin, Félix arrived in Toronto with "the right theorem but the wrong lemmas". Together they figured out what the right lemmas should be, working 15 to 18 hours a day until all the details were correct and the paper almost completely written.

The fundamental result upon which the other theorems in [37] rest is a characterization of the L.–S. category of the rationalization of a simply-connected space in terms of Sullivan models, based on the idea that the L.–S. category of a space should be related to a sort of "nilpotence" of its Sullivan model. Lemaire and Félix had discussed this possibility while waiting at the airport for Lemaire's flight back to Nice after the 1979 workshop. More precisely, Félix and Halperin showed that if $(\Lambda V, d)$ is the Sullivan minimal model of E and $(\Lambda(V \oplus W), D)$ is an extension of $(\Lambda V, d)$ that is quasi-isomorphic to the quotient cochain algebra $(\Lambda V/\Lambda^{>n} V, \bar{d})$, then $\mathrm{cat}_0(E) \leqslant n$ if and only if there is a retract of commutative cochain algebras $(\Lambda(V \oplus W), D) \to (\Lambda V, d)$.

The first important consequence that Félix and Halperin obtained from their algebraic characterization of L.–S. category is the *Mapping Theorem*: if E and E' are simply-connected spaces and there is a continuous map $E \to E'$ inducing an injection on rational homotopy groups, then $\mathrm{cat}_0(E) \leqslant \mathrm{cat}_0(E')$. The Mapping Theorem has proved extremely useful in a multitude of applications. Félix and Lemaire provided a short and elegant geometrical proof of the Mapping Theorem in 1984, extending its validity to the tame category as well [46].

Indeed Félix and Halperin applied the Mapping Theorem immediately to proving that if $\mathrm{cat}_0(E)$ is finite, then the rational Gottlieb groups $G_*(E) \otimes \mathbb{Q}$ are concentrated in odd depress and their total dimension is less than $\mathrm{cat}_0(E)$. Using the refined Mapping Theorem, together with the result concerning the Gottlieb groups, they then proved the beautiful *Dichotomy Theorem*: if E is a simply-connected space such that $\dim H^*(E; \mathbb{Q}) < \infty$, then either $\dim \pi_*(E) \otimes \mathbb{Q} < \infty$ or the sequence $\{\dim \pi_k(E) \otimes \mathbb{Q}\}_k$ grows exponentially. A sequence $\{a_k\}_k$ *grows exponentially* if there are constants $C_2 \geqslant C_1 > 1$ and an integer N such that

$$C_1^m \leqslant \sum_{k \leqslant n} |a_k| \leqslant C_2^n, \quad \forall n \geqslant N.$$

In the first case, the spaces are called *rationally elliptic* and were already very well understood by 1980, thanks to the articles of Halperin [63, 64], as well as the article of Friedlander and Halperin [51]. In the second case, the spaces are called *rationally hyperbolic* and very little was known about them at that time. Félix and Halperin quickly set in motion the process of learning more about hyperbolic spaces. They showed, for example, that $\{\dim H^k(\Omega E; \mathbb{Q})\}_k$ grows polynomially of order $\dim \pi_{\text{odd}}(E) \otimes \mathbb{Q}$ if E is rationally elliptic and exponentially if E is rationally hyperbolic. A sequence $\{a_k\}_k$ *grows polynomially of order at most* r if there is a constant A such that $|a_k| \leqslant Ak^r$ for all k. They also provided the first indication of the structure of the homotopy Lie algebra of a rationally hyperbolic space, when they proved that it is never nilpotent.

The year following the workshop organized by Félix and Lehmann was an exciting, dynamic period in the development of rational homotopy theory. It was also the last year of the "splendid isolation" of rational homotopy theory from the rest of mathematics, as we explain in Section 4.

4. The bridge builders and the consolidators

During the period of expansion of rational homotopy theory covered in Section 3, two algebraists, Luchezar Avramov of the University of Sofia and Jan-Erik Roos of the University of Stockholm, had discovered and begun to exploit a deep connection between local algebra and rational homotopy theory. In 1981 they established contact with the rational homotopy theorists, initiating a powerful synergy that led to a multitude of important results in both fields. Furthermore, rational homotopy theorists soon learned to apply their new techniques inspired by local algebra to solving problems in mod p homotopy theory.

We describe in this section the dramatic evolution of rational homotopy theory under the influence of local algebra and then in contact with mod p homotopy theory. The primary sources for this section include letters from Luchezar Avramov, Clas Löfwall, and Jan-Erik Roos, discussions with Yves Félix, Steve Halperin, and Jean-Claude Thomas, and the survey articles by Anick and Halperin [11], Avramov [14], and Avramov and Halperin [15], as well as Roos's *Mathematical Introduction* to the Proceedings of the 1983 Stockholm conference [93].

4.1. *The algebraic visionaries*

In 1974 Jan-Erik Roods found a copy of Lemaire's Springer Lecture Notes on loop space homology in a Stockholm bookstore, even before it had arrived at the math library of the University of Stockholm. Upon reading Lemaire's book, he was struck by the resemblance between Lemaire's work on Serre's question concerning Poincaré series of the rational homology of loop spaces and the work of local algebraists on the analogous question of Kaplansky and Serre for local rings. More precisely, Lemaire had studied the Poincaré series

$$\sum_{n\geqslant 0} \dim_{\mathbb{Q}} H_n(\Omega E; \mathbb{Q}) \cdot z^n$$

for E a finite, simply-connected CW-complex, while local algebraists were interested in the series

$$\sum_{n\geqslant 0} \dim_{\Bbbk} \operatorname{Ext}^n_R(\Bbbk, \Bbbk) \cdot z^n$$

for R a local, commutative, Noetherian ring with residue field $\Bbbk$. In both cases, the goal was to determine under what conditions the series represent rational functions.

Inspired by what he had read in Lemaire's book, Roos established a research program to study the homological properties of local rings, in particular those whose maximal ideal $\mathfrak{m}$ satisfied $\mathfrak{m}^3 = 0$, the first nontrivial case for Poincaré series calculations. He realized that in order to study local rings, it was useful, or even necessary, to work in the larger category of (co)chain algebras.

By 1976 Roos had proved that Serre's problem for complexes E such that $\dim E = 4$ and the Kaplansky–Serre problem for local rings $(R, \mathfrak{m})$ such that $\mathfrak{m}^3 = 0$ were equivalent [92]. He had thus demonstrated that not only were similar techniques useful in both rational homotopy theory and local algebra, but that there were also significant theorems common to both fields. It is interesting to note that the idea of linking the two fields first appeared in Serre's lecture notes on local algebra [97].

In 1976 Luchezar Avramov went to Stockholm to visit Roos. They discussed extensively the increasingly evident connections between the homology of local rings and the homology of loop spaces on finite CW-complexes. Avramov also lectured on Eilenberg–Moore-type spectral sequences and exact sequences of homotopy Lie algebras of local rings, which are defined as follows. Levin had shown in 1965 that if R is a local, commutative ring with residue field $\Bbbk$, then $\operatorname{Tor}^R(\Bbbk, \Bbbk)$ is a graded, divided powers Hopf algebra [84]. Since the dual of a graded, divided powers Hopf algebra is the universal enveloping algebra of a uniquely defined graded Lie algebra ($\operatorname{char}\Bbbk = 0$ [87], $\operatorname{char}\Bbbk > 2$ [7], $\operatorname{char}\Bbbk = 2$ [100]), it is possible to associate to any Noetherian, local commutative ring R a uniquely defined graded Lie algebra $\pi_*(R)$, the homotopy Lie algebra of R.

Based on results in characteristic 0 due to Gulliksen in the late 1960's [60], Avramov had proved before his trip to Stockholm that if $R \to S$ is a homomorphism of Noetherian, local, commutative rings with the same residue field $\Bbbk$ such that S is R-flat, then there is an exact sequence of groups

$$\cdots \to \pi^n(S \otimes_R \Bbbk) \to \pi^n(S) \to \pi^n(R) \xrightarrow{\delta_n} \pi^{n+1}(S \otimes_R \Bbbk) \to \cdots$$

He showed furthermore that if $\sup\{j \mid \operatorname{Tor}^R_j(\Bbbk, S) \neq 0\} < \infty$, then $\delta_{2n+1} = 0$ for all n, the exact equivalent of Halperin's result for rational fibrations mentioned in Section 3.1.

Avramov's results appeared in 1977 [13], one year before the publication of Halperin's analogous results. The two authors were unaware of each other's contributions while they were under preparation. Influenced by Tate, who had indicated in [107] the importance for his work of his contact with John Moore (see Section 1.2), Avramov had acquired the habit of reading articles in topology and spotted Halperin's article soon after it appeared. The similarity of their results confirmed for Avramov his intuition that rational homotopy invariants provided the correct analogy for homology invariants of local rings in arbitrary characteristic.

4.2. *The bridge to local algebra*

In 1980 the thesis of a student from MIT, David Anick, sparked interest among both rational homotopy theorists and local algebraists. He had constructed a finite, simply-connected CW-complex E of dimension $\leqslant 4$ such that the Poincaré series of the homology of ΩE was not rational, thus answering Serre's question [9]. Anick's construction interested the rational homotopy theorists because of its relation to the dichotomy between elliptic and hyperbolic spaces; see Section 3.3. Local algebraists were interested because of Roos's result, that allowed the transcription of Anick's space into a local ring $(R, \mathfrak{m})$ with $\mathfrak{m}^3 = 0$ and with irrational Poincaré series.

Shortly after Anick's result became known, Roos and his student Clas Löfwall discovered other examples of local rings with irrational Poincaré series that they obtained by completely different methods [85].

The converging interests of the rational homotopy theorists and the local algebraists finally led to direct contact between the two groups in 1981. Roos had begun sending preprints from his group in Stockholm to the rational homotopy theorists, inspiring Félix and Thomas to begin work on calculating the radius of convergence of the Poincaré series of a loop space [48]. They eventually succeeded in proving the following beautiful characterization, the proof of which relies heavily on results from [37]: a simply-connected space E is rationally elliptic if and only if the radius of convergence of the Poincaré series of ΩE is 1. If E is rationally hyperbolic, then the radius of convergence is strictly less than 1. Moreover, they found a relatively easily computable upper bound for the radius of convergence if E is a hyperbolic, formal space.

During the first half of 1981, a preprint containing the above results reached Avramov in Sofia, who arranged to visit Lille as soon as possible thereafter. His visit marked the beginning of his extremely fruitful collaboration with the rational homotopy theorists. Perhaps as a result of that visit, Félix and Thomas were able to conclude their article on the radius of convergence with a proof of a rational version of a conjecture of Golod and Gulliksen on the radius of convergence of Poincaré series of rings. More precisely, they showed that if A is a Noetherian, connected graded commutative algebra over a field $\Bbbk$ of characteristic zero and ρ_A denotes the radius of convergence of

$$P_A(z) = \sum_{n \geqslant 0} \dim \operatorname{Tor}_n^A(\Bbbk, \Bbbk) \cdot z^n,$$

then either $\rho_A = +\infty$ and A is a polynomial algebra; or $\rho_A = 1$, A is a complete intersection, and the coefficients of $P_A(z)$ grow polynomially; or $\rho_A < 1$, A is not a complete intersection, and the coefficients of $P_A(z)$ grow exponentially. Recall that a local, Noetherian ring is a *complete intersection* if its completion, in terms of powers of its maximal ideal, is isomorphic to the quotient of a regular, local ring by an ideal generated by a regular sequence. Avramov later generalized this result of Félix and Halperin to any characteristic.

In the autumn of 1981 Halperin was back in Europe, visiting Baues in Bonn. During Halperin's visit, Baues organized a conference that Roos attended. When Halperin talked about his work on splitting rational fibrations and his work with Félix on the Mapping Theorem and its consequences, Roos was impressed by the analogy with the split exact sequence of Avramov and intrigued by the Dichotomy Theorem, as similar results had been conjectured in local algebra, e.g., the conjecture of Golod and Gulliksen mentioned above.

During the Bonn conference, Halperin learned of Avramov's visit to Lille. Frustrated not to have been there, he decided to go with Félix and Thomas to see Avramov in Sofia. Together they rode the *Orient Express* to Sofia, where they spent a week in intense discussion with Avramov.

Since the contact with local algebraists seemed so promising, in June of the following year Lemaire and Thomas organized a conference in Luminy devoted to algebraic homotopy and local algebra. The morning sessions at the conference were devoted to synthesis talks, explaining algebraic problems and techniques to the topologists present and vice-versa. The written version of Avramov's talk on local algebra and rational homotopy theory provides an excellent and thorough introduction to the subject [14]. His article contains the first "dictionary" between rational homotopy theory and local algebra, explaining how to translate notions and techniques from one field to the other. Consequently, given a theorem in one field, applying the dictionary leads to a statement in the other field that stands a reasonable chance of being true, though the method of proof may be completely different.

In his article Avramov also emphasized the importance of minimal models in local ring theory. If K^R is the Koszul complex of a local, commutative ring R with residue field $\Bbbk$ such that the Yoneda algebra $\mathrm{Ext}^*_R(\Bbbk, \Bbbk)$ is Noetherian, then there is a minimal, commutative cochain algebra $(\Lambda V, d)$ over $\Bbbk$ quasi-isomorphic to K^R. Avramov called $(\Lambda V, d)$ the minimal model of R. He established its relevance by observing that in degrees greater than 1, the graded Lie algebra derived from $(\Lambda V, d_2)$ was isomorphic to the homotopy Lie algebra of R. Note that the underlying graded vector space of the graded Lie algebra derived from a free, commutative cochain algebra $(\Lambda V, d_2)$ is the desuspension of ΛV, while its bracket product is obtained by dualizing the differential d_2.

The proceedings of the Luminy conference also contain signs of links forming between rational homotopy theory and mod p homotopy theory. In one of the survey articles, for example, Dale Husemöller wrote about loop space decompositions and exponents in homotopy groups, using techniques based on graded Lie algebras mod p and the mod p Hurewicz homomorphism.

Shortly before the Luminy conference, Félix, Halperin, and Thomas completed an article [41] in which they continued the in-depth study of the homotopy Lie algebra of a rationally hyperbolic space begun by Félix and Halperin in [37]. They showed, for example, that if E is rationally hyperbolic, where the definition is slightly weakened to allow spaces of finite L.–S. category, instead of finite dimension, then its rational homotopy Lie algebra is not solvable. Moreover, they proposed as conjectures translations of their theorems into local algebra, where, for a local ring $(R, \mathfrak{m})$ with residue field $\Bbbk$, L.–S. category is replaced by $\dim_{\Bbbk}(\mathfrak{m}/\mathfrak{m}^2) - \operatorname{depth} R$ and infinite dimensional rational homotopy is replaced by R not being a complete intersection. Recall that

$$\operatorname{depth} R = \inf\{j \mid \mathrm{Ext}^j_R(\Bbbk, R) \neq 0\}.$$

In [41] Félix, Halperin and Thomas also mentioned a very important conjecture due to Avramov and Félix, stating that the homotopy Lie algebra of a rationally hyperbolic space should contain a free Lie algebra on at least two generators. This conjecture has motivated much interesting work in the study of the homotopy Lie algebra and has not as yet (1997) been proved.

Avramov and Halperin quickly proved a weaker version of one of the conjectures in [41], when they showed that R is a complete intersection if an only if its homotopy Lie algebra is nilpotent [15].

Also in 1982 Félix and Thomas proved the Lemaire–Sigrist conjecture on rational cone-length (see Section 3.2) for spaces of rational L.–S. category 2, i.e. they showed that $\mathrm{cat}_0(E) = 2$ if and only if $f(E) = 2$. They also provided detailed information on the structure of the homotopy Lie algebra of a space of rational L.–S. category 2, and proceeded to translate that information into results concerning the homotopy Lie algebra of a local ring $(R, \mathfrak{m})$ such that $\mathfrak{m}^3 = 0$.

An important event in the history of the interaction between rational homotopy theory and local algebra took place in 1983, when Roos organized a Summer School and Research Symposium at the University of Stockholm entitled *Algebra, Algebraic Topology and their Interactions*. As in Luminy the previous year, the morning sessions were devoted to survey lectures, while research-level talks occupied the afternoon sessions. Among the survey lecturers were Anick, Avramov, Gulliksen, Halperin, and Lemaire, as well as D. Eisenbud, M. Hochster, C. Lech, R. Sharp, and R. Stanley. Many of those who attended the conference remember it as a particularly exciting and dynamic ten days, during which algebraists and topologists learned a great deal from each other. Several fruitful collaborations formed during the conference as well.

Avramov and Halperin are the authors of the first article in the proceedings of the Stockholm conference, a thorough guide to translation between rational homotopy theory and local algebra [16]. It begins at a more elementary level than the survey article of Avramov in the proceedings of the Luminy conference, leading the reader from first principles of differential graded homological algebra to notions of homotopy fiber and loop space and on to the homotopy Lie algebra.

The proceedings of the Stockholm conference also include numerous results of the interaction between rational homotopy theory and local algebra. One of the most striking is the article by Halperin and Rikard Bøgvad, a young researcher who had been a student of Roos [23]. Using minimal model techniques, they proved two conjectures due to Roos, which are "translations" of each other. More precisely, they showed that

(1) if R is a local, commutative ring such that the Yoneda algebra $\mathrm{Ext}^*_R(\Bbbk, \Bbbk)$ is Noetherian, then R is a complete intersection;

(2) if E is a simply-connected, finite CW-complex such that the Pontryagin algebra $H_*(\Omega E; \mathbb{Q})$ is Noetherian, then E is rationally elliptic.

Their proof is based on a slightly weakened form of the Mapping Theorem that holds over a field of any characteristic, as well as on ideas from the article of Félix, Halperin and Thomas of the previous year [41].

The article by Félix and Thomas in the Stockholm proceedings constitutes another prime example of the synergy between rational homotopy theory and local algebra [50]. They studied the action induced in rational homology by the holonomy of a fibration $F \to E \to B$, i.e.

$$H_*(\Omega B; \mathbb{Q}) \otimes H_*(F; \mathbb{Q}) \to H_*(F; \mathbb{Q})$$

and its dual in cohomology

$$H_*(\Omega B; \mathbb{Q}) \otimes H^*(F; \mathbb{Q}) \to H_*(F; \mathbb{Q})$$

where F, E, and B have the homotopy type of CW-complexes of finite type and B is simply connected. Using Sullivan models, they obtained computational descriptions of these two actions, which they then applied to proving a number of interesting theorems, several of which are "translations" or generalizations of theorems in local algebra. For example, they translated a result of Gulliksen [59] into rational homotopy theory, proving that if the total dimensions of the rational homotopy of B and of the rational cohomology of E are finite, then $H_*(F;\mathbb{Q})$ is a Noetherian $H_*(\Omega B;\mathbb{Q})$-module.

In the spring of 1985 Halperin returned to Stockholm for a five-week visit. During his stay, he worked on applying minimal model techniques to answering on old question concerning the *deviations* of a local ring. The jth deviation, e_j, of a noetherian, local, commutative ring R with residue field $\Bbbk$ is $\dim_{\Bbbk} \pi^j(R)$. Assmus had shown in 1959 that R is a weak complete intersection if and only if $e_j = 0$ for all $j > 2$ [12], raising the question of whether any deviation could vanish if R were not a weak complete intersection. Halperin succeeded in answering this question, showing that if R is not a weak complete intersection, then $e_j \neq 0$ for all j. According to Halperin, he had been struggling to work out the details of the proof during his stay in Stockholm, until suddenly one evening at dinner with Löfwall, all the pieces fell into place.

The following winter it was Halperin's turn to host Swedish mathematicians in Toronto, when Bøgvad and Löfwall came to visit. While they were there, Halperin received a preprint from Félix and Thomas in which they proved a complicated, technical result relating L.–S. category of a space E and some other mysterious numerical invariant. After reading through the preprint, Bøgvad and Löfwall pointed out to Halperin that the technical computations of Félix and Thomas were actually calculations of a certain Tor over the enveloping algebra $U(\pi_*(\Omega E)\otimes\mathbb{Q})$ of the homotopy Lie algebra of E, which implied that their mysterious invariant was nothing but the depth of $U(\pi_*(\Omega E)\otimes\mathbb{Q})$. Suddenly the work of Félix and Thomas became easier to understand, and their calculations easier to carry out.

Félix, Halperin, Löfwall, and Thomas proceeded to work together on rewriting the paper from this new perspective and on trying understand the consequences of the results obtained. Carl Jacobsson, also of the University of Stockholm, pitched in as well to iron out the final details. By the late fall of 1986, the article now known in rational homotopy circles as the *Five Author* paper was ready for submission [38].

The Five Author paper is remarkable not only for the story of its genesis, but also for the importance of the results presented. It represents a great leap forward in understanding of the structure of the homotopy Lie algebra of a space or of a local ring. The principal innovation of the Five Author paper consists in exploiting the *radical* of the homotopy Lie algebra, i.e. the sum of all of its solvable ideals, which rational homotopy theorists had begun to study in 1983. The radical itself is in general not solvable.

Expressed in topological terms, the main theorem of the Five Author paper states that if E is a simply-connected CW-complex of finite type and $\mathrm{cat}(E) = m < \infty$, then the radical of the homotopy Lie algebra, $Rad(E)$ is finite dimensional and $\dim Rad(E)_{\mathrm{even}} \leqslant m$. This is a consequence of two further theorems, both of which are of great interest themselves. The first concerns the relations among the rational L.–S. category of a space and the depth and global dimension of its homotopy Lie algebra. Recall that the global dimension of a local ring R with residue field $\Bbbk$ is defined by

$$\mathrm{gl.dim}(R) = \sup\{j \mid \mathrm{Ext}^j_R(\Bbbk,\Bbbk) \neq 0\}.$$

The precise statement of this theorem in topological terms is then that if L is the homotopy Lie algebra of a simply-connected CW-complex of finite type E, then either

$$\operatorname{depth} UL < \operatorname{cat}_0(E) < \operatorname{gl.dim} UL$$

or

$$\operatorname{depth} UL = \operatorname{cat}_0(E) = \operatorname{gl.dim} UL.$$

The second theorem states that under the same hypotheses, if $\operatorname{depth} UL < \infty$, then $Rad(E)$ is finite dimensional and satisfies

$$\dim Rad(E)_{\text{even}} \leqslant \operatorname{depth} UL.$$

Moreover, if

$$\dim Rad(E)_{\text{even}} = \operatorname{depth} UL,$$

then $Rad(E) = L$.

Rational homotopy theorists have used the results cited above extensively since the middle of the 1980's. In particular, they have proved crucial to developing understanding of the homotopy Lie algebra of rationally hyperbolic spaces. The methods the five authors developed to prove their results have turned out to be extremely important as well. For example, since their goal was to relate $\operatorname{cat}_0(E)$ to $\operatorname{depth}(L)$, they needed to construct a model of the quotient cochain algebra $(\Lambda V/\Lambda^{>n} V, \bar{d})$, where $(\Lambda V, d)$ is the Sullivan minimal model of E. Their method for doing so, based on perturbation of a model for $(\Lambda V/\Lambda^{>n} V, \bar{d}_2)$ has been exploited repeatedly since then, as explained below.

4.3. *The bridge to* mod *p homotopy theory*

In 1986 Félix organized another conference in Louvain, attended this time not only by rational homotopy theorists and a few local algebraists, but also by nonrational homotopy theorists, such as Bill Dwyer, Haynes Miller, John Moore, Lionel Schwartz, and Paul Selick. Rational homotopy theory was opening itself to mod p homotopy theory, as it became increasingly clear that the algebraic methods of rational homotopy theory could be applied in a broader context with interesting consequences.

The article of Halperin and Lemaire in the proceedings of the Louvain conference is an important example of this new tendency among rational homotopy theorists [69]. They had noticed that in the articles mentioned above by Félix, Halperin and Thomas [41] and by Bøgvad and Halperin [23], as well as in the Five Author paper [38], the hypothesis that $\operatorname{cat}_0(E) \leqslant m$ for some finite m could be replaced by an apparently weaker hypothesis requiring only that there be a retraction of *differential* $(\Lambda V, d)$*-modules* $(\Lambda(V \oplus W), D) \to (\Lambda V, d)$, where $(\Lambda(V \oplus W), D)$ is a free extension of $(\Lambda V, d)$ that is quasi-isomorphic to the quotient cochain algebra $(\Lambda V/\Lambda^{>m} V, \bar{d})$ (see Section 3.3 and the discussion of [37]). The smallest such m is denoted $\operatorname{Mcat}_0(E)$. This observation, together with the desire to apply rational methods over fields of nonzero characteristic, inspired Halperin and Lemaire to define numerical invariants based on a new sort of algebraic model.

Halperin and Lemaire chose to model spaces over fields of positive characteristic by free, associative cochain algebras, since in positive characteristic it is in general impossible to find a commutative cochain algebra weakly equivalent to the cochains on a space. The method of adjoining new generators degree by degree to construct a cochain algebra weakly equivalent to a given cochain algebra generalizes easily to the noncommutative context. Thus, for any simply-connected space E of finite type, there exists a quasi-isomorphism of cochain algebras

$$(TV, d) \xrightarrow{\simeq} C^*(E; \Bbbk),$$

where TV is a free, associative (tensor) $\Bbbk$-algebra on a positively graded vector space of finite type and d increases wordlength by at least one. Note that such a model is far from unique, and that it has in general no geometric basis, unlike the Sullivan and Adams–Hilton models.

Given such a model (TV, d) for E, and a free extension $(T(V \oplus W), D)$ of (TV, d) that is quasi-isomorphic to the quotient cochain algebra $(TV/T^{>m}V, \bar{d})$, Halperin and Lemaire defined numerical invariants analogous to cat_0 and Mcat_0 as follows: $\mathrm{Acat}_{\Bbbk}(E) \leqslant m$ (respectively, $\mathrm{Mcat}_{\Bbbk}(E) \leqslant m$) if and only if there exists a retract of cochain algebras (respectively, differential (TV, d)-modules) $(T(V \oplus W), D) \to (TV, d)$. They proved that $\mathrm{Acat}_{\Bbbk}(E) \leqslant \mathrm{cat}(E)$ for all $\Bbbk$, while if $\Bbbk$ is of characteristic zero, then

$$\mathrm{Mcat}_{\Bbbk}(E) = \mathrm{Mcat}_0(E) \leqslant \mathrm{Acat}_{\Bbbk}(E) \leqslant \mathrm{cat}_0(E).$$

Once Halperin and Lemaire had shown that free, associative cochain algebra models could be useful for approximating L.–S. category over fields of positive characteristic, they set to work with Félix and Thomas on applying such models to generalizing the results in the Five Author paper. The article in which they presented their results [39], commonly known in rational homotopy circles as the *Four Author* paper, is the first substantial contribution of rational homotopy theory techniques, including those borrowed from local algebra, to mod p homotopy theory.

The main theorem in the Four Author paper is a generalization of one of the theorems in the Five Author paper. It states that if E is a simply-connected space such that $H_i(E; \mathbb{F}_p)$ is finite dimensional for all i, where $\mathbb{F}_p$ is the prime field of characteristic p, then depth $H_*(\Omega E; \mathbb{F}_p) \leqslant \mathrm{cat}\, E$. The key to the proof is a construction based on a free, associative cochain algebra model of E that is similar to the construction based on the Sullivan model used to prove the analogous theorem in the Five Author paper.

Under the hypotheses of the theorem above, $H_*(\Omega E; \mathbb{F}_p)$ is a connected, cocommutative graded Hopf algebras, so that the four authors were able to generalize the notion of radical to the mod p context as follows. Recall that a sub-Hopf algebra G of a connected, cocommutative graded Hopf algebra H is *normal* if $H_+ \cdot G = G \cdot H_+$ [87]. The *mod p radical* of E, $Rad_p(E)$, is the union of the normal, solvable sub-Hopf algebras of $H_*(\Omega E; \mathbb{F}_p)$. Just as in the characteristic zero case, the four authors showed that if depth $H_*(\Omega E; \mathbb{F}_p) = m < \infty$, then $Rad_p(E)$ is nilpotent. They also proved various other interesting results concerning the nature of $Rad_p(E)$, extending those of the Five Author paper.

Another noteworthy theorem from the Four Author paper generalizes a theorem of Serre that asserts that a noncontractible, finite, simply-connected CW-complex has infinitely many nontrivial homotopy groups [98]. Recall that E is an *n-stage Postnikov space* if there exists a tower of principal fibrations

$$E = E_n \xrightarrow{q_n} E_{n-1} \xrightarrow{q_{n-1}} \cdots \xrightarrow{q_1} E_0 = pt$$

such that the fiber F_i of each q_i is a product of Eilenberg–Mac Lane spaces,

$$F_i = \prod_{k \geqslant 1} K(G_{i,k}; k).$$

If in addition the localization of $G_{i,k}$ at p is a finitely generated module over the ring $\mathbb{Z}_{(p)}$ of integers localized at p, then E is of *finite type at p*. Félix, Halperin, Lemaire and Thomas showed that if E is a p-local CW-complex of finite conelength and of finite type at p, then E is contractible.

The Four Author paper marks the beginning of an highly productive period of study of the homology of loop spaces mod p, particularly by the team of Félix, Halperin, and Thomas, who referred to the wealth of results to be extracted as a "gold mine". One of their most significant contributions [43] consisted in the proof of a homological version of the well-known conjecture of Moore that if E is a finite, simply-connected CW-complex such that $\{\dim H_k(\Omega E; \mathbb{Q})\}_k$ grows at most polynomially, then there exists r such that p^r annihilates all the p-torsion in $\pi_*(E)$. Félix, Halperin, and Thomas showed that if E is a simply-connected space such that

(1) $\mathrm{cat}(E) < \infty$;
(2) each $H_i(E; \mathbb{Z}_{(p)})$ is finitely generated module over $\mathbb{Z}_{(p)}$; and
(3) $\{\dim H_k(\Omega E; \mathbb{F}_p)\}_k$ grows at most polynomially,

then there is an integer r such that p^r annihilates all the torsion in $H_*(\Omega E; \mathbb{Z}_{(p)})$. To prove this theorem, they studied spectral sequences of *elliptic* Hopf algebras, i.e. Hopf algebras that are finitely generated and nilpotent, establishing in the process a sequence of important and deep results concerning elliptic differential Hopf algebras. They had shown in an earlier article [42] that if a simply-connected space E satisfied conditions (1)–(3), then $H_*(\Omega E; \mathbb{F}_p)$ was an elliptic Hopf algebra.

Anick, too, made a landmark contribution to the study of loop space homology via rational homotopy techniques [10]. He observed that the Adams–Hilton model (TV, d) of a space E over any ring could be endowed with a coproduct $\psi : (TV, d) \to (TV, d) \otimes (TV, d)$ that was a map of chain algebras and a model for the diagonal map $E \to E \times E$. Furthermore, such a coproduct is necessarily coassociative and cocommutative up to chain homotopy. The triple (TV, d, ψ) is thus an example of what he called a *Hopf algebra up to homotopy.*

Let R be a subring of $\mathbb{Q}$ that contains $1/n$ for all integers $n < p$. Anick's main theorem states that every Hopf algebra up to homotopy (TV, d, ψ) over R such that $V = \bigoplus_{i=r}^{rp-1} V_i$ is isomorphic as an algebra to the enveloping algebra of a free chain Lie algebra. Moreover, ψ is chain homotopic to the natural coproduct on the enveloping algebra. Thus, in particular, the Adams–Hilton model over R of an r-connected CW-complex of dimension at most rp, called an *(r, p)-mild* complex, is the enveloping algebra of a differential graded Lie algebra.

As a corollary of this profound theorem, Anick proved a conjecture of Wilkerson asserting that if E is a finite simply-connected CW-complex, then pth powers vanish in

$\tilde{H}^*(\Omega E; \mathbb{F}_p)$ for $p \gg 0$. He even obtained a precise lower bound on p, showing that if E is an (r, p)-mild complex, then pth powers vanish in $\tilde{H}^*(\Omega E; \mathbb{F}_p)$.

Anick's theorem opened the door to even wider use of standard rational homotopy techniques in solving mod p homotopy problems. Let E be an (r, p)-mild complex, and let R be a subring of $\mathbb{Q}$ containing $\frac{1}{2}$, as well as $1/n$ for all $n < p$. In his important paper on universal enveloping algebras, Halperin observed that if one computed the (differential, graded) Chevalley–Eilenberg–Koszul cochain complex of the chain Lie algebra associated to E by Anick's theorem, then one obtained a free commutative cochain algebra weakly equivalent to the R-cochains of E [68]. He then defined an appropriate notion of minimality over R, and explained how to construct R-minimal commutative cochain algebras quasi-isomorphic to a given commutative cochain algebra over R. This generalization of Sullivan's rational models has proved quite useful in recent years, for extending rational homotopy theory results over subrings of $\mathbb{Q}$.

Halperin employed the observation above in establishing an extension of Milnor and Moore's theorem mentioned in Section 1.3 to principal ideal domains. More precisely, he showed that if R is a subring of $\mathbb{Q}$ containing $\frac{1}{2}$ and $1/n$ for all $n < p$ and E is an (r, p)-mild complex such that $H_*(\Omega E; R)$ is R-torsion free, then $H_*(\Omega E; R)$ is isomorphic to the universal enveloping algebra of its Lie algebra of primitives. He showed, furthermore, that if p is an odd prime, there is a functor $\mathbf{F}$ from the category of (r, p)-mild complexes to that of chain Lie algebras over $\mathbb{F}_p$ such that $H_*(\Omega E; \mathbb{F}_p)$ is naturally isomorphic to $U\mathbf{F}(E)$ for any (r, p)-mild complex E.

4.4. *Consolidating the structure of rational homotopy theory*

During the period described above of intense interaction with local algebra and, later, mod p homotopy theory, work on refining and deepening knowledge of rational homotopy theory continued. The articles of Halperin on torsion gaps in homotopy constitute a prime example of such effort [66, 67]. Félix had conjectured that if E was a finite, simply-connected CW-complex of dimension n such that $\pi_k(E) \otimes \mathbb{Q} \neq 0$ and $\pi_l(E) \otimes \mathbb{Q} \neq 0$ but $\pi_i(E) \otimes \mathbb{Q} = 0$ for all $k < i < l$, then $l - k < n$. In his first article on torsion gaps, written in 1986, Halperin proved the conjecture when $H_{\mathrm{odd}}(E; \mathbb{Q}) = 0$. More generally, he showed that there is an upper bound, expressed in terms of n and $\mathrm{cat}(E)$, for the degree of l, limiting both the location and length of such torsion gaps. The proofs of these results depend on classical, Sullivan model constructions and arguments.

In his second paper on torsion gaps, Halperin proved Félix's conjecture in complete generality. The proof is again purely based on Sullivan model constructions. In the course of the proof, Halperin obtained a very nice characterization of those minimal models that have finite dimensional cohomology.

I would like to end this section on a more personal note, as I think the experience I describe illustrates the unity of the rational homotopy community, as well as the importance of the interaction with local algebra. At the end of the paper of Félix and Halperin on rational L.–S. category there is a list of unsolved problems and open questions [37]. One of the questions on that list asks whether Ganea's conjecture, that $\mathrm{cat}(E \times S^n) = \mathrm{cat}(E) + 1$ for every finite complex E, is true rationally. In the mid 1980's a student of Halperin, Barry Jessup, decided to try to answer this question. By the summer of 1987, he had succeeded in proving that $\mathrm{Mcat}_0(E \times S^n) = \mathrm{Mcat}_0(E) + 1$ [77].

In the fall of 1987, I was a student of Anick at MIT, one of the few he had before leaving mathematics for medicine in 1991, and had also decided to work on the rational version of Ganea's conjecture. I went with Anick to Toronto to attend the Ontario Topology Seminar, where Halperin and Jessup were both present. Anick and I learned then of Jessup's result. The proof of the rational version of Ganea's conjecture was thus reduced to the problem of showing that $\mathrm{Mcat}_0 = \mathrm{cat}_0$, for which there was some evidence, as mentioned in Section 4.2. When asked for his opinion, Halperin expressed skepticism about whether $\mathrm{Mcat}_0 = \mathrm{cat}_0$, indicating exactly where he thought the problems would lie.

In January of 1988, after several weeks of fruitless calculations, I accompanied Anick on a month-long visit to the University of Stockholm, where he was a guest lecturer. During our stay, Löfwall patiently and carefully explained to me the Sullivan model constructions underlying the proofs in the Five Author paper. That was the crucial knowledge that enabled me to prove a few months later that $\mathrm{Mcat}_0 = \mathrm{cat}_0$ and thus to obtain, in conjunction with Jessup, a proof of the rational version of Ganea's conjecture [71].

Once he heard of my result, Halperin invited me to Toronto, where he proceeded to take me through every step of the proof at the blackboard during two very long days, until he was absolutely convinced that there was no mistake. It was a wonderful and reassuring, if draining, experience for a young researcher.

Epilogue

Since the end of the 1980's rational homotopy theory has evolved tremendously, expanding at a rapid rate. Relatively little research is currently devoted to problems in pure rational homotopy theory, though there remain a few intriguing outstanding conjectures, such as that of Avramov and Félix; see Section 4.2. In fact, much of the work of the former rational homotopy theorists is now oriented towards mod p homotopy theory, with the goal of either modelling spaces algebraically over $\mathbb{F}_p$ or translating algebraic methods directly to topology. Topologists with roots in rational homotopy theory are also solving problems in symplectic geometry, in Hochschild and cyclic homology, and in generalized Morse theory, among other fields. Closer to their origins, there are others working in tame homotopy theory or in abstract algebraic homotopy theory.

After an initial decade of prolific, internal development, followed by a decade of intense interaction with local algebra and mod p homotopy theory, rational homotopy theory has been gradually merging into mainstream homotopy theory over the past several years. Whether rational homotopy theory will still be an actively independent branch of homotopy theory ten years from now is difficult to predict. If it is not, however, its influence will certainly still be felt, both through the insight it provides into the nature of spaces "modulo torsion" and through the algebraic modelling methods developed by its practitioners.

Bibliography

[1] J.F. Adams, *On the cobar construction*, Proc. Nat. Acad. Sci. USA **42** (1956), 409–412.
[2] J.F. Adams and P.J. Hilton, *On the chain algebra of a loop space*, Comment. Math. Helv. **30** (1956), 305–330.

[3] C. Allday, *On the rank of space*, Trans. Amer. Math. Soc. **166** (1972), 173–185.
[4] C. Allday, *On the rational homotopy of fixed point sets of torus actions*, Topology **17** (1978), 95–100.
[5] C. Allday and S. Halperin, *Lie group actions on spaces of finite rank*, Quart. J. Math. Oxford **29** (1978), 63–76.
[6] M. André, *Cohomologie des algèbres différentielles où opère une algèbre de Lie*, Tôhoku Math. J. **14** (1962), 263–311.
[7] M. André, *Hopf algebras with divided powers*, J. Algebra **18** (1971), 19–50.
[8] P. Andrews and M. Arkowitz, *Sullivan's minimal models and higher order Whitehead products*, Canadian J. Math. **30** (1978), 961–982.
[9] D.J. Anick, *A counter-example to a conjecture of Serre*, Ann. of Math. **115** (1982), 1–33.
[10] D.J. Anick, *Hopf algebras up to homotopy*, J. Amer. Math. Soc. **2** (1989), 417–453.
[11] D.J. Anick and S. Halperin, *Commutative rings, algebraic topology, graded Lie algebras and the work of Jan-Erik Roos*, J. Pure Appl. Algebra **38** (1985), 103–109.
[12] E. Assmus, *On the homology of local rings*, Illinois J. Math. **3** (1959), 187–199.
[13] L. Avramov, *Homology of local flat extensions and complete intersection defects*, Math. Ann. **228** (1977), 27–37.
[14] L. Avramov, *Local algebra and rational homotopy*, Homotopie Algébrique et Algèbre Locale, Astérisque **113–114** (1984), 16–43.
[15] L. Avramov and S. Halperin, *On the structure of the homotopy Lie algebra of a local ring*, Homotopie Algébrique et Algèbre Locale, Astérisque **113–114** (1984), 153–155.
[16] L. Avramov and S. Halperin, *Through the looking glass: a dictionary between rational homotopy theory and local algebra*, Algebra, Algebraic Topology and Their Interactions, Lecture Notes in Mathematics vol. 1183, Springer (1986), 1–27.
[17] H.J. Baues, *Identitäten für Whitehead-Produkte höherer Ordnung*, Math. Z. **146** (1976), 239–265.
[18] H.J. Baues, *Rationale homotopietypen*, Manuscripta Math. **20** (1977), 119–131.
[19] H.J. Baues and J.-M. Lemaire, *Minimal models in homotopy theory*, Math. Ann. **225** (1977), 219–242.
[20] R. Body, *Regular rational homotopy types*, Comment. Math. Helv. **50** (1975), 89–92.
[21] R. Body and R. Douglas, *Homotopy types within a rational homotopy type*, Topology **13** (1974), 209–214.
[22] R. Body and R. Douglas, *Rational homotopy and unique factorization*, Pacific J. Math. **75** (1978), 331–338.
[23] R. Bøgvad and S. Halperin, *On a conjecture of Roos*, Algebra, Algebraic Topology and their Interactions, Lecture Notes in Mathematics vol. 1183, Springer (1986), 120–126.
[24] R. Bott, *A topological obstruction to integrability*, Proc. Symp. Pure Math. **16** (1970), 127–131.
[25] A.K. Bousfield and V.K.A.M. Gugenheim, *On PL de Rham Theory and Rational Homotopy Type*, Memoirs of Amer. Math. Soc. **179** (1976).
[26] E. Cartan, *Sur les nombres de Betti des espaces de groupes clos*, C. R. Acad. Sci. Paris **187** (1928), 196–197.
[27] E. Cartan, *Sur les invariants intégraux de certains espaces homogènes clos*, Ann. Soc. Math. Polon. **8** (1929), 181–225.
[28] H. Cartan, *Notions d'algèbre différentielle; application aux groupes de Lie et aux variétés où opère un groupe de Lie*, Colloque de topologie algébrique (espaces fibrés), Tenu à Bruxelles, Centre Belge de Recherches Mathématiques, Louvain (1950), 15–28.
[29] H. Cartan, *La transgression dans un groupe de Lie et dans un espace fibré principal*, Coloque de topologie algébrique (espaces fibrés), Tenu à Bruxelles, Centre Belge de Recherches Mathématiques, Louvain (1950), 57–72.
[30] K.-T. Chen, *Iterated path integrals*, Bull. Amer. Math. Soc. **83** (1977), 831–879.
[31] P. Deligne, P. Griffiths, J. Morgan and D. Sullivan, *Real homotopy theory of Kähler manifolds*, Invent. Math. **29** (1975), 245–274.
[32] G. de Rham, *Intégrales multiples et analysis situs*, C. R. Acad. Sci. Paris **188** (1929), 1651–1652.
[33] G. de Rham, *Sut l'analysis situs des variétés à n dimensions*, J. Math. Pures Appl. **10** (1931), 115–200.
[34] N. Dupont, *A counter-example to a conjecture of Lemaire and Sigrist*, Topology, to appear.
[35] S. Eilenberg and C. Chevalley, *Cohomology theory of Lie groups and Lie algebras*, Trans. Amer. Math. Soc. **63** (1948), 85–124.
[36] Y. Félix, *Classification homotopique des espaces rationnels de cohomologie donnée*, Bull. Soc. Math. Belgique **31** (1979), 75–86.

[37] Y. Félix and S. Halperin, *Rational L.-S. category and its applications*, Trans. Amer. Math. Soc. **273** (1982), 1–37.

[38] Y. Félix, S. Halperin, C. Jacobsson, C. Löfwall and J.-C. Thomas, *The radical of the homotopy Lie algebra*, Amer. J. Math. **110** (1988), 301–322.

[39] Y. Félix, S. Halperin, J.-M. Lemaire and J.-C. Thomas, *Mod p loop space homology*, Invent. Math. **95** (1989), 247–262.

[40] Y. Félix, S. Halperin and J.-C. Thomas, *L.-S. catégorie et suite spectrale de Milnor–Moore*, Bull. Soc. Math. France **111** (1983), 89–96.

[41] Y. Félix, S. Halperin and J.-C. Thomas, *The homotopy Lie algebra for finite complexes*, Publ. Math. I.H.E.S. **56** (1982), 387–410.

[42] Y. Félix, S. Halperin and J.-C. Thomas, *Hopf algebras of polynomial growth*, J. Algebra **125** (1989), 408–417.

[43] Y. Félix, S. Halperin and J.-C. Thomas, *Torsion in loop space homology*, J. Reine Angew. Math. **432** (1992), 77–92.

[44] Y. Félix, S. Halperin and J.-C. Thomas, Rational homotopy theory, Preprint, University of Toronto (1996).

[45] Y. Félix, J.-M. Lemaire and F. Sigrist, *Sur les cônes itérés et la catégorie de Lusternik–Schnirelmann rationnelle*, C. R. Acad. Sci. Paris **290** (1980), 905–907.

[46] Y. Félix and J.-M. Lemaire, *On the mapping theorem for Lusternik–Schnirelmann category*, Topology **24** (1985), 41–43.

[47] Y. Félix and J.-C. Thomas, *Homotopie rationnelle: dualité et complémentarité des modèles*, Bull. Soc. Math. Belgique **33** (1981), 7–19.

[48] Y. Félix and J.-C. Thomas, *The radius of convergence of Poincaré series of loop spaces*, Invent. Math. **68** (1982), 257–274.

[49] Y. Félix and J.-C. Thomas, *Sur la structure des espaces de L. S. catégorie 2*, Illinois J. Math. **30** (1986), 574–593.

[50] Y. Félix and J.-C. Thomas, *Sur l'opération d'holonomie rationnelle*, Algebra, Algebraic Topology and Their Interactions, Lecture Notes in Mathematics vol. 1183, Springer (1986), 136–169.

[51] J. Friedlander and S. Halperin, *An arithmetic characterization of the rational homotopy groups of certain spaces*, Invent. Math. **53** (1979), 117–133.

[52] W. Greub, S. Halperin and R. Vanstone, *Connections, Curvature and Cohomology III*, Academic Press, New York (1976).

[53] P.-P. Grivel, *Formes différentielles et suites spéctrales*, Ann. Inst. Fourier (Grenoble) **29** (1979), 17–37.

[54] P.A. Griffiths and J.W. Morgan, *Rational Homotopy Theory and Differential Forms*, Progress in Mathematics vol. 16, Birkhäuser, Basel (1981).

[55] D. Gromoll and W. Meyer, *Periodic geodesics on compact Riemannian manifolds*, J. Diff. Geom. **3** (1969), 493–510.

[56] K. Grove and S. Halperin, *Contributions of rational homotopy to global problems in geometry*, Publ. Math. I.H.E.S. **56** (1982), 171–177.

[57] K. Grove and S. Halperin, *Dupin hypersurfaces, group actions and the double mapping cylinder*, J. Differential Geom. **26** (1987), 429–459.

[58] V.K.A.M. Gugenheim, L. Lambe and J. Stasheff, *Algebraic aspects of Chen's twisting cochain*, Illinois J. Math. **34** (1990).

[59] T. Gulliksen, *On the Hilbert series of the homology of differential graded algebras*, Math. Scand. **46** (1980), 15–22.

[60] T.H. Gulliksen and G. Levin, *Homology of Local Rings*, Queen's Papers in Pure and Applied Mathematics vol. 20, Queen's University, Kingston, Ontario (1969).

[61] A. Haefliger, *Rational homotopy type of the space of sections of a nilpotent bundle*, Trans. Amer. Math. Soc. **273** (1982), 609–620.

[62] A. Haefliger, *Des espaces homogènes à la résolution de Koszul*, Ann. Inst. Fourier (Grenoble) **37** (1987), 5–13.

[63] S. Halperin, *Finiteness in minimal models*, Trans.Amer. Math. Soc. **230** (1977), 173–199.

[64] S. Halperin, *Rational fibrations, minimal models, and fibrings of homogeneous spaces*, Trans. Amer. Math. Soc. **244** (1978), 199–224.

[65] S. Halperin, *Lectures on Minimal Models*, Mémoire de la Société Mathématique de France (N.S.) vol. 9/10 (1983).

[66] S. Halperin, *Torsion gaps in the homotopy of finite complexes*, Topology **27** (1988), 367–375.

[67] S. Halperin, *Torsion gaps in the homotopy of finite complexes II*, Topology **30** (1991), 471–478.

[68] S. Halperin, *Universal enveloping algebras and loop space homology*, J. Pure Appl. Algebra **83** (1992), 237–282.

[69] S. Halperin and J.-M. Lemaire, *Notions of category in differential algebra*, Algebraic Topology: Rational Homotopy, Lecture Notes in Mathematics vol. 1183, Springer, Berlin, 138–153.

[70] S. Halperin and J. Stasheff, *Obstructions to homotopy equivalences*, Adv. in Math. **32** (1979), 233–279.

[71] K. Hess, *A proof of Ganea's conjecture for rational spaces*, Topology **30** (1991), 205–214.

[72] G. Hirsch, *Un isomorphisme attaché aux structures fibrées*, C. R. Acad. Sci. Paris **227** (1948), 1328–1330.

[73] G. Hirsch, *Homology invariants and fibre bundles*, Proceedings of the International Congress of Mathematicians, Vol. 2, Cambridge, MA, 1950, Amer. Math. Soc. (1952), 383–389.

[74] G. Hirsch, *Sur les groupes d'homologie des expaces fibrés*, Bull. Soc. Math. Belgique **6** (1954), 79–96.

[75] H. Hopf, *Über die Topologie der Gruppen-Manifaltigkeiten und ihrer Verallgemeinerungen*, Ann. of Math. **42** (1941), 22–52.

[76] I.M. James, *On category in the sense of Lusternik–Schnirelmann*, Topology **17** (1978), 331–348.

[77] B. Jessup, *Rational L.–S. category and a conjecture of Ganea*, Trans. Amer. Math. Soc. **317** (1990), 655–660.

[78] J.-L. Koszul, *Homologie et cohomologie des algèbres de Lie*, Bull. Soc. Math. France **78** (1950), 1–63.

[79] J.-L. Koszul, *Sur un type d'algèbres différentielles en rapport avec la transgression*, Colloque de topologie algébrique (espaces fibrés), Tenu à Bruxelles, Centre Belge de Recherches Mathématiques, Louvain (1950), 73–82.

[80] D. Lehmann, *Théorie homotopique des formes différentielles*, Astérisque **45** (1977).

[81] J.-M. Lemaire, *Algèbres Connexes et Homologie des Espaces de Lacets*, Lecture Notes in Mathematics vol. 422, Springer (1974).

[82] J.-M. Lemaire and F. Sigrist, *Dénombrement de types d'homotopie rationnelle*, C. R. Acad. Sci. Paris **287** (1978), 109–112.

[83] J.-M. Lemaire and F. Sigrist, *Sur les invariants d'homotopie rationnelle liés à la L. S. catégorie*, Comment. Math. Helv. **56** (1981), 103–122.

[84] G. Levin, *Homology of local rings*, Ph.D. Thesis, University of Chicago (1965).

[85] C. Löfwall and J.-E. Roos, *Cohomologie des algèbres de Lie graduées et séries de Poincaré–Betti non rationnelles*, C. R. Acad. Sci. Paris **290** (1980), 733–736.

[86] M. Majewski, *A cellular Lie algebra model for spaces and its equivalence with the models of Quillen and Sullivan*, Thesis, Free University of Berlin (1996).

[87] J. Milnor and J. Moore, *On the structure of Hopf algebras*, Ann. of Math. **81** (1965), 211–264.

[88] J. Neisendorfer, *Lie algebras, coalgebras, and rational homotopy theory for nilpotent spaces*, Pacific J. Math. **74** (1978), 429–460.

[89] H. Poincaré, *Analysis Situs*, Journal de l'Ecole Polytechnique **1** (1895), 1–121.

[90] D. Quillen, *Homotopical Algebra*, Lecture Notes in Mathematics vol. 43, Springer, Berlin (1967).

[91] D. Quillen, *Rational homotopy theory*, Ann. of Math. **90** (1969), 205–295.

[92] J.-E. Roos, *Relations between the Poincaré–Betti series of loop spaces and local rings*, Séminaire d'Algèbre Paul Dubreil, Lecture Notes in Mathematics vol. 740, Springer, Berlin, 285–322.

[93] J.-E. Roos, *A mathematical introduction*, Algebra, Algebraic Topology and their Interactions, Lecture Notes in Mathematics vol. 1183, Springer, Berlin (1986), iii–viii.

[94] J.-E. Roos, *Homology of free loop spaces, cyclic homology and nonrational Poincaré–Betti series in commutative algebra*, Algebra – Some Current Trends, Lecture Notes in Mathematics vol. 1352, Springer, Berlin (1988), 173–189.

[95] H. Samelson, *Beiträge zur Topologie der Gruppen-Mannigfaltigkeiten*, Ann. of Math. **42** (1941), 1091–1137.

[96] J.-P. Serre, *Groupes d'homotopie et classes de groupes abéliens*, Ann. of Math. **58** (1953), 258–294.

[97] J.-P. Serre, *Algèbre Locale. Multiplicités, 3rd ed.*, Lecture Notes in Mathematics vol. 11, Springer, Berlin (1975).

[98] J.-P. Serre, *Cohomologie modulo 2 des complexes d'Eilenberg–Mac Lane*, Comment. Math. Helv. **27** (1953), 198–232.

[99] K. Shibata, *On Haefliger's model for the Gelfand–Fuks cohomology*, Japan J. Math. **17** (1981), 379–415.

[100] G. Sjödin, *Hopf algebras and derivations*, J. Algebra **64** (1980), 218–229.

[101] D. Sullivan, *Differential forms and the topology of manifolds*, Proceedings of the International Conference on Manifolds and Related Topics in Topology, Tokyo, University of Tokyo Press (1973).

[102] D. Sullivan, *Cartan–de Rham homotopy theory*, Astérisque **32–33** (1976), 227–254.
[103] D. Sullivan, *Infinitesimal computations in topology*, Publ. I.H.E.S. **47** (1977), 269–331.
[104] D. Sullivan and M. Vigué-Poirrier, *The homology theory of the closed geodesic problem*, J. Differential Geometry **11** (1976), 633–644.
[105] R. Swan, *Thom's theory of differential forms on simplicial sets*, Topology **14** (1975), 271–273.
[106] D. Tanré, *Homotopie Rationelle: Modèles de Chen, Quillen, Sullivan*, Lecture Notes in Mathematics vol. 1025, Springer, Berlin (1983).
[107] J. Tate, *Homology of Noetherian rings and local rings*, Illinois J. Math. **1** (1957), 14–27.
[108] R. Thom, *Opérations en cohomologie réelle*, Séminaire H. Cartan 1954/1955, Exposé 17, ENS, Paris, reprinted by W.A. Benjamin, New York (1967).
[109] R. Thom, *L'homologie des espaces functionnels*, Colloque de topologie algébrique, Tenu à Louvain, Centre Belge de Recherches Mathématiques, Louvain (1956), 29–40.
[110] H. Whitney, *Geometric Integration Theory*, Princeton University Press (1957).

CHAPTER 28

History of Homological Algebra

Charles A. Weibel

Department of Mathematics, Rutgers University, New Brunswick, NJ 08903-2101, USA
E-mail: weibel@math.rutgers.edu

Homological algebra had its origins in the 19-th century, via the work of Riemann (1857) and Betti (1871) on "homology numbers", and the rigorous development of the notion of homology numbers by Poincaré in 1895. A 1925 observation of Emmy Noether [145] shifted the attention to the "homology groups" of a space, and algebraic techniques were developed for computational purposes in the 1930's. Yet homology remained a part of the realm of topology until about 1945.

During the period 1940–1955, these topologically-motivated techniques for computing homology were applied to define and explore the homology and cohomology of several algebraic systems: Tor and Ext for Abelian groups, homology and cohomology of groups and Lie algebras, and the cohomology of associative algebras. In addition, Leray introduced sheaves, sheaf cohomology and spectral sequences.

At this point Cartan and Eilenberg's book [41] crystallized and redirected the field completely. Their systematic use of derived functors, defined via projective and injective resolutions of modules, united all the previously disparate homology theories. It was a true revolution in mathematics, and as such it was also a new beginning. The search for a general setting for derived functors led to the notion of Abelian categories, and the search for nontrivial examples of projective modules led to the rise of algebraic K-theory. Homological algebra was here to stay.

Several new fields of study grew out of the Cartan–Eilenberg revolution. The importance of regular local rings in algebra grew out of results obtained by homological methods in the late 1950's. The study of injective resolutions led to Grothendieck's theory of sheaf cohomology, the discovery of Gorenstein rings and Local Duality in both ring theory and algebraic geometry. In turn, cohomological methods played a key role in Grothendieck's rewriting of the foundations of algebraic geometry, including the development of derived categories. Number theory was infused with new results from Galois cohomology, which in turn led to the development of étale cohomology and the eventual solution of the Weil Conjectures by Deligne.

Simplicial methods were introduced in the 1950's by Dold, Kan, Moore and Puppe. They led to the rise of homotopical algebra and non-Abelian derived functors in the 1960's. Among its many applications, perhaps André–Quillen homology for commutative rings

HISTORY OF TOPOLOGY
Edited by I.M. James

and higher algebraic K-theory are the most noteworthy. Simplicial methods also played a more recent role in the development of Hochschild homology, topological Hochschild homology and cyclic homology.

This completes a quick overview of the history we shall discuss in this article. Now let us turn to the beginnings of the subject.

1. Betti numbers, torsion coefficients and the rise of homology

Homological algebra in the 19-th century largely consisted of a gradual effort to define the "Betti numbers" of a (piecewise linear) manifold. Beginning with Riemann's notion of genus, we see the gradual development of numerical invariants by Riemann, Betti and Poincaré: the Betti numbers and Torsion coefficients of a topological space. Indeed, the subject did not really move beyond these numerical invariants until about 1930. And it was not concerned with anything except invariants of topological spaces until about 1945.

1.1. *Riemann and Betti*

The first step was taken by Riemann (1826–1866) in his great 1857 work "Theorie der Abel'schen Funktionen" [155, VI]. Let C be a system of one or more closed curves C_j on a surface S, and consider the contour integral $\int_C X\,\mathrm{d}x + Y\,\mathrm{d}y$ of an exact differential form. Riemann remarked that this integral vanished if C formed the complete boundary of a region in S (Stokes' theorem), and this led him to a discussion of "connectedness numbers". Riemann defined S to be $(n+1)$-*fold connected* if there exists a family C of n closed curves C_j on S such that no subset of C forms the complete boundary of a part of S, and C is maximal with this property. For example, S is "simply connected" (in the modern sense) if it is 1-fold connected. As we shall see, the connectness number of S is the homology invariant $1 + \dim H_1(S; \mathbb{Z}/2)$.

Riemann showed that the connectedness number of S was independent of the choice of maximal family C. The key to his assertion is the following result, which is often called "Riemann's lemma" [155, p. 85]: Suppose that A, B and C are three families of curves on S such that A and B form the complete boundary of one region of S, and A and C form the complete boundary of a second region of S. Then B and C together must also form the boundary of a third region, obtained as the symmetric difference of the other two regions (obtained by adding the regions together, and then subtracting any part where they overlap).

If we write $C \sim 0$ to indicate that C is a boundary of a region then Riemann's lemma says that if $A + B \sim 0$ and $A + C \sim 0$ then $B + C \sim 0$. This, in modern terms, is the definition of addition in mod 2 homology! Indeed, the C_j in a maximal system form a basis of the singular homology group $H_1(S; \mathbb{Z}/2)$.

Riemann was somewhat vague about what he meant by "closed curve" and "surface", but we must remember that this paper was written before Möbius discovered the "Möbius surface" (1858) or Peano studied pathological curves (1890). There is another ambiguity in this lemma, pointed out by Tonelli in 1875: every curve C_j must be used exactly once.

Riemann also considered the effect of making cuts (*Querschnitte*) in S. By making each cut q_j transverse to a curve C_j (see [155, p. 89]), he showed that the number of cuts needed to make S simply connected equals the connectivity number. For a compact Riemann surface, he shows [155, p. 97] that one needs an even number $2p$ of cuts. In modern language, p is the *genus* of S, and the interaction between the curves C_j and cuts q_j forms the germ of Poincaré Duality for $H_1(S; \mathbb{Z}/2)$.

Riemann had poor health, and frequently visited Italy for convalescence between 1858 and his death in 1866. He frequently visited Enrico Betti (1823–1892) in Pisa, and the two of them apparently discussed the idea of extending Riemann's construction to higher-dimensional manifolds. Two documents with very similar definitions survive.

One is an undated "Fragment on Analysis Situs" [155, XXVIII], discovered among Riemann's effects, in which Riemann defines the n-dimensional connectedness of a manifold M: replace "closed curve" with n-dimensional subcomplex (*Streck*) without boundary, and "bounding a region" with "bounding an $(n+1)$-dimensional subcomplex". Riemann also defined higher dimensional cuts (submanifolds whose boundary lies on the boundary of M) and observed that a cut of dimension $\dim(M) - n$ either drops the n-dimensional connectivity by one, or raises the $(n-1)$-dimensional connectivity by one. In fairness, we should point out that Riemann's notion of connectedness is not independent of the choice of basis, because his notion that A and B are similar (*veränderlich*) is not the same as A and B being homologous; a counterexample was discovered by Heegaard in 1898.

The other document is Betti's 1871 paper [25]. The ideas underlying this paper are the same as those in Riemann's fragment, and Betti states that his proof of the independence of the homology numbers from the choice of basis is based upon the proof in Riemann's 1857 paper. However, Heegaard observed in 1898 that Betti's proof of independence is not correct in several respects, starting from the fact that a meridian on a torus is not closed in Betti's sense.

Betti also made the following assertion [25, p. 148], which presages the Poincaré Duality theorem:

> In order to render a finite n-dimensional space simply connected, by removing simply connected sections, it is necessary and sufficient to make p_{n-1} linear cuts, ..., p_1 cuts of dimension $n-1$,

where $p_i + 1$ is the i-th connectivity number. Heegaard found mistakes in Betti's proof here too, and Poincaré observed in 1899 [148, p. 289] that the problem was in (Riemann and) Betti's definition of similarity: it is not enough to just consider the set underlying A, one must also account for multiplicities.

1.2. *Poincaré and Analysis Situs*

Inspired by Betti's paper, Poincaré (1854–1912) developed a more correct homology theory in his landmark 1895 paper "Analysis Situs" [148]. After defining the notion, he fixes a piecewise linear manifold (*variété*) V. Then he considers formal integer combinations of oriented n-dimensional submanifolds V_i of V, and introduces a relation called a *homology*, which can be added like ordinary equations: $\sum k_i V_i \sim 0$ if there is an $(n+1)$-dimensional submanifold W whose boundary consists of k_1 submanifolds like V_1, k_2 submanifolds like V_2, etc.

Poincaré calls a family of n-dimensional submanifolds V_i linearly independent if there is no homology (with integer coefficients) connecting them. In honor of Enrico Betti, Poincaré defined the *n-th Betti number* of V to be $b_n + 1$, where b_n is the size of a maximal independent family. Today we call b_n the n-th Betti number, because it is the dimension of the rational vector space $H_n(V; \mathbb{Q})$. For geometric reasons, he did not bother to define the n-th Betti number for $n = 0$ or $n = \dim(V)$.

With this definition, Poincaré stated his famous Duality Theorem [148, p. 228]: for a closed oriented (m-dimensional) manifold, the Betti numbers equally distant from the extremes are equal, viz., $b_i = b_{m-i}$. Unfortunately, there was a gap in Poincaré's argument, found by Heegaard in 1898. Poincaré published a new proof in 1899, using a triangulation of V and restricting his formal sums $\sum k_i V_i$ to linear combinations of the simplices in the triangulation. Of course this restriction yields "reduced" Betti numbers which could potentially be different from the Betti numbers he had defined in 1895. Using simplicial subdivisions, he sketched a proof that these two kinds of Betti numbers agreed. (His sketch had a geometric gap, which was filled in by J.W. Alexander in 1915.) This 1899 paper was the origin of the simplicial homology of a triangulated manifold.

Poincaré's 1899 paper also contains the first appearance of what would eventually (after 1929) be called a chain complex. Let V be an oriented polyhedron. In [148, p. 295], he defined boundary matrices ε^q as follows. The (i, j) entry describes whether or not the j-th $(q-1)$-dimensional simplex in V lies on the boundary of the i-th q-dimensional simplex: $\varepsilon^q_{ij} = \pm 1$ if it is ($+1$ if the orientation is the same, -1 if not) and $\varepsilon^q_{ij} = 0$ if they do not meet. Poincaré called the collection of these matrices the *scheme* of the polyhedron, and demonstrated on p. 296 that $\varepsilon^{q-1} \circ \varepsilon^q = 0$. This is of course the familiar condition that the matrices ε^q form the maps in a chain complex, and today Poincaré's scheme is called the simplicial chain complex of the oriented polyhedron V.

Another major result in *Analysis Situs* is the generalization of the notion of Euler characteristic to higher dimensional polyhedra V. If α_n is the number of n-dimensional cells, Poincaré showed that the alternating sum $\chi(V) = \sum(-1)^n \alpha_n$ is independent of the choice of triangulation of V (modulo the gap filled by Alexander). On p. 288 he showed that $\chi(V)$ is the alternating sum of the Betti numbers b_n (in the modern sense); because of this result $\chi(V)$ is today called the *Euler–Poincaré characteristic* of V. Finally, when V is closed and $\dim(V)$ is odd, he used Duality to deduce that $\chi(V) = 0$.

In 1900, Poincaré returned once again to the subject of homology, in the *Second complément à l'Analysis Situs*. This paper is important from our perspective because it introduced linear algebra and the notion of torsion coefficients. To do this, Poincaré considered the sequence of integer matrices (or *tableaux*) T_p which describe the boundaries of the p-simplices in a polyhedron; this sequential display of integer matrices was the second occurrence of the notion of chain complex.

In Poincaré's framework, one performs elementary row and column operations upon all the matrices until the matrix T_p had been reduced to the block form

$$T_p = \begin{pmatrix} I & 0 & 0 \\ 0 & K_p & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad K_p = \begin{pmatrix} k_1 & & \\ & k_2 & \\ & & \ddots \end{pmatrix}, \quad 1 < k_1,\ k_1 | k_2,\ k_2 | k_3, \ldots$$

Here I denotes an identity matrix. The p-th Betti number b_p is the difference between the number of zero columns in T_p and the number of nonzero rows in T_{p+1} [148, p. 349]. The p-th *torsion coefficients* were defined as the integers k_1, k_2, etc. in the matrix K_{p+1} [148, p. 363].

In modern language, $H_n(V;\mathbb{Z})$ is a finitely generated Abelian group, so it has the form $\mathbb{Z}^{b_n} \oplus \mathbb{Z}/k_1 \oplus \mathbb{Z}/k_2 \oplus \cdots$ with $k_1|k_2$, $k_2|k_3$, etc. Here b_n is the Betti number, and the p-th torsion coefficients are the orders of the finite cyclic groups $\mathbb{Z}/k_i$. Of course, since homology was not thought of as a group until 1925 (see [145]), this formulation would have looked quite strange to Poincaré!

1.3. *Homology of topological spaces (1900–1935)*

The next 25 years were a period of consolidation and clarification of Poincaré's ideas. For example, the Duality theorem for the mod 2 Betti numbers, even for nonoriented manifolds, appeared in the 1913 paper [186] by O. Veblen (1880–1960) and J.W. Alexander (1888–1971). The topological invariance of the Betti numbers and torsion coefficients of a manifold was established by Alexander in 1915. In 1923, Hermann Künneth (1892–1975) calculated the Betti numbers and torsion coefficients for a product of manifolds in [119]; his results have since become known as the *Künneth Formulas.*

Until the mid 1920's, topologists studied homology via incidence matrices, which they could manipulate to determine the Betti numbers and torsion coefficients. This changed in 1925, when Emmy Noether (1882–1935) pointed out in her 14-line report [145], and in her lectures in Göttingen, that homology was an *Abelian group*, rather than just Betti numbers and torsion coefficients, and perceptions changed forever. The young H. Hopf (1894–1971), who had just arrived to spend a year in Göttingen and meet P. Alexandroff, realized how useful this viewpoint was, and the word spread rapidly. Inspired by the new viewpoint, the 1929 paper [140] by L. Mayer (1887–1948) introduced the purely algebraic notions of chain complex, its subgroup of cycles and the homology groups of a complex. Slowly the subject became more algebraic.

During the decade 1925–1935 there was a general movement to extend the principal theorems of algebraic topology to more general spaces than those considered by Poincaré. This led to several versions of homology. Some people who invented homology theories in this decade were: Alexander [3], Alexandroff (1896–1982), Čech (1893–1960) [43], Lefschetz (1884–1972) [121], Kolmogoroff (1903–1987), Kurosh (1908–1971) and Vietoris (1891–!). In 1940, Steenrod (1910–1989) developed a homology theory for compact metric spaces [176], and his theory also belongs to this movement.

In each case, the homology theory could be described as follows: given topological data, the inventors gave an ad hoc recipe for constructing a chain complex, and defined their homology groups to be the homology of that chain complex. In each case, they showed that the result is independent of choices, and provides the usual Betti numbers for compact manifolds. One theme in many recipes was homology with coefficients in a compact topological group; this kind of homology remained in vogue until the early 1950's, by which time it had become superfluous. We shall pass over most of this decade, as it played little part in the development of homological algebra per se.

One theory we should mention is the "de Rham homology" of a smooth manifold, which was introduced by G. de Rham (1903–1990) in his 1931 thesis [50]. Elie Cartan (1869–1951) had just introduced the cochain complex of exterior differential forms on a smooth manifold M in a series of papers [36, 37] and had conjectured that the Betti number b_i of M is the maximum number of closed i-forms ω_j such that no nonzero linear combination

$\sum \lambda_j \omega_j$ is exact. When de Rham saw Cartan's note [36] in 1929, he quickly realized that he could solve Cartan's conjecture using a triangulation on M and the bilinear map

$$(C, \omega) \mapsto \int_C \omega.$$

Here C is an i-cycle for the triangulation and ω is a closed i-form. Indeed, Stokes' formula shows that $\int_C \omega = 0$ if either ω is an exact form or if C is a boundary. De Rham showed the converse was true: if we fix C then $\int_C \omega = 0$ if and only if C is a boundary, while if we fix ω then $\int_C \omega = 0$ if and only if ω is exact. De Rham's theorem proves Cartan's conjecture, since if we write $H^i_{\mathrm{dR}}(M)$ for the quotient of all closed forms by the exact forms, then it gives a nondegenerate pairing between the vector spaces $H_i(M; \mathbb{R})$ and $H^i_{\mathrm{dR}}(M)$.

Of course, $H^i_{\mathrm{dR}}(M)$ is just the i-th cohomology of Cartan's complex, and we now refer to it as the "de Rham cohomology" of M. But cohomology had not been invented in 1931, and no one seems to have realized this fact until Cartan and Chevalley in the 1940's, so de Rham was forced to state his results in terms of homology. Much later, the de Rham cohomology of Lie groups would then play a critical role in the development of the cohomology of Lie algebras (see [45] and the discussion below).

1.4. *The rise of algebraic methods (1935–1945)*

The year 1935 was a watershed year for topology in many ways. We shall focus upon four developments of importance to homology theory.

The *Hurewicz maps* $h : \pi_n(X) \to H_n(X; \mathbb{Z})$ were constructed and studied by Witold Hurewicz (1904–1956) in 1935. Hurewicz also studied *aspherical* spaces, meaning spaces such that $\pi_n(X) = 0$ for $n \neq 1$. He noticed in [105] that if X and X' are two finite-dimensional aspherical spaces with $\pi_1(X) = \pi_1(X')$ then X and X' are homotopy equivalent. From this he concluded that the homology $H_n(X; \mathbb{Z})$ of such an X depended only upon its fundamental group $\pi_1(X)$. This observation forms the implicit definition of the cohomology of a group, a definition only made explicit a decade later (see below).

The homology of the classical Lie groups was calculated in 1935 by Pontrjagin [149] (Betti numbers only, using combinatorial proofs) and more fully by R. Brauer [32] (ring structure, using de Rham homology). These calculations led directly to the modern study of Hopf algebras, as follows. H. Hopf introduced H-spaces in the paper [102], written in 1939, and showed that the Brauer–Pontrjagin calculations were a consequence of the fact that the homology ring $H_*(M; \mathbb{Q})$ of any H-space M is an exterior algebra on odd generators; today we would say that Hopf's result amounted to an early classification of finite-dimensional graded "Hopf algebras" over $\mathbb{Q}$.

The third major advance was the determination of Universal Coefficient groups for homology, that is, a coefficient group A_u which would determine the homology groups $H_*(X; A)$ for arbitrary coefficients A. For finite complexes, where matrix methods apply, J.W. Alexander had already shown in 1926 [3] that $H_*(X; \mathbb{Z}/n)$ was determined by $H_*(X; \mathbb{Z})$, the case $n = 2$ having been done as early as 1912 [186]. In the 1935 paper [44], E. Čech discovered that $\mathbb{Z}$ is a Universal Coefficient group for homology: assume that there is a chain complex C_* of free Abelian groups, whose homology gives the integral homology of a space X (the space is introduced only for psychological reasons). Then for every

Abelian group A and every complex X, $H_n(X; A)$ is the direct sum of two subgroups, determined explicitly by $H_n(X; \mathbb{Z})$ and $H_{n-1}(X; \mathbb{Z})$, respectively.

In fact, Čech's Universal Coefficient Theorem gave explicit presentations for these subgroups, which today we would recognise as presentations for $H_n(X; \mathbb{Z}) \otimes A$ and $\mathrm{Tor}_1(H_{n-1}(X; \mathbb{Z}), A)$. Thus Čech was the first to introduce the general tensor product and torsion product Tor of Abelian groups into homological algebra. However, such a modern formulation of Čech's result (and the name Tor, due to Eilenberg around 1950) did not appear in print before 1951 ([39, Exposé 10]; see also [67, p. 161]). We note a contemporary variant in passing: Steenrod proved a Universal Coefficient Theorem for cohomology with coefficients in a compact topological group in 1936; see [175]; in this context the Universal Coefficient group is the character group $\mathbb{R}/\mathbb{Z}$ of $\mathbb{Z}$.

The fourth great advance in 1935 was the discovery of cohomology theory and cup products, simultaneously and independently by Alexander and Kolmogoroff. The drama of their back-to-back presentations at the Moscow International Conference on Topology in September 1935 is nicely described in Massey's article [136] in this book. The Alexander–Kolmogoroff formulas defining the cup product were completely ad hoc, and also not exactly correct; the rectification was quickly discovered by Čech and Hassler Whitney (1907–1989), and corrected by Alexander. All three authors published articles about the cup product in the Annals of Mathematics during 1936–1938. Whitney's article [190] had the most enduring impact, for it introduced the modern "co" terminology: *coboundary* (δ) and *cocycle*, as well as the notation $a \smile b$ and $a \frown b$, prophetically suggesting that "we might call $\smile$ 'cup' and $\frown$ 'cap' ". Whitney's article also implicitly introduced the notion of what we now call a differential graded algebra, via the "Leibniz axiom" that if a and b are homogeneous of degrees p and q then:

$$\delta(a \smile b) = (\delta a) \smile b + (-1)^p a \smile \delta b.$$

During the next decade, while the world was at war, the algebraic machinery slowly fell into place.

In the 1938 paper [191], Hassler Whitney discovered the tensor product construction $A \otimes B$ for Abelian groups (and modules). Up to that time, this operation had only been known (indirectly) in special cases: the tensor product of vector spaces, or the tensor product of A with a finitely generated Abelian group B. Whitney took the name from the following classical example in differential geometry: if T is the tangent vector space of a manifold at a point, then $T \otimes T$ is the vector space of (covariant) "tensors of order 2". The full modern definition of the tensor product (using left and right modules) appeared in Bourbaki's influential 1942 treatment [30], as well as in the 1944 book [11] by Artin, Nesbitt and Thrall.

The concept of an exact sequence first appeared in Hurewicz' short abstract [106] of a talk in 1941. This abstract discusses the long exact sequence in cohomology associated to a closed subset $Y \subset X$, in which the operation $\delta: H^q(X - Y) \to H^{q+1}(X, Y)$ plays a key role.

In the 1942 paper [59], Eilenberg (1915–1998) and Mac Lane (1909–) gave a treatment of the Universal Coefficient Problem for cohomology, naming Hom and Ext for the first time. Using these, they showed that Čech homology with coefficients in any Abelian group A is determined by Čech cohomology with coefficients in $\mathbb{Z}$. This application further established the importance of algebra in topology. We will say more about this discovery in the next section.

In 1944, S. Eilenberg defined singular homology and cohomology in [57]. First, he introduced the singular chain complex $S(X)$ of a topological space, and then he defined $H_*(X; A)$ and $H^*(X; A)$ to be the homology and cohomology of the chain complexes $S(X) \otimes A$ and $\mathrm{Hom}(S(X), A)$, respectively. The algebra of chain complexes was now firmly entrenched in topology. Eilenberg's definition of $S(X)$ was only a minor modification of Lefschetz' 1933 construction in [121], replacing the notion of oriented simplices by the use of simplices with ordered vertices; this trick avoided the issue of equivalence relations on oriented simplices which introduced "degenerate" chains of order 2. (See [136].)

We close our description of this era with the 1945 paper by Eilenberg and Steenrod [66]. This paper outlined an axiomatic treatment of homology theory, rederiving the whole of homology theory for finite complexes from these axioms. They also pointed out that singular homology and Čech homology satisfy the axioms, so they must agree on all finite complexes. The now-familiar axioms introduced in this paper were: functoriality of H_q and ∂; homotopy invariance; long exact homology sequence for $Y \subset X$; excision; and the *dimension axiom*: if P is a point then $H_q(P) = 0$ for $q \neq 0$. We refer the reader to May's article [139] for subsequent developments on generalized homology theories, which are characterized by the Eilenberg–Steenrod axioms with the dimension axiom replaced by Milnor's *wedge axiom* [141].

2. Homology and cohomology of algebraic systems

During the period 1940–1950, topologists gradually began to realize that the homology theory of topological spaces gave invariants of algebraic systems. This process began with the discovery that group extensions came up naturally in cohomology. Then came the discovery that the cohomology of an aspherical space Y and of a Lie group G only depended upon algebraic data: the fundamental group $\pi = \pi_1(Y)$ and the Lie algebra $\mathfrak{g}$ associated to G, respectively. This led to thinking of the homology and cohomology groups of Y and G as intrinsic to π and $\mathfrak{g}$, and therefore algebraically definable in terms of the group π and the Lie algebra $\mathfrak{g}$.

2.1. Ext *of Abelian groups*

If A and B are Abelian groups, an *extension* of B by A is an Abelian group E, containing B as a subgroup, together with an identification of A with E/B. The set $\mathrm{Ext}(A, B)$ of (equivalence classes of) extensions appeared as a purely algebraic object, as a special case of the more general problem of group extensions (see below), decades before it played a crucial role in the development of homological algebra.

Here is the approach used by Reinhold Baer in 1934 [16]. Suppose that we fix a presentation of an Abelian group A by generators and relations: write $A = F/R$, where F is a free Abelian group, say with generators $\{e_i\}$, and R is the subgroup of relations. If E is any extension of B by A, then by lifting the generators of A to elements $a(e_i)$ of E we get an element $a(r)$ of B for every relation r in R. Brauer thought of this as a function from the defining relations of A into B, so he called the induced homomorphism $a : R \to B$ a *relations function*. Conversely, he observed that every relations function a gives rise to a factor set, and hence to an extension $E(a)$, showing that two relations functions a and a' gave the same extensions if and only if there are elements b_i (corresponding to a function $b: F \to B$) so that $a'(r) = b(r) + a(r)$ [16, p. 394]. Finally, Baer observed (p. 395) that the formal sum $a + a'$ of two relations functions defined an addition law on $\mathrm{Ext}(A, B)$, making it into an Abelian group. In his honor, we now call the extension $E(a + a')$ the "Baer sum" of the extensions.

Baer's presentation $A = F/R$ amounted to a free resolution of A, and his formulas were equivalent to the modern calculation of $\mathrm{Ext}(A, B)$ as the cokernel of $\mathrm{Hom}(F, B) \to \mathrm{Hom}(R, B)$. But working with free resolutions was still a decade away [104, 72], and using them to calculate $\mathrm{Ext}(A, B)$ was even further in the future [67].

We now turn to 1941. That year, Saunders Mac Lane gave a series of lectures on group extensions at the University of Michigan. According to [133], most of the lectures concerned applications to Galois groups and class field theory, but Mac Lane ended with a calculation of the Abelian extensions of $\mathbb{Z}$ by $A = \mathbb{Z}[1/p]$. Samuel Eilenberg, who had recently emigrated from Poland and was an Instructor at Michigan, could not attend the last lecture and asked for a private lecture. Eilenberg immediately noticed that the group $\mathbb{Z}[1/p]$ was dual to the topological p-adic solenoid group Σ, which Eilenberg had been studying, and that Mac Lane's algebraic answer $\mathrm{Ext}(\mathbb{Z}[1/p], \mathbb{Z}) = \widehat{\mathbb{Z}}_p/\mathbb{Z}$ coincided with the Steenrod homology groups $H_1(S^3 - \Sigma; \mathbb{Z})$ calculated in Steenrod's 1940 paper [176]. After an all-night session, followed by several months of puzzling over this observation, they figured out how Ext plays a role in cohomology; the result was the paper [59].

Time has recognized their result as the Universal Coefficient Theorem for singular cohomology, but singular cohomology had not yet been invented in 1942. In addition, the notation then in vogue, and used in [59], was the opposite of today's conventions (which date to 1945 [66, 41]) in several respects. They wrote $H^q(A)$ for the homology groups they worked with, and wrote $H_q(A)$ for the cohomology groups under consideration. And since they were reworking many of Baer's observations about extensions, they wrote $\mathrm{Ext}\{B, A\}$ for what we call $\mathrm{Ext}(A, B)$.

Here is a translation of their Universal Coefficient Theorem into modern language. Given an infinite but star-finite CW complex K, they formed the cochain complex $C^*(K)$ of *finite* cocycles with integer coefficients; each $C^q(K)$ is a free Abelian group. Define the cohomology $\mathcal{H}^*$ of K using $C^*(K)$, and define the homology $H_*(K; A)$ of K with coefficients in A using the chain complex $\mathrm{Hom}(C^*(K), A)$. Then $H_q(K; A)$ is the product of $\mathrm{Hom}(\mathcal{H}^q, A)$ and the group $\mathrm{Ext}(\mathcal{H}^{q+1}, A)$ of Abelian group extensions. Of course, the proof in [59] only uses the algebraic properties of $C^*(K)$. Since the early 1950's [67, 41] it has been traditional to state this result the other way: given a chain complex C_* of free Abelian groups, one sets $H_q = H_q(C_*)$ and describes the cohomology of the cochain complex $\mathrm{Hom}(C_*(K), A)$ as the product of $\mathrm{Hom}(H_q, A)$ and $\mathrm{Ext}(H_{q-1}, A)$.

In order to find a universal coefficient formula for the cohomology $H^q(K; A)$ of $C^*(K) \otimes A$, they discovered the "adjunction" isomorphism

$$\mathrm{Hom}(A \otimes B, C) \cong \mathrm{Hom}\big(B, \mathrm{Hom}(A, C)\big);$$

see [59, (18.3)]. This is an isomorphism which varies naturally with the Abelian groups A, B and C. With this in hand, they reformulated Čech's Universal Coefficient Theorem: $H_q(K; A)$ is the direct sum of $H_q \otimes A$ and a group T that we would write as $T =$

$\mathrm{Hom}_{\mathrm{cont}}(A^*, H_{q-1})$, where A^* is the Pontrjagin dual of A. In fact, T is $\mathrm{Tor}(A, H_{q-1})$; see [41, p. 138].

The notion that $\mathrm{Hom}(A, B)$ varies naturally, contravariantly in A and covariantly in B, was central to the discussion in [59]. In order to have a precise language for speaking of this property for Hom, and for homology and cohomology, Eilenberg and Mac Lane concocted the notions of *functor* and *natural isomorphism* in 1942. They expanded the language to include *category* and *natural transformation* in 1945; see [61]. Although these concepts were used in several papers, the new language of Category Theory did not gain wide acceptance until the appearance of the books [67, 41] in the 1950's.

2.2. *Cohomology of groups*

The low dimensional cohomology of a group π was classically studied in other guises, long before the notion of group cohomology was formulated in 1943–1945. For example,

$$H^0(\pi; A) = A^G, \qquad H_1(\pi; \mathbb{Z}) = \pi/[\pi, \pi]$$

and (for π finite) the character group

$$H^2(\pi, \mathbb{Z}) = H^1(\pi; \mathbb{C}^\times) = \mathrm{Hom}(\pi, \mathbb{C}^\times)$$

were classical objects.

The group $H^1(\pi, A)$ of *crossed homomorphisms* of π into a representation A is just as classical: Hilbert's "Theorem 90" (1897) is actually the calculation that $H^1(\pi, L^\times) = 0$ when π is the Galois group of a cyclic field extension L/K, and the name comes from its role in the study of crossed product algebras [33].

The study of $H^2(\pi; A)$, which classifies extensions over π with normal subgroup A via factor sets, is equally venerable. The idea of factor sets appeared as early as Hölder's 1893 paper [101, Section 18], again in Schur's 1904 study [161] of projective representations $\pi \to PGL_n(\mathbb{C})$ (these determine an extension E over π with subgroup $\mathbb{C}^\times$, equipped with an n-dimensional representation) and again in Dickson's 1906 construction of crossed product algebras. O. Schreier's 1926 paper [160] was the first systematic treatment of factor sets; Schreier did not assume that A was Abelian. In 1928, R. Brauer used factor sets in [31] to represent central simple algebras as crossed product algebras in relation to the Brauer group; this was clarified in [33]. In 1934, R. Baer gave the first invariant treatment of extensions (i.e. without using factor sets) in [16]. He noticed that when A was Abelian, Schreier's factor sets could be added termwise, so that the extensions formed an Abelian group. Extensions with A Abelian were also studied by Marshall Hall in 1938 [85].

The next step came in 1941, when Heinz Hopf submitted a surprising 2-page announcement [103] to a topology conference at the University of Michigan. In it he showed that the fundamental group $\pi = \pi_1(X)$ determined the cokernel of the Hurewicz map $h : \pi_2(X) \to H_2(X; \mathbb{Z})$. If we present π as the quotient $\pi = F/R$ of a free group F by the subgroup R of relations, Hopf gave the explicit formula:

$$\frac{H_2(X; \mathbb{Z})}{h(\pi_2(X))} \cong \frac{R \cap [F, F]}{[F, R]}.$$

In particular, if $\pi_2(X) = 0$ this shows exactly how $H_2(X; \mathbb{Z})$ depends only upon $\pi_1(X)$; this formula is now called *Hopf's formula* for $H_2(\pi; \mathbb{Z})$.

Communication with Switzerland was difficult during World War II, and Hopf's paper arrived too late to be presented at the conference, but his result made a big impression upon Eilenberg. According to Mac Lane [133], Eilenberg suggested that they try to get rid of that non-invariant presentation of $\pi(X)$. Since they had just learned in 1942 [59] that homology determined cohomology, was it more efficient to describe the effect of $\pi_1(X)$ on $H^2(X; \mathbb{Z})$? Mac Lane states that this line of investigation provided the justification for the abstract study of the cohomology of groups, and "was the starting point of homological algebra" [132, p. 137].

The actual definition of the homology and cohomology of a group π first appeared in the 1943 announcement [60] by Eilenberg and Mac Lane (the full paper appeared in 1945). At this time (March 1943–1945) Eilenberg and Mac Lane were housed together at Columbia, working on war-related applied mathematics [134]. Independently in Amsterdam, Hans Freudenthal (1905–1990) discovered homology and cohomology of groups using free resolutions; his paper [72] was probably smuggled out of the Netherlands in late 1944. Also working independently of Eilenberg–Mac Lane and Freudenthal, but in Switzerland, homology was defined in Hopf's 1944 paper [104], and (based on Hopf's paper) the cohomology ring was defined in Beno Eckmann's 1945 paper [56]. We will discuss these approaches, beginning with [60].

Given π, Eilenberg and Mac Lane choose an aspherical space Y with $\pi = \pi_1(Y)$. Using Hurewicz' observation that the homology and cohomology groups of Y (with coefficients in A) were independent of the choice of Y, Eilenberg and Mac Lane took them as the definition of $H_n(\pi; A)$ and $H^n(\pi; A)$. To perform computations, Eilenberg and Mac Lane chose a specific abstract simplicial complex $K(\pi)$ for the aspherical space Y. Its n-cells correspond to ordered arrays $[x_1, \ldots, x_n]$ of elements in the group. Thus one way to calculate the cohomology groups of π was to use the cellular cochain complex of $K(\pi)$, whose n-chains are functions $f : \pi^q \to A$ from q copies of π to A. Eckmann's paper [56] also defines $H^q(X; A)$ as the cohomology of this ad hoc cochain complex, and defines the cohomology cup product in terms of this complex. Both papers showed that $H^2(G; A)$ classifies group extensions.

At the same time, Hopf gave a completely different definition in [104]. First Hopf considers a module M over any ring R, and constructs a resolution F_* of M by free R-modules. If I is an ideal of R, he considers the homology of the kernel of $F_* \to F_*/I$ and shows that it is independent of the choice of resolution. In effect, this is the modern definition of the groups $\mathrm{Tor}_*^R(M, R/I)$! Hopf then specializes to the group ring $R = \mathbb{Z}[\pi]$, the augmentation ideal I and $M = \mathbb{Z}$, and defines the homology of π to be the result. That is, Hopf's definition is literally (in modern notation)

$$H_n(\pi; \mathbb{Z}) = \mathrm{Tor}_n^{\mathbb{Z}[\pi]}(\mathbb{Z}, \mathbb{Z}).$$

Finally, Hopf showed that if Y is an aspherical cell space with $\pi = \pi_1(Y)$ then $H_n(Y; \mathbb{Z}) = H_n(\pi; \mathbb{Z})$. His proof has since become standard: the cellular chain complex F_* for the universal cover of Y is a free $\mathbb{Z}[\pi]$-resolution of $\mathbb{Z}$, and F_*/I is the cellular chain complex of Y. Thus the homology of F_*/I simultaneously computes the Betti homology of Y and the group homology of π, as claimed.

Freudenthal's method [72] was similar to Hopf's, but less general. He considered a free $\mathbb{Z}[\pi]$-module resolution F_* of $\mathbb{Z}$, and showed that the homology of $F_* \otimes_\pi A$ is independent of F_* for every Abelian group A. Like the others, Freudenthal constructed one such resolution from an aspherical polytope Y with $\pi = \pi_1(Y)$.

At first, calculations of group homology were restricted to those groups π which were fundamental groups of familiar topological spaces, using the bar complex. In his 1946 Harvard thesis [129], R. Lyndon found a way to calculate the cohomology of a group π, given a normal subgroup N such that $H^*(N)$ and $H^*(\pi/N; A)$ were known. His procedure started with $H^p(\pi/N; H^q(N))$ and proceeded through successive subquotients, ending with graded groups associated to a filtration on $H^*(\pi)$. Serre quickly realized [162] that Lyndon's procedure amounted to a spectral sequence, and completed the description with Hochschild in the 1953 paper [99]. Since then, it has been known as the Lyndon–Hochschild–Serre spectral sequence.

One application of the new definitions was *Galois cohomology*, so named in Hochschild's 1950 study [95] of local class field theory. If L is a finite Galois extension of a field K with Galois group G, this referred to the cohomology of G with coefficients in $L^\times$, or in a related G-module such as the idèle class group of L. For example, the Normal Basis Theorem implies that the additive group L is a free G-module over L, so $H^q(G; L) = 0$ for $q \neq 0$ [58]. In the late 1940's, it was observed that the factor sets of Brauer [31] and Brauer–Noether [33] were 2-cocycles, and the Brauer–Noether results translated immediately into the following theorem about the Brauer group: $H^2(G; L^\times)$ is isomorphic to the kernel $Br(L/K)$ of the map $Br(K) \to Br(L)$, and is generated by the central simple algebras which are split over L. This observation was mentioned in Eilenberg's 1948 survey [58] of the field. A careful writeup was given by Serre in Cartan's 1950/1951 seminar [39].

The 1952 paper [98] explored the connection to Class Field Theory, translating Tsen's theorem (1933) into the vanishing of $H^q(G; K^\times)$ for $q \neq 0$ when k and K are function fields of curves over an algebraically closed field. This paper also marked the first appearance of *Shapiro's lemma*, a formula for the cohomology of an induced module which is due to Arnold Shapiro.

While studying Galois cohomology in his 1952 thesis [182], John Tate discovered that there is a natural isomorphism $H^r(G; \mathbb{Z}) \cong H^{r+2}(G; C_L)$, where C_L is the idèle class group of a number field L. Moreover, the reciprocity law gave a similar relation between $H_1(G; \mathbb{Z}) = G/[G, G]$ and a subgroup of $H^0(G; C_L)$. This led him to define the *Tate cohomology* $\widehat{H}^*(G, A)$ of any finite group G and any G-module A, indexed by all integers; see the 1954 paper [183]. Tate did this by splicing together the cohomology of G ($\widehat{H}^r(G; A) = H^r(G; A)$ for $r > 0$) and the homology of G (reindexing via H_n as $\widehat{H}^{-n-1}$ for $n \geqslant 1$), and using ad hoc definitions for $\widehat{H}^0$ and $\widehat{H}^{-1}$.

The 1950/1951 Seminaire Cartan [39] saw the next major advances in group homology. In Exposés 1 and 2, Eilenberg gave an axiomatic characterization of homology and cohomology theories for a group π, and used a fixed free resolution of the π-module $\mathbb{Z}$ to establish the existence of both a homology and a cohomology theory. The key axioms Eilenberg introduced to prove uniqueness were:

(1) if A is a free π-module then $H_q(\pi; A) = 0$ for $q > 0$, and

(2) if A is an injective π-module then $H^q(\pi; A) = 0$ for $q > 0$.

In Exposè 4 of the same seminar, H. Cartan proved what we now call the Comparison Theorem for chain complexes: given a free resolution C_* and an acyclic resolution C'_* of $\mathbb{Z}$, there is a chain map $C_* \to C'_*$ over $\mathbb{Z}$, unique up to chain homotopy. This made Eilenberg's construction natural in the choice of C_*, and allowed Cartan the freedom to construct cup products in group cohomology via resolutions.

After the 1950–1951 Seminaire Cartan [39], the germs of a complete reworking of the subject were in place. Cartan and Eilenberg began to collaborate on this reworking, not realizing that the resulting book [41] would take five years to appear.

2.3. *Associative algebras*

Before the cohomology theory of associative algebras was defined, the special cases of derivations and extensions had been studied. Derivations and inner derivations of algebras (associative or not) over a field k were first studied systematically in 1937 by N. Jacobson [107], who was especially interested in the connection to Galois theory over k when $\mathrm{char}(k) \neq 0$.

Hochschild studied derivations of associative algebras and Lie algebras in the 1942 paper [93]. He showed that every derivation of an associative algebra A is inner if and only if A is a *separable* algebra, meaning that not only is A semisimple, but the ℓ-algebra $A \otimes_k \ell$ is semisimple for every extension field $k \subseteq \ell$. In addition, he showed that if A is semisimple over a field of characteristic zero, M is an A-bimodule, and $f : A \otimes A \to M$ is a bilinear map satisfying the *factor set condition*:

$$a\, f(b, c) + f(a, bc) = f(a, b)\, c + f(ab, c),$$

then there is a linear map $e : A \to M$ so that $f(a, b) = a\, e(b) + e(a)b - e(ab)$.

Upon seeing the Eilenberg–MacLane treatment of the cohomology of groups in 1945, Hochschild observed [94] that the same formulas gave a purely algebraic definition of the cohomology of an associative algebra A over a field, with coefficients in a bimodule M. The degree q part $C^q(A; M)$ of his ad hoc cochain complex is the vector space of multilinear maps from A to M, i.e. linear maps $A^{\otimes q} \to M$. For example, if $e : k \to M$ has $e(1) = m$ then $\delta(e)(a) = am - ma$ is an inner derivation, and a 1-cocycle is a map $f : A \to M$ such that $f(ab) = af(b) + f(a)b$. Thus the construction makes $H^1(A; M)$ into the quotient of all derivations by inner derivations, and the first of Hochschild's 1942 results becomes: $H^1(A; M)$ vanishes for every M if and only if A is a *separable* algebra. Hochschild also showed that $H^2(A; M)$ measures algebra extensions E of A by M, meaning that M is a square-zero ideal and $E/M \cong A$; a trivial extension is one in which the algebra map $E \to A$ splits. Since a 2-cocycle is just a map $f : A \otimes A \to M$ satisfying the factor set condition mentioned above, Hochschild's second 1942 result becomes: if A is semisimple then $H^2(A; M)$ vanishes for every M, and hence every nilpotent algebra extension of A must be split.

2.4. *Lie algebras*

Since Elie Cartan's 1929 theorem [37] that every connected semisimple Lie group is diffeomorphic to the product of a compact Lie group G and $\mathbb{R}^n$, the cohomology of Lie groups was reduced to that of compact Lie groups. Cartan conjectured in 1928 [36] and de Rham observed that the de Rham cohomology $H^*_{\text{dR}}(G; \mathbb{R})$ of G may be computed using left invariant differentials, and it was gradually noticed that the Lie algebra $\mathfrak{g}$ of left invariant vector fields (or tangent vectors at the origin of G) determines the cohomology of G. We have seen how this was solved in 1935 by Brauer and Pontrjagin.

Chevalley and Eilenberg were able to use this observation to define the cohomology of any Lie algebra in their 1948 paper [45]. After reviewing de Rham cohomology, they calculate that the (differential graded) algebra of left invariant differential forms on a Lie group G is isomorphic to the dual algebra $C^*(\mathfrak{g})$ of the exterior algebra $\wedge^*\mathfrak{g}$. Translating the de Rham differential into this context gave the differential $\delta : C^q(\mathfrak{g}) \to C^{q+1}(\mathfrak{g})$ defined by

$$(\delta\omega)(x_1, \dots, x_{q+1}) = \frac{1}{q+1} \sum (-1)^{k+l+1} \omega\big([x_k, x_l], \dots, \hat{x}_k, \dots, \hat{x}_l, \dots\big).$$

This makes $C^*(\mathfrak{g})$ into a differential graded algebra, and they define the cohomology ring $H^*_{\text{Lie}}(\mathfrak{g}; \mathbb{R})$ of the Lie algebra $\mathfrak{g}$ to be the cohomology of $C^*(\mathfrak{g})$. (They then state that in other characteristics one can and should omit the constant $1/(q+1)$.) Thus if G is compact and connected then their construction of Lie algebra cohomology has the isomorphism $H^*_{\text{dR}}(G; \mathbb{R}) \cong H^*_{\text{Lie}}(\mathfrak{g}; \mathbb{R})$ as its birth certificate.

It is immediate that a 1-cocycle is a map $\mathfrak{g} \to \mathbb{R}$ vanishing on the subalgebra $[\mathfrak{g}, \mathfrak{g}]$. Since there are no 1-coboundaries we see that $H^1_{\text{Lie}}(\mathfrak{g}; k)$ is the dual space of $\mathfrak{g}/[\mathfrak{g}, \mathfrak{g}]$. This purely algebraic feature is present, but had been downplayed in the cohomology of compact connected Lie groups, because it follows from the fact that $G/[G, G]$ is a torus.

In order to study the cohomology $H^*_{\text{dR}}(G/H; \mathbb{R})$ of the homogeneous spaces G/H of G, Chevalley and Eilenberg also defined the cohomology $H^*_{\text{Lie}}(\mathfrak{g}; V)$ of a representation V of $\mathfrak{g}$. This was defined similarly, as the cohomology of the chain complex $C^*(\mathfrak{g}; V)$ of (vector space) maps from $\wedge^*\mathfrak{g}$ to V. Translated from the corresponding de Rham differential on the manifold G/H, the formula for the differential $\delta\omega$ resembled the one displayed above, but it had an extra alternating sum of terms $x_k\omega(\dots, \hat{x}_k, \dots)$.

According to Jacobson's 1937 paper [107], a derivation from a Lie algebra $\mathfrak{g}$ into a $\mathfrak{g}$-module V is a linear map $D : \mathfrak{g} \to V$ such that $D([x, y]) = x(Dy) - y(Dx)$. It is an inner derivation if $D(x) = xv$ for some $v \in V$. It is immediate from the Chevalley–Eilenberg complex $C^*(\mathfrak{g}; V)$ that $H^1_{\text{Lie}}(\mathfrak{g}; V)$ is the quotient of all derivations from $\mathfrak{g}$ into M by the inner derivations.

The 1948 paper [45] also contains the theorem that Lie extensions of $\mathfrak{g}$ by V are in one-to-one correspondence with elements of $H^2(\mathfrak{g}; V)$, a result inspired by Eilenberg's role in the earlier classification of group extensions via $H^2(G; A)$ in [60]. Indeed, the proof was similar: cocycles in the complex $C^*(\mathfrak{g}; V)$ are recognised as factor sets for extensions.

Now suppose that $\mathfrak{g}$ is any semisimple Lie algebra over a field k of characteristic zero. J.H.C. Whitehead (1904–1960) had discovered some algebraic lemmas about linear maps on $\mathfrak{g}$ in 1936–1937 (see [189]), in order to give a purely algebraic proof of Weyl's 1925 theorem that every representation is completely reducible. Whitehead's lemmas also appeared in Hochschild's 1942 paper [93] on derivations. Whitehead's "first lemma" said that every derivation from $\mathfrak{g}$ into any representation V was inner, even though he proved this result before the notion of derivation was known. Chevalley and Eilenberg translated Whitehead's "first lemma" as the statement that $H^1_{\text{Lie}}(\mathfrak{g}; V) = 0$ for all V.

Whitehead's "second lemma" concerned alternating bilinear maps $f: \mathfrak{g} \wedge \mathfrak{g} \to V$ satisfying a *factor set condition*, which we would now write as $\delta f(x, y, z) = 0$. Whitehead proved that for every such f there was always a linear map $e: \mathfrak{g} \to V$ so that $f(x, y) = xe(y) - ye(x) + e([x, y])$. Chevalley and Eilenberg translated this as the statement that $H^2_{\mathrm{Lie}}(\mathfrak{g}; V) = 0$ for all V. In both of these results, the first step was an analysis of the trivial representation $V = k$. For the second step, they used another result of Whitehead to show that when $V \neq k$ is a simple representation then $H^q_{\mathrm{Lie}}(\mathfrak{g}; V) = 0$ for all q. This last step shows that the only interesting cohomology groups of $\mathfrak{g}$ are those with trivial coefficients, and these are interesting because $H^q_{\mathrm{Lie}}(\mathfrak{g}; k) = H^q(G; k)$.

The analogy with the cohomology of compact Lie groups was pursued further by Koszul (1921–) in the 1950 paper [118]. He introduced the notion of a reductive Lie algebra $\mathfrak{g}$, and showed that (in characteristic zero) its cohomology is an exterior algebra.

2.5. *Sheaves and spectral sequences*

Jean Leray (1906–1998) was a prisoner of war during World War II, 1940–1945. He organized a university in his prison camp and taught a course on algebraic topology. At the end of his imprisonment, he invented *sheaves* and *sheaf cohomology* [122], as well as *spectral sequences* for computing his sheaf cohomology [123].

As we saw above, the essential features of a spectral sequence had also been noted independently by R. Lyndon [129], as a way to calculate the cohomology of a group. The algebraic construction of spectral sequences was codified by Koszul [117] in 1947, using Cartan's suggestion that the central object should be a filtered chain complex. Koszul's presentation clarified things so much that Leray immediately adopted Koszul's framework.

In 1947/1948, and again in 1949/1950, Leray gave a course at the Collège de France on this new cohomology theory. Part I was a review of his theory of spectral sequences, using Koszul's framework. Part II introduced the notion of a sheaf, fine sheaves, and the cohomology with compact supports of a locally compact topological space relative to a differential graded sheaf. The details of this course eventually appeared in Leray's detailed 1950 article [124].

The next year (1948/1949), Henri Cartan ran a Seminar [38] on algebraic topology, with 17 exposés published as unbound mimeographed notes. Exposés XII–XVII were devoted to an exposition of Leray's theory using Cartan's version of sheaves, but were later surpressed when Cartan's viewpoint on sheaves changed. The same subject was revisited by H. Cartan two years later in Exposés 14–20 of the 1950/1951 Cartan Seminar [39], where he and his students reworked the theory of sheaves, and sheaf cohomology, based on Leray's notion of a "fine" sheaf.

In Exposé 16 Cartan gave axioms for sheaf cohomology theory on a paracompact space X (with supports in a family Φ of closed subspaces of X, which we shall omit from our notation here). His axioms were:

- $H^0(X, F)$ is the group $\Gamma(F)$ of global sections of the sheaf F (with support in Φ);
- $H^q(X, F)$ depends functorially on F and vanishes for negative q;
- A natural long exact cohomology sequence exists for each short exact sequence of sheaves; and
- If F is a "fine" sheaf then $H^q(X, F) = 0$ for all $q \neq 0$.

Cartan was now able to mimic the proof of existence and uniqueness for group cohomology given earlier in Exposés 1–4 of the same Seminar by Cartan and Eilenberg. To prove uniqueness, he observed that every sheaf F may be embedded in a fine sheaf, specifically into a sheaf he called $F \otimes S$, which we would describe as the sum of the skyscraper sheaves $x_* x^*(F)$ over all points x of X. To prove existence, Cartan fixed a resolution

$$0 \to \mathbb{Z} \to C_0 \to \cdots$$

of $\mathbb{Z}$ by torsion-free fine sheaves, and set $H^q(X, F) = H^q(\Gamma(C \otimes F))$. Observing that some choices of C happen to give differential graded algebras, he was able to define a product structure

$$H^p(X, F) \otimes H^q(X, F') \to H^{p+q}(X, F \otimes F')$$

on sheaf cohomology.

In the remaining exposés of [39], Cartan, Eilenberg and Serre returned to Leray's spectral sequences, codifying the machinery and studying its multiplicative structure. Much of this material was reproduced in the 1953 Hochschild–Serre paper [99] in order to redo Lyndon's spectral sequence [129]. The usefulness of this approach to spectral sequences was decisively demonstrated by Serre in his 1951 thesis [163].

A completely different approach to spectral sequences was given by W. Massey in 1952 [135]. Massey defines an *exact couple* to be a pair of (graded) modules D and E, equipped with maps fitting into an exact sequence

$$D \xrightarrow{i} D \xrightarrow{j} E \xrightarrow{k} D \xrightarrow{i} D.$$

One forms its *derived couple* by considering $D_1 = i(D)$ and the homology E_1 of E with respect to the differential jk. By an iterative process, one obtains a sequence of derived couples, and the sequence of modules E_r forms a spectral sequence. The exact couple approach to spectral sequences has since become very popular with topologists, but less so with algebraists.

Godement's 1958 book [79] summarized and refined all these developments, becoming the standard reference for sheaves, sheaf cohomology and spectral sequences for many years. In Godement's approach, the focus moved away from Cartan's notion of "fine" sheaf and towards the new notions of *flasque* and *soft* (*mou*) sheaves. One trick introduced by Godement, but implicit in Cartan's 1950/1951 seminar [39], was that by iteration of the canonical embedding of F into $F \otimes S$ one could get a resolution of F by injective sheaves which is functorial in F; nowadays it is called the *Godement resolution* of F.

3. The Cartan–Eilenberg revolution

As we have mentioned, Cartan and Eilenberg began collaborating during the 1950/1951 Seminaire Cartan [39], rewriting the foundations of all the ad hoc algebraic homology and cohomology theories that had arisen in the previous decade. Coining the term *Homological Algebra* for this newly unified subject, and using it for the title of the 1956 textbook [41], they revolutionized the subject.

The first occurrence of the notation Tor_n and Ext^n, as well as the concepts of projective module, derived functor and hyperhomology appeared in this book. In his review of their book, Hochschild stated that "The appearance of this book must mean that the experimental phase of homological algebra is now surpassed".

Before we describe the innovations in their book further, let us back up and review the evolution of the two main tools that were now available, namely chain complexes and resolutions.

3.1. *Chain complexes*

The algebra of chain complexes had been slowly evolving since their formal introduction in 1929 by Mayer [140]. We have already mentioned Hurewicz' 1941 discovery of the notion of exact sequence [106], and the application of this notion in the 1945 axiomatization of homology theory [66].

The next step was taken in 1947 by Kelley and Pitcher [115], who coined the term "exact sequence" and first systematically studied chain complexes from an algebraic point of view. They showed that direct limits preserve exact sequences (axiom AB5 holds), but that inverse limits do not (axiom AB5* fails). If A_* is a subcomplex of B_*, with quotient C_*, they constructed the boundary map $\partial : H_q(C) \to H_{q-1}(A)$ and proved that the long homology sequence

$$\cdots \to H_n(A) \to H_n(B) \to H_n(C) \xrightarrow{\partial} H_{n-1}(A) \to \cdots$$

is exact. Since they restricted themselves to positive complexes (indexed by positive integers q), their sequence ended in $H_0(B) \to H_0(C) \to 0$.

The yoga of chain complexes was further developed in Eilenberg and Steenrod's 1952 book [67]; cf. [66]. They indexed their chain complexes by all integers, and observed that cochain complexes could be identified as chain complexes via the reindexing $C_q = C^{-q}$. The familiar "five-lemma" occurs for the first time on [67, p. 16]. (Its companion, the "snake lemma", first appeared in [41].) Eilenberg and Steenrod's book also introduced the "Mayer–Vietoris" sequence for a space $X = U \cup V$, associated to the excision isomorphism $H_*(U, U \cap V) \cong H_*(X, V)$.

3.2. *Free and injective resolutions*

Free resolutions have long been used in algebra, starting with David Hilbert (1862–1943) in his 1890 paper [91] on iterated syzygies of a finitely generated graded module M over a polynomial ring $R = k[x_1, \ldots, x_n]$. A choice of $b_0 = \dim(M \otimes_R k)$ homogeneous generators of M defines a surjection $R^{b_0} \to M$, and its kernel is the first *syzygy module* of M. (There is a grading on R^{b_0} which we are ignoring.) Hilbert proved that the syzygy was also finitely generated (the Hilbert Basis Theorem), so one could use induction to define the higher syzygy modules. Hilbert's syzygy theorem states that the $(n+1)$-st syzygy is always zero, i.e. the n-th syzygy is R^{b_n} for some b_n. Since the number of generators b_i of the syzygies is chosen minimally, they are independent of the choices of generators: today we know this is so because b_i is the dimension of the vector space $\mathrm{Tor}_i^R(M, k)$. By analogy with topology, the b_i are called the *Betti numbers* of M.

As we have remarked, Baer [16] implicitly used free resolutions of an Abelian group in 1934 to study the groups $\mathrm{Ext}(A, B)$. The next explicit use of free resolutions was by Hopf in 1944 [104]. As we have mentioned above, he used them to describe the homology of a group, and implicitly gave a definition of the modules $\mathrm{Tor}_i^R(M, R/I)$ for any ideal I of any ring R. Based on Hopf's work, Cartan and Eilenberg used free $\mathbb{Z}[\pi]$-resolutions of a π-module A in 1950 [39] to give an axiomatic description for the group homology $H_*(G; A)$.

Injective R-modules were introduced and studied in 1940 by R. Baer [17]. Baer called them "complete" Abelian groups over the ring R; the name *injective* apparently first arose in Eilenberg's 1950 survey paper [39]. Baer's paper contains the proposition that every module is a submodule of an injective module, and what is now called "Baer's criterion" for M to be injective: every map from an ideal into M must extend to a map from R into M. Finally, Baer characterized semisimple rings as those for which every module is injective.

In the 1948 paper [130], Mac Lane formulated the projective and injective lifting properties for the category of Abelian groups, and showed that these properties describe free and divisible Abelian groups, respectively. He did not discover the notion of projective module because he did not apply these lifting properties to categories of modules. Using this, he showed that one could compute $\mathrm{Ext}(A, B)$ by embedding the Abelian group B in a divisible group D; this amounts to the use of an injective resolution of B.

3.3. *Cartan and Eilenberg: the book*

We now turn to the contents of the book [41] itself. On p. 6 it introduced an entirely new concept: the definition of a *projective* module. It proved on p. 11 that every R-module is projective if and only if R is semi-simple, complementing Baer's characterization of semisimplicity in terms of injective modules; later in the book (p. 111), this was viewed as the characterization of rings of global dimension 0.

In Chapter II the authors introduced the notion of left exact functors (such as Hom) and right exact functors (such as $\otimes_R$). In the central Chapter V, they introduced the notions of *projective resolutions* $\cdots \to P_0 \to M$ and *injective resolutions* $M \to I^0 \to \cdots$ of a module M, and used these to define the derived functors $L_nT(M) = H_nT(P_*)$ and $R^nT(M) = H^nT(I^*)$ of an additive functor T. This material was clearly based on the ideas in the 1950/1951 Seminaire Cartan [39].

In Chapter VI, the authors defined $\mathrm{Tor}_n^R(M, N)$ and $\mathrm{Ext}_R^n(M, N)$ as the derived functors of $M \otimes_R N$ and $\mathrm{Hom}_R(M, N)$. Then they defined the projective and injective dimension of M as the length of the shortest projective and injective resolution, and characterized these dimensions in terms of the vanishing of $\mathrm{Ext}_R^n(M, -)$ and $\mathrm{Ext}_R^n(-, M)$, respectively. This led them to define the (left and right) *global dimension* of R as the largest n such that Ext_R^n is nonzero, and the *weak global dimension* (now called the Tor-*dimension*) as the largest n such that Tor_n^R is nonzero.

Chapters VIII–XIII unified the homology of augmented algebras, Hochschild's homology and cohomology of an associative algebra Λ (as Tor and Ext groups over the enveloping algebra $\Lambda \otimes \Lambda^{\mathrm{op}}$), the homology and cohomology of a group π (as Tor and Ext groups over the group ring $\mathbb{Z}[\pi]$), and the homology and cohomology of a Lie algebra $\mathfrak{g}$ (as Tor and Ext groups over the enveloping algebra $U\mathfrak{g}$).

Chapters XV–XVI contained a very readable introduction to spectral sequences for filtered chain complexes, and applications to computing Ext and Tor. Again, this material was based on the ideas in the 1950/1951 Seminaire Cartan [39].

The final chapter (XVII) concerned the hyperhomology of a functor T applied to a chain complex A. This was the precursor to the discovery (in 1963) of the Derived Category by Grothendieck and Verdier [187]. First they defined double complexes they called "projective" and "injective" resolutions of A; since 1966 [88] we call them *Cartan–Eilenberg resolutions* of A. Then they defined the hyperhomology $\mathbb{L}_*T(A)$ and hypercohomology $\mathbb{R}^*T(A)$ to be the (co)homology of the total complex of T applied to the double complex resolutions.

Until 1970, [41] was the bible on homological algebra, although Mac Lane's 1963 book [132] was also popular. These texts helped the subject become standard course material. Grothendieck's 1957 Tôhoku paper [81], which we shall describe below, and later his multi-volume tome [83] on the foundation of sheaf cohomology in Algebraic Geometry, were also heavy favorites. A second generation of texts appeared in 1970/1971: Rotman's *Notes on Homological Algebra* [159] and Hilton and Stammbach's book [92].

3.4. *Abelian categories*

As soon as Cartan and Eilenberg began their undertaking, limiting themselves to functors defined on modules, it was clear that there was more than a formal analogy with the cohomology of sheaves, and that their methods worked in a more general setting. The search for that setting led to the notion of an Abelian category.

The first attempt to formulate a setting in which homological algebra made sense was by Mac Lane in 1948 [130]. In this paper Mac Lane introduced what he called "Abelian categories," but which were actually additive categories with special objects resembling the objects $\mathbb{Z}$ and $\mathbb{Q}/\mathbb{Z}$ in the category **Ab** of Abelian groups. The category of Abelian semigroups was an Abelian category in Mac Lane's sense. This notion never caught on, though.

The appendix to [41] contained the next attempt, by D. D. Buchsbaum. It was actually a summary without proofs of his 1955 thesis [35], written under Eilenberg. In attempting to formulate a general setting in which the theory in Cartan–Eilenberg could be generalized, he needed categories which had a natural notion of an exact sequence. To this end, Buchsbaum introduced the notion of an *exact category*, which is an Abelian category without the requirement that direct sums exist. To handle functors of more than one variable, he introduced the extra axiom (V) that direct sums $A \oplus B$ exist, which is equivalent to the definition of an Abelian category. Buchsbaum also introduced axioms that the category has *enough projectives* or *enough injectives*. These axioms, unnecessary for the categories of modules considered in [41], allowed Buchsbaum to carry over verbatim the Cartan–Eilenberg construction of derived functors to exact categories.

The name *Abelian category* is due to A. Grothendieck [81] and A. Heller [89]. Grothendieck's paper was motivated by the observation that the category $\mathbf{Sh}(X)$ of sheaves of Abelian groups on a topological space X was an Abelian category with enough injectives, so that sheaf cohomology could be defined as the right derived functors of the global sections functor, while Heller was more concerned with a formal analogy to stable homotopy (where syzygy modules correspond to loop spaces, and projective modules correspond to contractible spaces).

Grothendieck's 1957 "Tôhoku" paper [81] introduced a hierarchy of axioms (AB3)–(AB6) and (AB3*)–(AB6*) that an Abelian category may or may not satisfy. Axioms (AB3) and (AB3*) specify that set-indexed coproducts and products exist, respectively. The Abelian category $\mathbf{Sh}(X)$ satisfies axiom (AB5), that filtered colimits of exact sequences are exact, but not axiom (AB4*), which states that a product of surjections is a surjection.

Given this framework, Grothendieck proceeded to generalize Cartan and Eilenberg's treatment of derived functors, introducing the names ∂-functor and *universal* ∂-functor, as well as the notion of T-acyclic objects (in [41, p. 122] flat modules were defined as Tor-acyclic modules; Grothendieck showed that Godement's flasque sheaves were Γ-acyclic sheaves). The primary computational tool introduced by Grothendieck was a special case of the hypercohomology spectral sequence for the composition TU of two functors (see the last page of [41]). Grothendieck observed that if T and U were left exact, and if U sends injective modules to T-acyclic modules then we could write the spectral sequence as

$$\left(R^pT\right)\left(R^qU\right) \Longrightarrow R^{p+q}(TU).$$

Several of the spectral sequences in [41] were seen to be simple special cases of Grothendieck's spectral sequence, but so were the Leray spectral sequences associated to a continuous map $f: Y \to X$ and a sheaf F on Y:

$$H^p\left(X, R^q f_* F\right) \Longrightarrow H^{p+q}(Y, F).$$

Even the simplest of lemmas (such as the snake lemma) were painfully difficult to prove in a general Abelian category, because one could not chase elements that did not exist. This technique of diagram-chasing was justified in 1960, when Saul Lubkin [128], A.P. Heron (1960 Oxford thesis) and J.P. Freyd (1960 Princeton thesis) proved that every small Abelian category admits an exact embedding into the category of Abelian groups. Shortly thereafter, Freyd and Barry Mitchell proved a stronger version: every small Abelian category admits a full exact embedding into the category of modules over some ring (see [73]). With this result, and P. Gabriel's 1962 thesis [74], the subject was near maturity.

4. After the Cartan–Eilenberg revolution

Upon the publication of Cartan and Eilenberg [41], there was an explosion of research in homological algebra. Some results appeared to be fairly isolated curiosities at the time, but later became important, such as Yoneda's definition of Ext^n groups by long exact sequences in 1954 [192], the 1961 study of $\lim^1$ by J.E. Roos [157], the 1962 Eilenberg and Moore paper [64] on spectral sequences for complete filtered complexes, Giraud's 1965 work [78] on non-Abelian H^1 in a Grothendieck Topos, or Boardman's influential 1981 preprint [26] on conditional convergence in spectral sequences. In this article we shall focus upon the strands of thought that have led to flourishing new fields of study.

4.1. *Projective modules*

When the notion of projective module was introduced in the book [41], there were not many examples of projective modules which were not free. By [41, p. 157], all finitely generated projective modules over a local ring are free. By [41, p. 13], all projective modules over a principal ideal domain (or more generally a Bezóut domain) are free. Kaplansky later showed [112] that *all* projective modules over a local ring are free, as a consequence of the general result that any infinitely generated projective module is a direct sum of countably generated projective modules.

If I is an ideal of an integral domain R, Cartan and Eilenberg showed that I was projective if and only if it was *invertible*: $I \cdot I^{-1} = R$. Moreover, if $\dim(R) = 1$ then invertible ideals have at most two generators, so $I \oplus I^{-1} \cong R \oplus R$. Since every ideal in a Dedekind domain is invertible – their isomorphism classes forming the *Picard class group* of R – and the integers in a number ring were Dedekind domains whose class groups were classical objects of study, some examples of nonfree projective modules were already known in the late 19-th century.

For some rings, it was possible to classify all projective modules. A *Prüfer domain* is a commutative domain in which every finitely generated ideal is invertible; this generalization of Dedekind domains is named for H. Prüfer, who initiated their study in 1923. Kaplansky [111] showed in 1952 that if R is a Prüfer domain then every finitely generated torsion-free module – hence every projective module – is a direct sum of invertible ideals; see [41, pp. 13, 133].

For other rings, the classification was much harder. In Serre's classic 1955 paper [164, p. 243], he stated that it was unknown whether or not every projective R-module was free when R is a polynomial ring over a field. This became known as the "Serre problem", and was not solved (affirmatively) until 1976, by Quillen [154] and Suslin [177].

In the period 1958–1962 there was a flurry of examples of nonfree projective modules, coming from algebraic geometry [29, 167], arithmetic [19], group rings [178] and topological vector bundles [179]. Much of this was based upon the dictionary in Serre's 1955 paper [164], between projective modules and topological vector bundles. Grothendieck's Riemann–Roch theorem, published in 1958 [29], showed that the "projective class group" $K(R)$ of stable isomorphism classes of projective modules was useful, especially for rings coming from algebra and algebraic geometry. Bass, Serre and Swan began a study of the projective class group $K(R)$; by 1964 it was renamed $K_0(R)$ in view of its parallels to topological K-theory, and this led to the rise of algebraic K-theory in the 1960's.

4.2. *Homological algebra and ring theory*

The left and right global dimension of a ring were early targets. In the 1955 paper [12], M. Auslander (1926–1994) showed that the left and right global dimension of a Noetherian ring agree, and equal the weak global dimension. Then M. Harada [86] showed (1956) that the rings with weak global dimension 0 are precisely the von Neumann regular rings, so the weak dimension and global dimension need not agree. Examples in which the left and right global dimensions of a ring are different were not known until a decade later, and were found by Osofsky [146].

4.2.1. *Regular local rings.* A *regular* local ring is a commutative Noetherian local ring R whose maximal ideal $\mathfrak{m}$ is generated by a regular sequence, or equivalently, such that $\dim(\mathfrak{m}/\mathfrak{m}^2) = \dim(R)$. Regular local rings had become important in Algebraic Geometry because they were the local coordinate rings of smooth algebraic varieties. In 1956, Auslander and Buchsbaum [13] and Serre [166] used homological methods to characterize regular local rings as those (Noetherian) local rings R with finite global dimension. If R is local with residue field k and $\dim_k(\mathfrak{m}/\mathfrak{m}^2) = n$, Serre proved that $\mathrm{Tor}_n^R(k,k) \neq 0$. Hence, $\mathrm{gl.dim}(R) \geqslant n$, and $n \geqslant \dim(R) \geqslant \mathrm{depth}(R)$. Auslander and Buchsbaum proved that the depth of R is an upper bound for the finite values of $\mathrm{pd}_R(M)$, so if pd_R is always finite we must have equality: $\mathrm{gl.dim}(R) = \dim_k(\mathfrak{m}/\mathfrak{m}^2) = \dim(R)$. In particular, if $\mathrm{gl.dim}(R) < \infty$ then R must be regular.

Since localization cannot increase global dimension, a corollary is that any localization of a regular local ring is again a regular ring. This nonhomological statement, proven by homological methods, firmly established homological algebra as a central tool in ring theory; the alternate nonhomological proof of this localization result, due to Nagata [143], is very long and hard.

Also in [13], Auslander and Buchsbaum proved that 2-dimensional regular local rings are Unique Factorization Domains (UFDs). A few years later, in 1959, Auslander and Buchsbaum [15] used similar homological methods to prove that every regular local ring is a Unique Factorization Domain.

Two timely courses on this material, by Serre in France and Kaplansky in the U.S., had a lasting impact upon the field.

In 1957/1958, Serre taught a course on multiplicities at the Collège de France [169]. Part of that course focussed upon the simple inequality $\mathrm{pd}_R(M) \leqslant \mathrm{pd}_R(S) + \mathrm{pd}_S(M)$ for a module M over an R-algebra S (an exercise in [41, p. 360]). Auslander and Buchsbaum realized (1958) that Serre's methods could be used to study the connection between the codimension and multiplicity over a local ring; see [14]. This led them to the *Auslander–Buchsbaum Equality*: if M is a finitely generated module over a local ring R and $\mathrm{pd}_R(M) < \infty$ then $\mathrm{depth}(R) = \mathrm{depth}(M) + \mathrm{pd}_R(M)$.

In Fall 1958, Kaplansky taught a course [113] on homological algebra at the University of Chicago. Several students attending this course would later make important contributions to the subject: H. Bass, S. Chase, E. Matlis and S. Shanuel.

Kaplansky's course was organized around three "change of rings" theorems, describing how homological dimension changes when one passes from a ring R to a quotient ring $R/(x)$. They allowed him to prove the theorems of Serre and Auslander–Buchsbaum without having to first develop Ext or Tor. Early in the course, Shanuel noticed that there was an elegant relation between different projective resolutions of the same module. Kaplansky seized upon this result as a way to define projective dimension, and christened it "Shanuel's lemma". Subsequently it was discovered that H. Fitting had proven Shanuel's lemma in 1936 [71] (with "projective" replaced by "free") as part of his study of the fitting invariants of a module.

4.2.2. $\mathrm{Tor}_*(k,k)$ *for local rings.* Consider a local ring R with maximal ideal $\mathfrak{m}$ and residue field $k = R/\mathfrak{m}$. Cartan and Eilenberg had shown that $\mathrm{Tor}_*^R(k,k)$ was a graded-commutative k-algebra [41, XI.4 and XI.5]. Its Hilbert function is just the sequence of Betti numbers $b_i = \dim \mathrm{Tor}_i^R(k,k)$, and it is natural to consider the *Poincaré–Betti series*

$$P_R(t) = \sum_{i=0}^{\infty} b_i t^i.$$

Note that the first Betti number is $b_1 = \dim(\mathfrak{m}/\mathfrak{m}^2)$. For example, if R is a regular ring, it was well known that $\mathrm{Tor}_*^R(k,k)$ was an exterior algebra, so that $P_R(t) = (1+t)^{b_1}$.

Serre showed in 1955 [166] that one always had $P_R(t) \geqslant (1+t)^{b_1}$, i.e. that b_i is at least $\binom{b_1}{i}$. In particular, if $i = b_1$ then $b_i \geqslant 1$ and so $\mathrm{Tor}_{b_1}^R(k,k) \neq 0$. As we mentioned above, this was the key step in Serre's proof that local rings of finite global dimension are regular. In his 1956 study [184], Tate showed that k had a free R-module resolution F_* which was a graded-commutative differential graded algebra, and used this to show that if R is not regular then $P_R(t) \geqslant (1+t)^{b_1}/(1-t^2)$, i.e. that b_i is at least $\binom{b_1}{i} + \binom{b_1}{i-2} + \cdots$. This is the best lower bound. In case R is the quotient of a regular local ring by a regular sequence of length r (contained in the square of the maximal ideal), Tate showed that the Poincaré–Betti series of R is the rational function $P_R(t) = (1+t)^{b_1}/(1-t^2)^r$.

Based upon Tate's results, Serre stated in his lecture notes [169, p. 118] that it was not known whether or not $P_R(t)$ was always a rational function. This problem remained open for over twenty years, until it was settled negatively in 1982 by David Anick [8]. Anick's example was an Artinian algebra R with $\mathfrak{m}^3 = 0$. Constructing a finite simply-connected CW complex X whose cohomology ring was R, a 1979 result of Roos [158] showed that the Poincaré–Betti series of the loop space ΩX,

$$H(t) = \sum \dim H_i(\Omega X)\, t^i,$$

was not a rational function either. This settled a second problem of Serre, also posed in [169, p. 118].

4.2.3. *Matlis duality.* In his 1958 thesis [137] under Kaplansky, Eben Matlis studied the structure of injective modules over a Noetherian ring R, and showed that they can be written uniquely as direct sums of copies of the injective hulls $E(R/\mathfrak{p})$, as $\mathfrak{p}$ ranges over the prime ideals of R. This put injective resolutions on an equal footing with projective resolutions.

Let $\mathcal{A}$ denote an additive category of modules over a ring R. A *dualizing functor* on $\mathcal{A}$ is an exact contravariant R-linear functor D from $\mathcal{A}$ to itself such that $D(D(M)) = M$. Matlis' thesis [137] also showed that the category $\mathcal{A}$ of modules of finite length over a local Noetherian ring R has a unique dualizing functor: $D(M) = \mathrm{Hom}_R(M, E)$, where E is the injective hull of $R/\mathfrak{m}$.

This turned attention to other kinds of duality, and to modules of finite injective dimension. The goal here was to find the analogue of Serre's Duality Theorem for projective space $X = \mathbb{P}^d$ [164]: if F is a coherent sheaf on X then the dual of the vector space $H^i(X; F)$ is $\mathrm{Ext}_X^{d-i}(F, \omega_X)$, where $\omega_X = \Omega_X^d$ is the sheaf of differential d-forms on X.

It would turn out that that the good class of rings from this perspective would be Gorenstein rings. In a 1957 Séminaire Bourbaki talk on duality [82, exp. 2], Grothendieck defined a commutative ring R (or scheme) of finite type over a field to be "Gorenstein" if it is Cohen–Macaulay and a certain R-module ω_R is locally free of rank 1. A few years later, Bass proved a theorem characterizing rings of finite self-injective dimension [19], and Serre remarked that the two definitions agreed in a geometric context. Bass consolidated these notions in the 1963 paper [20], giving the modern definition: a commutative Noetherian ring R is called *Gorenstein* if all its local rings have finite injective dimension. Bass proved that this is equivalent to several other conditions, such as R being Cohen–Macaulay and a system of parameters generates an irreducible ideal in each local ring.

Nowadays we have the notion of the *canonical module* ω_R of a ring (see below), and if R is a Cohen–Macaulay local ring, then R is Gorenstein if and only if R is its own canonical module: $\omega_R = R$. For example, in Matlis Duality for a zero-dimensional ring, the role of ω_R is played by E, and R is Gorenstein exactly when $E = R$.

4.2.4. *Local cohomology and duality.* In 1961, Grothendieck ran a Harvard seminar on Local Cohomology, based upon his 1957 Séminaire Bourbaki talk on duality [82, exp. 2]; the notes were eventually published in [84]. From the viewpoint of schemes, the local cohomology of a sheaf is the same as cohomology with supports. From the viewpoint of Noetherian local rings, the local cohomology $H^*_{\mathfrak{m}}(M)$ of a module M are the derived functors of the $\mathfrak{m}$-primary submodule functor

$$H^0_{\mathfrak{m}}(M) = \varinjlim \operatorname{Hom}_R(R/\mathfrak{m}^n, M),$$

so

$$H^i_{\mathfrak{m}}(M) = \varinjlim \operatorname{Ext}^i_R(R/\mathfrak{m}^n, M).$$

Grothendieck showed that the depth of M is characterized as the smallest i such that $H^i_{\mathfrak{m}}(M) \neq 0$, and that if R is a Cohen–Macaulay ring then $H^i_{\mathfrak{m}}(R) \neq 0$ only for $i = \dim(R)$. Moreover, R is a Gorenstein ring if and only if the module $H^{\dim(R)}_{\mathfrak{m}}(R)$ is *dualizing* in Matlis' sense, meaning that it is the injective hull of $R/\mathfrak{m}$.

The highlight of the seminar was the Duality Theorem: if R is a complete Gorenstein ring of dimension d, then $H^i_{\mathfrak{m}}(M)$ is dual to $\operatorname{Ext}^{d-i}_R(M, R)$, in the sense that Matlis' dualizing functor D interchanges them. For a more general local ring, the duality is more complicated. If R is complete and Cohen–Macauley, one considers the functors $T^i(M) = D(H^i_{\mathfrak{m}}(M))$, and shows that they equal $\operatorname{Ext}^{d-i}_R(M, \omega_R)$, where $\omega_R = D(H^d_{\mathfrak{m}}(R))$. More generally, Grothendieck also observed that the $T^i(M)$ may be interpreted as $\operatorname{Ext}^{d-i}_R(M, K_R)$ for a suitable dualizing cochain complex K_R on R [84, Section 6.8]. This led to the development of the derived category $D(R)$, which we shall describe shortly.

This material on duality took awhile to absorb, and a ring-theoretic derivation of these results only appeared in 1970 [170]. Gradually the notion of a *canonical module* ω_R became the organizing principal for duality theory, and R is Gorenstein exactly when $\omega_R = R$. If R is Cohen–Macaulay, the canonical module is defined [90] to be a maximal Cohen–Macaulay R-module of finite injective dimension, and the functor $D(M) = \operatorname{Hom}_R(M, \omega_R)$ is dualizing on the category of maximal Cohen–Macaulay R-modules.

In 1971, Sharp [171] used local cohomology (and duality) to show that if R is a complete Cohen–Macauley local ring then the Gorenstein modules are precisely the direct sums of ω_R. He also showed that the final term in the Cousin complex of an R-module M is $H^{\dim(M)}_{\mathfrak{m}}(M)$.

In 1976 Hochster and Roberts [100] studied the local cohomology of a graded ring R in characteristic $p > 0$, and found that the structure of the local cohomology $H^i_{\mathfrak{m}}(R)$ was amazingly simplified under certain assumptions, such as the purity of the Frobenius homomorphism $F : R \to R$. They were also able to lift these characteristic p results to certain rings of characteristic 0, beginning a rennaisance in the study of Cohen–Macaulay rings.

4.3. *Cohomology theories in Algebraic Geometry*

During the early 1950's, the foundations of Algebraic Geometry were reworked by O. Zariski and others, focussing upon the role played by the algebras of regular functions. In his classic 1955 paper "GAGA" [164], Serre observed that if U is affine, with coordinate ring R, then there is an equivalence between finitely generated R-modules and coherent sheaves of modules on U. Hence, restriction to an affine open V of U is an exact functor on coherent modules, because it corresponds to localization of modules. This implies that if F is coherent and U is affine then the Čech cohomology $\check{H}^q(U, F)$ vanishes. Using this, Serre defined the cohomology groups $H^q(X, F)$ of a coherent module on any variety X as the Čech cohomology relative to a covering of X by affine open subvarieties U. All this was in the spirit of the Cartan Seminars on sheaf theory in 1948–1950, but with the homological underpinnings of Cartan–Eilenberg available, Serre's presentation in terms of the Zariski topology was much simpler.

Serre also proved in [165] that if X is a projective variety over $\mathbb{C}$ the groups $H^q(X, F)$ were the same as the analytically defined Betti cohomology, leaving little doubt that using the Zariski topology was a good approach to cohomology.

Grothendieck then observed that Serre's construction was a special case of the derived functor sheaf cohomology (for the Zariski topology) that he had developed in his 1957 paper [81]. Chapter III of "EGA" [83] was devoted to the Zariski cohomology theory of coherent sheaves on a scheme, using the right derived functors Rf_* associated to a morphism $f : X \to Y$.

As part of the preliminaries to this development, Grothendieck wrote a primer on spectral sequences and hypercohomology in 1961 [83, 0_{III}]. This was a reworking of the corresponding material in [41, 81] into a more workable form, and made these tools widely available to algebraic geometers.

4.3.1. *Galois cohomology.*

We have already mentioned that Hochschild [95] coined the term "Galois cohomology" in 1950 for the group cohomology of the Galois groups $G = \mathrm{Gal}(K/k)$, where K is a (possibly infinite) Galois field extension of k. As we have already mentioned, Hochschild [95] and Tate [182, 183] applied Galois cohomology to class field theory in the early 1950's.

In the mid 1950's Tate began to systematically study what he called the "Galois cohomology" of the Galois groups $G = \mathrm{Gal}(K/k)$, where K is a (possibly infinite) Galois field extension of k, such as the separable closure of k. Such a group has a topology induced by its finite quotients:

$$G = \varprojlim G/H_F,$$

where F ranges over all the finite extensions of K contained in k and $H_F = \mathrm{Gal}(K/F)$. As a topological group, G is compact, Hausdorff and totally disconnected; today we call such groups *profinite*. Moreover, each H_F is an open subgroup of finite index in G.

In 1954, Kawada and Tate [114] used Galois cohomology to calculate the cohomology of a variety. To an étale covering U of X they associated a subgroup of the Galois group of $k(U)/k(X)$. This would later be recognized as the first use of what would later be called étale cohomology.

After years of gestation, a published account of Galois cohomology appeared in the 1958 paper [120] by Serge Lang and John Tate. One considers a G-module A which is *discrete* in the sense that the action $G \times A \to A$ is continuous (when A has the discrete topology), and defines the *Galois cohomology* $H^*(G, A)$ to be the cohomology of the complex $C^*(G, A)$ of continuous cochains, that is, maps $\phi : G^n \to A$ which are continuous. An almost immediate observation is that

$$H^*(G, A) = \varinjlim_{H} H^*(G/H, A^H),$$

as H ranges through the open subgroups of finite index in G.

Tate's applications lay in the cases where $\mathbf{A}$ is an Abelian group scheme defined over k; the G-module in this case is $A = \mathbf{A}(\bar{k})$, the group of rational points over the separable closure $\bar{k}$ of k.

One of the most important examples is the group scheme $\mathbf{A} = \mathbf{G}_m$, for which the G-module A is $\bar{k}^\times = \mathbf{G}_m(\bar{k})$ of units of $\bar{k}$. Hilbert's "Theorem 90" states that for every finite Galois extension F/k we have $H^1(\mathrm{Gal}(F/k), F^\times) = 0$; taking the direct limit over all such F and setting $G = \mathrm{Gal}(\bar{k}/k)$ yields the infinite version $H^1(G, \bar{k}^\times) = 0$. As we have seen, it was already known that $H^2(\mathrm{Gal}(F/k), F^\times)$ is the relative Brauer group $\mathrm{Br}(F/k)$; taking the direct limit over all such F shows that $H^2(F, \mathbf{G}_m)$ is the classical Brauer group $\mathrm{Br}(F)$ introduced in 1928 by Richard Brauer [31] and by Brauer and Noether [33].

Serre's 1962 course *Cohomologie Galoisienne* [168], published in 1964, has remained the standard reference on the Galois cohomology over number fields.

4.3.2. *Étale cohomology.* In 1958, Grothendieck found a common generalization of Galois cohomology and Zariski cohomology and used it to define the étale cohomology of schemes. A *Grothendieck topology* is a category $\mathcal{T}$ such that each object X is equipped with a family of morphisms $\{U_i \to X\}$, called *coverings*, subject to certain axioms. From this viewpoint, a *sheaf* F is a contravariant functor on $\mathcal{T}$ such that for each covering, each $s \in F(X)$ is uniquely determined by elements $s_i \in F(U_i)$ which agree in each $F(U_i \times_X U_j)$. The category of sheaves of Abelian groups on $\mathcal{T}$ is an Abelian category with enough injectives, and Grothendieck defined the cohomology groups $H^*(\mathcal{T}, F)$ to be the right derived functors of $F \mapsto F(X)$. When X is a topological space and $\mathcal{T}$ is the poset of open subspaces then sheaf has its usual meaning, and we recover the usual sheaf cohomology on X.

To define the *étale topology* on a scheme X, Grothendieck took the category of all schemes U which are étale over X, with the set-theoretic notion of covering. If F is a sheaf for this topology, the above construction defines the étale cohomology groups $H^*(X_{\mathrm{et}}, F)$ of F on X. When X is the spectrum of a field k and $G = \mathrm{Gal}(\bar{k}/k)$, a discrete G-module A is the same as an étale sheaf on X, so the étale cohomology of X with coefficients A agrees with Tate's Galois cohomology $H^*(G, A)$.

In Fall 1961, Grothendieck presented his ideas in a course at Harvard. The following semester (Spring 1962), M. Artin ran a seminar covering Grothendieck Topologies, as well as some material on étale cohomology (such as cohomological dimension). The published notes [9] of this seminar, as well as Giraud's 1963 Bourbaki talk [77] made the ideas available to a wide audience.

The next year (1962/1963), when the seminar continued in France, Artin and Grothendieck worked out the fundamental structure theorems of étale cohomology: proper and smooth base change, specialization, cohomology with compact supports and duality. The following year, more results were obtained (such as purity and the Lefschetz trace formula), with the seminar notes eventually appearing in 1972 as [10].

One of Grothendieck's early successes with étale cohomology was his cohomological proof of the rationality of the Zeta function $Z_X(t)$ of a scheme of finite type over the finite field $\mathbb{F}_q$. He proved that each factor $P_i(t)$ of $Z_X(t)$ is the characteristic polynomial of the Frobenius operator acting on an l-adic cohomology group, namely

$$H^i(X, \mathbb{Q}_l) = \varinjlim H^i_{\mathrm{et}}\big(X, \mathbb{Z}/(l^\nu)\big).$$

In 1972, Deligne used étale cohomology to prove the "Riemann hypothesis" over $\mathbb{F}_q$ [49]: the eigenvalues of the Frobenius on $H^i(X, \mathbb{Q}_l)$ (and hence the zeroes and poles of the zeta function) were algebraic integers with absolute value $q^{i/2}$. This completed the proof of the celebrated Weil conjectures, and firmly established the importance of étale cohomology.

4.4. *Derived categories*

After Grothendieck's 1961 Harvard seminar on Local Cohomology, described above, Grothendieck realized that in order to extend these results to arbitrary schemes he needed some results in homological algebra which were not yet available. This was overcome by Verdier's 1963 thesis [187] on Derived Categories.

The derived category $D(\mathcal{A})$ of an Abelian category $\mathcal{A}$ is the category obtained from the category $\mathbf{Ch}(\mathcal{A})$ of (co)chain complexes by formally inverting the quasi-isomorphisms, i.e. the maps $C \to C'$ which induce isomorphisms on (co)homology. To describe it, Verdier introduced the notion of a *triangulated category*. The quotient category $K(\mathcal{A})$ of $\mathbf{Ch}(\mathcal{A})$, whose morphisms are the chain homotopy equivalence classes of maps, is triangulated; $D(\mathcal{A})$, which is formed from $K(\mathcal{A})$ by a calculus of fractions, is also triangulated. If $F: \mathcal{A} \to \mathcal{B}$ is an additive functor then under reasonable conditions there is a functor $\mathbf{R}F : D(\mathcal{A}) \to D(\mathcal{B})$ with the property that if an A in $\mathcal{A}$ is considered as a complex then the cohomology of the complex $\mathbf{R}F(A)$ give the ordinary right derived functors $R^*F(A)$.

The topologist D. Puppe had already (1962) defined the notion of a *stable category* in [150]. This is just a graded triangulated category without the "octahedral" axiom. Since Puppe only discussed $K(\mathcal{A})$ and not $D(\mathcal{A})$, and did not deal with the total derived functors $\mathbf{R}F$, his notion never caught the attention of the algebraists.

In the Summer of 1963, after Hartshorne proposed to run a seminar at Harvard on duality theory, Grothendieck wrote a series of "prenotes", sketching the construction of a functor $f^! : D(Y\text{-}\mathbf{mod}) \to D(X\text{-}\mathbf{mod})$ associated to a reasonable morphism $f: X \to Y$ of schemes, together with a natural trace morphism $\mathbf{R}f_* f^!(A) \to A$. The so-called "Séminaire Hartshorne" was held at Harvard in 1963/1964, based upon these prenotes, and the seminar notes appeared as [88]. An appendix to [88], written by Deligne in 1966, constructs $f^!$ for every separated morphism of finite type between Noetherian schemes.

During the 1966/1967 Séminaire de Geometrie Algebrique [24], Grothendieck used the triangulated category $\mathrm{Perf}(X)$ of *perfect* complexes of $\mathcal{O}_X$-modules to develop a global theory of intersections and a Riemann–Roch theorem for arbitrary Noetherian schemes. By definition, a complex is perfect if it is locally quasi-isomorphic to a bounded complex of vector bundles, and the alternating sum of these vector bundles gives a well-defined element in the Grothendieck group $K(X)$, at least if X is quasi-projective or smooth. If $f: X \to Y$ is proper, the machinery of triangulated categories yields an exact functor $Rf_*: \mathrm{Perf}(X) \to \mathrm{Perf}(Y)$ and, hence, a homomorphism $K(X) \to K(Y)$.

In 1978, Bernstein, I. Gelfand and S. Gelfand [23] used derived categories to classify vector bundles on projective space $\mathbb{P}^n$ over a field k in terms of graded modules over the exterior algebra Λ on $n+1$ variables. The crucial step in their classification was the discovery of an isomorphism between the (bounded) derived categories of graded modules $D^b_{\mathrm{gr}}(\Lambda)$ and $D^b_{\mathrm{gr}}(R)$, where R is the polynomial algebra on $n+1$ variables. This result showed that $D^b(\mathcal{A})$ did not determine the "heart" category $\mathcal{A}$, a result which came as a bit of a surprise.

The problem of multiple hearts for a triangulated category was revisited in 1982 by Bernstein, Beilinson and Deligne [21]. These authors used triangulated categories to study $\mathcal{D}$-modules and perverse sheaves on a stratified space. In 1988, Beilinson, Ginsburg and Schechtman [22] generalized the results of [23, 21] by proving that many filtered triangulated categories have two hearts, which are in Koszul duality.

In the mid of 1980's, derived categories found yet another application. The notion of a tilting module had come up in the study of representations of finite algebras. Cline, Parshall and Scott [46] showed in 1986 that if T is a tilting module for A, and $B = \mathrm{Hom}_A(T, T)$, then $D^b(A) \cong D^b(B)$.

Early work on derived categories was often restricted to either bounded or bounded below complexes, because of the need to work with injective (or projective) resolutions. In 1988, Spaltenstein [174] showed that every unbounded complex was quasi-isomorphic to a "fibrant" complex, and that one could use fibrant complexes to compute derived functors. This result has led to several new developments which continue to this day.

5. Simplicial methods

During the 1940's, Eilenberg kept encountering things called "abstract complexes" which resembled the triangulated polyhedra (or "geometric simplicial complexes") introduced by Poincaré, except that a simplex was not always determined by its faces. For example, the abstract complex $K(\pi)$ of [60] and the singular complex $S(X)$ of [57] had this property. To describe this phenomenon, Eilenberg and Zilber [68] introduced the notions of a *semi-simplicial* complex and a *complete semi-simplicial* complex in 1950. The Eilenberg–Zilber notion of a complete semi-simplicial complex is identical to our modern notion of a *simplicial set* K: it is a sequence $K_0, K_1, \ldots$ of sets together with face maps $\partial_i : K_q \to K_{q-1}$ and degeneracy maps $s_i : K_q \to K_{q+1}$ $(0 \leqslant i \leqslant q)$ satisfying certain axioms; a semi-simplicial complex is just a simplicial set without the degeneracy maps.

A word about changing terminology is in order. The term "complete semi-simplicial complex" was awkward and was quickly abbreviated to "c.s.s. complex". During the 1950's the term c.s.s. complex prevailed, although the short-lived term "*FD*-complex" was also used in [63, 52]. Largely due to the influence of John Moore, the adjective "complete" began to be omitted, starting with 1954, while the notion of "semi-simplicial complex" languished in obscurity. By the early 1960's the term "semi-simplicial set" had replaced "c.s.s. complex". By the late 1960's, even the prefix "semi" was dropped, influenced by May's 1967 book [138]; since then "simplicial set" has been the universally used term.

Returning to the early 1950's, we mention two results which showed the power of the new simplicial methods. The "Eilenberg–Zilber theorem" was proven in 1953 [69] as an application of c.s.s. complexes to products: the (simplicial) map $S(X \times Y) \simeq S(X) \otimes S(Y)$, implicitly defined by Alexander and Whitney in 1935, is a homotopy equivalence. In 1955, the homotopy theory of c.s.s. complexes satisfying an extension condition was developed by Daniel Kan [109]; a simplicial set satisfying Kan's extension condition is now called a *Kan complex*.

5.1. *Homotopical algebra*

The homological study of simplicial Abelian groups was launched in 1954 by Eilenberg and Mac Lane [63], as part of their algebraic program to find the cohomology of Eilenberg–Mac Lane spaces $K(\pi, n)$. This program was analyzed with typical thoroughness in the 1954/1955 Seminaire Cartan [40]. In exposés 18 and 19 of that seminar, John Moore showed that every simplicial group K is a Kan complex, and that one could compute its homotopy groups as the homology of a chain complex N_* of groups, where $N_q \subset K_q$ is the intersection of kernels of all the face maps except ∂_q. The complex N_* quickly became known as the *Moore complex* of K.

In 1956/1957, A. Dold [52] and D. Kan [110] independently discovered that the Moore complex provided an equivalence between the category of simplicial Abelian groups and the category of non-negative chain complexes of Abelian groups. This Dold–Kan correspondence was later codified in [53]. Under the correspondence, Moore's result states that simplicial homotopy corresponds to homology. With this correspondence at hand, simplicial techniques could be brought to bear on any homological problem.

Dold and Puppe [53] announced in 1958 that with simplicial methods one could define the derived functors of a non-additive functor T (say of modules); their detailed paper appeared in 1961. The key idea was that one could consider a projective resolution P_* of a module M as a simplicial module via the Dold–Kan correspondence. Since the notion of simplicial homotopy does not involve addition, we may take the homotopy groups of $T(P_*)$ as the derived functors $L_iT(M)$ of T. A variant is obtained by placing M in degree $n > 0$; the derived functors $L_iT(M, n)$ are the homotopy groups of $T(P[n])$, where the simplicial module $P[n]$ corresponds to the chain complex P_* shifted n places. For example, the i-th homology $H_i(K(\pi, n); \mathbb{Z})$ of an Eilenberg–Mac Lane space $K(\pi, n)$ is just $L_iT(\pi, n)$ for the group ring functor $T(\pi) = \mathbb{Z}[\pi]$.

It is possible to generalize the Dold–Puppe construction and define the left derived functors of any functor T from any category $\mathcal{C}$ to an Abelian category, as long as $\mathcal{C}$ is closed under finite limits and has enough projective objects. This observation evolved during the late 1960's, finding voice in M. André's 1967 book [5], Quillen's 1967 book [151] on homotopical algebra, and in the later papers [6, 152]. In fact there are three standard constructions, which agree in reasonable situations.

André's construction [5] uses a subcategory of "acyclic models" in $\mathcal{C}$. In the category of functors on $\mathcal{C}$, one finds a resolution $T_* \to T$ which is aspherical on the "model" objects. Then one defines $L_iT(A)$ to be $\pi_iT_*(A)$, or H_i of the chain complex associated to the simplicial module $T_*(A)$.

Quillen's construction is simpler: one finds a simplicial "resolution" $P_* \to A$ of each A in $\mathcal{C}$, and defines $L_iT(A)$ to be $H_iT(P_*)$. The work comes in deciding what a "resolution" is: P_* should be *cofibrant* and $P_* \to A$ should be an *acyclic fibration* in the terminology of [151]. In many algebraic applications, fibrations are defined by a relative lifting property, so all "relatively projective" objects are cofibrant.

During 1965–1969, Barr and Beck [18] developed the idea of cotriple resolutions as a functorial way to obtain resolutions for computing non-Abelian derived functors. Suppose that there is a forgetful functor $U : \mathcal{C} \to \mathcal{S}$ with a left adjoint F. Then the functor FU is called a *cotriple*, and the iterates $P_i = (FU)^{i+1}(A)$ often form a simplicial "resolution" $P_* \to A$. Again, one takes $L_i T(A) = H_i T(P_*)$.

5.1.1. *Cohomology of commutative rings.* In analogy with Hochschild's (co)homology theory for associative algebras, it is reasonable to ask for a (co)homology theory for commutative rings. Let $k \to A$ be a map of commutative rings, and M an A-module. Then Hochschild's group $H^1(A; M)$ is the A-module $\mathrm{Der}_k(A, M)$ of all derivations $A \to M$ which vanish on k (as there are no inner derivations), $H_1(A; M)$ is $M \otimes \Omega_{A/k}$ and $H^2(A; M)$ classifies all associative k-algebra extensions B of A by M which are *k-split*, meaning that $B \cong A \oplus M$ as a k-module (this condition is obvious when k is a field). What was wanted was a theory with the same H^1 and H_1, but such that H^2 was the group $\mathrm{Exalcomm}_k(A, M)$ classifying all commutative k-algebra extensions of A by M.

The functors H^1 and H^2 were first studied by P. Cartier [42] in 1956, in the case that $A = K$ is a field extension of k, and partially extended to commutative rings in 1961 by Nakai [144]. In a 1961 course at Harvard, Grothendieck defined $\mathrm{Exalcomm}_k(A, M)$ and constructed a 6-term cohomology sequence for $k \to A \to B$ [83, 0_{IV}(18.4.2)].

When k is a field, Harrison [87] used a subcomplex of the Hochschild complex (1962) to define k-modules $H^*_{\mathrm{harr}}(A, M)$ with $H^1_{\mathrm{harr}} = H^1$ and $H^2_{\mathrm{harr}} = \mathrm{Exalcomm}_k$, equipped with a 9-term cohomology sequence. When k is perfect, and A is the local ring (at some point) of a variety over k, Harrison proved the following two results: (1) A is regular if and only if $H^2_{\mathrm{harr}}(A, -) = 0$, and (2) A is a complete intersection if and only if $\dim H^1_{\mathrm{harr}}(A, A/\mathfrak{m}) - \dim H^2_{\mathrm{harr}}(A, A/\mathfrak{m}) = \dim A$.

The next step was taken in the 1964 paper [125] by two Ph.D. students of Tate, Lichtenbaum and Schlessinger. Let k be any commutative ring. For each commutative ring map $f : k \to A$, they defined a 3-term chain complex $\mathbf{L}^\cdot$, called the *cotangent complex* of f, and – for $i = 0, 1, 2$ – set $T_i(A/k, M) = H_i(\mathbf{L}^\cdot \otimes M)$, $T^i(A/k, M) = H^i \mathrm{Hom}(\mathbf{L}^\cdot, M)$. When k is a field the $T^i(A/k, M)$ agreed with Harrison's $H^{i+1}_{\mathrm{harr}}(A, M)$, and in general $T^1(A/k, M) = \mathrm{Exalcomm}_k(A, M)$. Their infinitesimal criterion for A/k to be smooth, in terms of the vanishing of $T^1(A/k)$, was used by Grothendieck (1967) to great advantage in EGA [83, IV.17]. If k is Noetherian, R is a localization of $k[x, \ldots, y]$ and $A = R/I$, they showed that $T^2(A/k, -) = 0$ if and only if A is a complete intersection, i.e. I is defined by a regular sequence in R. Schlessinger's thesis applied the T^i to deformation theory, while Lichtenbaum's thesis was concerned with applications to relative intersection theory.

In 1967, M. André [5–7] and Quillen [152] discovered what we now call André–Quillen cohomology. If $k \to A$ and M are as above, their groups $D^i(A/k, M)$ agree with the Lichtenbaum–Schlessinger groups $T^i(A/k, M)$ for $i = 0, 1, 2$. It comes with a long exact sequence for $k \to A \to B$ (generalizing Harrison's) and generalizations of the Lichtenbaum–Schlessinger results for smoothness and local complete intersections. In this theory, the central role is played by a simplicial A-module $\mathbb{L}_{A/k}$, called the *cotangent complex* of A relative to k, because of the similarity (using the Dold–Kan correspondence) to the Lichtenbaum–Schlessinger complex $\mathbf{L}^{\bullet}$. This complex is well-defined in the derived category of chain complexes of A-modules, and one has

$$D^i(A/k, M) = H^i \mathrm{Hom}_A(\mathbb{L}_{A/k}, M)$$

and

$$D_i(A/k, M) = H_i(\mathbb{L}_{A/k} \otimes_A M).$$

Formally, the $D^i(A/k, M)$ are the non-Abelian derived functors of the functor

$$T(B) = \mathrm{Der}_k(B, M) \cong \mathrm{Hom}_A(A \otimes_B \Omega_{B/k})$$

on the category $\mathcal{C}$ of commutative k-algebras over A. According to the above prescription, the definition starts with an acyclic simplicial resolution $P_* \to A$ in $\mathcal{C}$, and has

$$D^i(A/k, M) = H^i \mathrm{Der}_k(P_*, M).$$

Defining the simplicial A-module $\mathbb{L}_{A/k} = A \otimes_P \Omega_{P/k}$, a little algebra yields the above formulas.

5.1.2. *Higher algebraic K-theory.* In order to find a possible definition of the higher K-groups $K_n(R)$ of a ring R, Swan was led in 1968 to consider the non-Abelian derived functors of the general linear group GL on the category of rings [180]. This required a slight generalization of derived functor, since the category of groups is not an Abelian category. In this context we have a functor G from a category $\mathcal{C}$, such as the category of rings, to the category of groups or sets.

Swan's original construction followed André's method, finding an acyclic resolution $G_* \to GL$ in the functor category and setting $K_n(R) = \pi_{n-2}G_*(R)$ for $n \geqslant 2$. In 1969 Gersten gave a cotriple construction [75], using the cotriple associated to the forgetful functor from rings to sets, while both Keune [116] and Swan [181] gave constructions using free resolutions $P_* \to R$ to define $K_n(R) = \pi_{n-2}GL(P_*)$ for $n \geqslant 2$. By 1970, Swan had proven [181] that all three constructions yielded the same functors $K_n(R)$.

Historically, however, the important construction was given by Quillen in 1969 [153]. He showed how to modify the classifying space $BGL(R)$ of $GL(R)$ to obtain a topological space $BGL(R)^+$ with the same homology as $BGL(R)$, and defined $K_n(R) = \pi_n BGL(R)^+$ for $n \geqslant 1$. The equivalence of Quillen's topological definition with the homological Swan–Gersten definition was established in 1972 by combining partial results obtained by several authors [4]. Since then the field of higher algebraic K-theory has taken on a life of its own, but that is another story.

5.2. *Hochschild and cyclic homology*

We have already described the 1945 development [94] of Hochschild homology of an algebra A over a field k. The next step was to let A be an algebra over an arbitrary commutative base ring k. In his 1956 paper [96], Hochschild began a systematic study of exact sequences of R-modules which are k-split (split as sequences of k-modules). This became part of a "relative" homological algebra movement.

Hochschild, Kostant and Rosenberg showed in 1962 [97] that if A is smooth of finite type over a field k, then there is a natural isomorphism $\Omega^*_{A/k} \cong H_*(A, A)$. It follows that for such A there is an analogue $d : \Omega^n_A \to \Omega^{n+1}_A$ of de Rham's operator for manifolds. In 1963, Rinehart [156] mimicked this construction for all algebras, constructing a chain map B inducing an operator $HH_n(A, A) \to H_{n+1}(A, A)$. This attempt to define an analogue of de Rham cohomology was before its time: twenty years later, Alain Connes [48] as well as Feigin and Tsygan [185, 70] would both seize upon B and make it the foundation of cyclic homology, unaware of Rinehart's earlier work.

We end our quick tour by mentioning an important application, discovered by Gerstenhaber in the 1964 paper [76]. A *deformation* of an associative algebra A is a $k[[t]]$-algebra structure on the $k[[t]]$-module $A[[t]]$ whose product agrees modulo t with the given product on A. Reducing a deformation modulo t^2 yields a k-split algebra extension of A by A, so giving the "infinitesimal" part of the deformation is equivalent to giving an element of $H^2(A, A)$. Gerstenhaber showed that there is a whole sequence of obstructions to deformations of A, lying in the Hochschild cohomology group $H^3(A, A)$. If A is smooth of finite type, the Hochschild–Kostant–Rosenberg theorem implies that the obstructions belong to $\Omega^3_{A/k}$.

5.2.1. Cotor *for coalgebras.* Hochschild homology was also involved in the early development of (differential graded) coalgebras over a field. This field was heavily influenced by its applications to topology, in part because the homology of a topological space X is a graded coalgebra, via the diagonal map $H_*(X) \to H_*(X \times X) \cong H_*(X) \otimes H_*(X)$. Moreover, the normalized chain complex $C_*(X)$ is a differential graded coalgebra.

In 1956, J.F. Adams [1] discovered a recipe for the homology of the loop space ΩX when X is simply connected. To describe it, he considered $C_*(X)$ as a differential graded coalgebra. Mimicking the Eilenberg–Mac Lane bar construction, Adams defined a differential graded algebra F_*, called the *cobar construction*, and showed that $H_*(\Omega X) \cong H_*(F_*)$. This purely algebraic construction attracted the attention of topologists to the algebraic structure of coalgebras and their comodules.

Now if C is a coalgebra one can define the *cotensor product* $M \square_C N$ of comodules M and N. Its right derived functors are called the *cotorsion* products $\mathrm{Cotor}^C(M, N)$ of M and N. In 1966 [65], Eilenberg and Moore defined and studied the cotensor product over a DG coalgebra $C = C_*$. Under mild flatness hypotheses, they constructed what we now call the "Eilenberg–Moore spectral sequence", which has E^2 equal to $\mathrm{Cotor}^{HC}_{pq}(H(M), H(N))$ and converges to $\mathrm{Cotor}^C(M, N)$. The importance of this is illustrated by the case when C is the normalized chain complex of a simply connected topological space X, and M and N are the chain complexes of spaces E and X' over X. If $E \to X$ is a Serre fibration, they prove that $\mathrm{Cotor}^C(M, N)$ is the homology of the fiber space $E' = E \times_X X'$, so this provides a powerful method to calculate homology. Of course when X' and E are contractible then $E' \simeq \Omega X$, and they recover Adams' cobar construction.

Eilenberg and Moore also studied the dual construction for tensor products of differential graded modules M, N over a differential graded algebra R. In this case the spectral sequence is

$$E_2^{pq} = \mathrm{Tor}_{H(R)}\big(H(N), H(M)\big) \Rightarrow \mathrm{Tor}_R(N, M).$$

Using the cochain algebras in the above topological situation, Eilenberg and Moore proved that

$$H^*(E') \cong \mathrm{Tor}_{C^*(X)}\big(C^*(E), C^*(X')\big),$$

so the spectral sequence converges to $H^*(E')$. This spectral sequence was described and studied in [173] by Larry Smith, who showed that this spectral sequence often collapsed.

Here is one application. Suppose that Y is simply connected and we take $X = Y \times Y$, with X' the diagonal copy of Y, and E the path space of Y. Then $E' = \Omega Y$ and if $C^*(Y)$ takes coefficients in a field k the Künneth formula yields

$$C^*(X) \simeq C^*(Y) \otimes C^*(Y).$$

Since the Eilenberg–Moore spectral sequence collapses in this case it yields an isomorphism between $H^*(\Omega Y)$, and the Hochschild cohomology $HH^*(C^*(Y), k)$ of the differential graded algebra $C^*(Y)$.

5.2.2. *Mac Lane cohomology and topological Hochschild homology* Let A be an associative ring and M an A-bimodule. As we have mentioned above, the Hochschild cohomology group $H^2(A, M)$ only measures ring extensions of A by M whose underlying Abelian group is $A \oplus M$. (One takes k to be $\mathbb{Z}$.) In order to measure *all* ring extensions of A by M, Mac Lane introduced what we now call *Mac Lane cohomology* in the 1956 paper [131]. One may naturally define a differential graded ring $Q = Q_*(A)$ and an augmentation $Q \to A$. By definition, $HML_*(A, M)$ and $HML^*(A, M)$ are the Hochschild homology $H_*(Q, M)$ and cohomology $H^*(Q, M)$. As required, ring extensions correspond to elements of the group $HML^2(A, M)$.

A variant for k-algebras and their extensions was invented in 1961 by U. Shukla [172], and is called *Shukla homology*. Shukla proved two comparison results: when k is a field, Shukla homology recovers Hochschild homology; when $k = \mathbb{Z}$, Shukla homology agrees with a homology theory defined by Mac Lane in 1958 (which is not Mac Lane homology, as asserted by Shukla).

Both Mac Lane cohomology and Shukla homology were almost completely forgotten for thirty years, except for some calculations by Breen in [34]. In 1991, an innocuous paper by Jibladze and Pirashvili [108] proved that the Mac Lane homology of a ring A (and a module M) is $\mathrm{Tor}_*^{\mathcal{F}}(A\otimes, M\otimes)$ in the functor category $\mathcal{F} = \mathcal{F}(A)$ of functors from the category of fin. gen. free A-modules to the category of A-modules. Similarly, the Mac Lane cohomology of A is $\mathrm{Ext}_{\mathcal{F}}(A\otimes, M\otimes)$. This was to lead to an unexpected connection to algebraic K-theory and manifolds.

In the late 1970's, F. Waldhausen introduced a variant of algebraic K-theory, which he called *stable K-theory* [188]. His construction was designed to study the homotopy theory of the diffeomorphism group of a manifold, and could be applied to a ring spectrum A as well as ordinary rings. Following this lead in the early 1980's, M. Bökstedt [27] introduced a variant $THH_*(A, A)$ of Hochschild homology for ring spectra, called *Topological Hochschild Homology*. It is roughly obtained by replacing rings by ring spectra and tensor products over k by smash products. In 1987, Waldhausen announced that stable K-theory of A was isomorphic to $THH(A)$, but the proof [54] took several years to appear.

Then in 1992, Pirashvili and Waldhausen [147] used the functor category interpretation to prove that the Mac Lane homology group $HML(A, A)$ was the same as $THH(A)$. This showed that homological algebra could be applied to calculate the topological invariants of Waldhausen and Bökstedt. A new and active field of research has been born out of this discovery.

5.3. *Cyclic homology*

Cyclic homology arose simultaneously in several applications in the early 1980's.

While studying applications of C^*-algebras to differential geometry in 1981, Alain Connes was led to study Hochschild cochains which were invariant under cyclic permutations of its arguments [47, 48]. Realizing that such "cyclic" cochains were preserved by the Hochschild coboundary gave him a new cohomology theory, rapidly christened $HC^*(A)$ and called the *cyclic cohomology* of A. Meanwhile, Boris Tsygan [185] was studying the homology of the Lie algebra $\mathfrak{gl}(A)$ over a field k of characteristic zero, and discovered that the Hopf algebra $H_*(\mathfrak{gl}(A); k)$ was the tensor algebra on the homology groups $K_i^+(A)$ of the complex of all Hochschild chains invariant under cyclic permutation; the proof, and the cohomology version, appeared in the 1983 paper [70] by Feigin and Tsygan. This description of $H_*(\mathfrak{gl}(A); k)$ was discovered independently in 1983 by Loday and Quillen [127], and their paper made the new subject of cyclic homology accessible to a large audience.

Both Connes and Tsygan discovered the following key structural sequence relating cyclic homology to Hochschild homology; Rinehart's operator [156] is the composition BI

$$\cdots HC_{n+1}(A) \xrightarrow{S} HC_{n-1}(A) \xrightarrow{B} H_n(A, A) \xrightarrow{I} HC_n(A) \cdots .$$

Using this sequence, Connes and others rediscovered and clarified the connection with de Rham cohomology; for smooth algebras $HC_n(A)$ is a product of de Rham cohomology groups, together with $\Omega^n_{A/k}/\mathrm{d}\Omega^{n-1}_{A/k}$.

In retrospect, cyclic homology had been hinted at in several places in the late 1970's: pseudo-isotopy theory [55], the homology of S^1-spaces and in algebraic K-theory [126]. Other applications soon arose. For example, Goodwillie showed in [80] that the cyclic homology (over $\mathbb{Q}$) of a nilpotent ideal I is isomorphic to the algebraic K-theory of I. Because of its diverse applications to other areas of mathematics, cyclic homology became quickly established as a flourishing field in its own right.

It is impossible to give an accurate historical perspective on current developments. As tempting as it is, I shall refrain from doing so. Perhaps in fifty years the history of homological algebra will be unrecognizable to us today. Let us hope so!

Bibliography

[1] J.F. Adams, *On the cobar construction for coalgebras*, Proc. Nat. Acad. Sci. **42** (1956), 409–412.

[2] J.W. Alexander, *A proof of the invariance of certain constants in Analysis Situs*, Trans. Amer. Math. Soc. **16** (1915), 148–154.

[3] J.W. Alexander, *Combinatorial analysis situs*, Trans. Amer. Math. Soc. **28** (1926), 301–329.

[4] D. Anderson, *Relationship among K-theories*, Lecture Notes in Math. vol. 341, Springer, Berlin (1973), 57–72.

[5] M. André, *Méthode Simplicial en Algèbre Homologique et Algèbre Commutative*, Lecture Notes in Math. vol. 32, Springer (1967).

[6] M. André, *Homology of simplicial objects*, Proc. Sympos. Pure Math. vol. 17, Amer. Math. Soc., Providence, RI (1970), 15–36.
[7] M. André, *Homologie des Algèbres Commutatives*, Springer, Berlin (1974).
[8] D. Anick, *A counterexample to a conjecture of Serre*, Ann. of Math. **115** (1982), 1–33.
[9] M. Artin, *Grothendieck Topologies*, Mimeographed Seminar Notes, Harvard University (1962).
[10] M. Artin, A. Grothendieck and J.-L. Verdier, *Théorie des Topos et Cohomologie Etale des Schémas (SGA4)*, Lecture Notes in Math. vols 269, 270, 305, Springer, Berlin (1972/1973).
[11] E. Artin, C. Nesbitt and R. Thrall, *Rings with Minimum Condition*, Univ. Michigan Press (1944).
[12] M. Auslander, *On the dimension of modules and algebras III. Global dimension*, Nagoya Math. J. **9** (1955), 67–77.
[13] M. Auslander and D. Buchsbaum, *Homological dimension in local rings*, Trans. Amer. Math. Soc. **85** (1957), 390–405; *Homological dimension in Noetherian rings*, Proc. Nat. Acad. Sci. **42** (1956), 36–38.
[14] M. Auslander and D. Buchsbaum, *Codimension and multiplicity*, Ann. of Math. **68** (1958), 625–657.
[15] M. Auslander and D. Buchsbaum, *Unique factorization in regular local rings*, Trans. Amer. Math. Soc. **45** (1959), 733–734.
[16] R. Baer, *Erweiterungen von Gruppen und ihren Isomorphismen*, Math. Z. **38** (1934), 375–416.
[17] R. Baer, *Abelian groups that are direct summands of every containing Abelian group*, Bull Amer. Math. Soc. (1940), 800–806.
[18] M. Barr and J. Beck, *Homology and Standard Constructions*, Lecture Notes in Math. vol. 80, Springer, Berlin (1969).
[19] H. Bass, *Projective modules over algebras*, Ann. of Math. **73** (1961), 532–542.
[20] H. Bass, *On the ubiquity of Gorenstein rings*, Math. Z. **82** (1963), 8–28.
[21] A. Beilinson, J. Bernstein and P. Deligne, *Faisceaux pervers*, Astérisque **100** (1982).
[22] A. Beilinson, V. Ginsburg and V. Schechtman, *Koszul duality*, J. Geom. Phys. **5** (1988), 317–350.
[23] I. Bernšteĭn, I. Gelfand and S. Gelfand, *Algebraic vector bundles on $\mathbb{P}^n$ and problems of linear algebra*, Funkt. Anal. Prilozhen. **12** (1978), 66–67.
[24] P. Berthelot, A. Grothendieck and L. Illusie, *Théorie des Intersections et Théorème de Riemann–Roch (SGA6)*, Lecture Notes in Math. vol. 225, Springer, Berlin (1971).
[25] E. Betti, *Sopra gli spazi di un numero qualunque di dimensioni*, Ann. Mat. Pura Appl. **2/4** (1871), 140–158.
[26] J.M. Boardman, *Conditionally convergent spectral sequences*, Preprint (1981).
[27] M. Bökstedt, *Topological Hochschild homology*, Preprint (1985), unpublished.
[28] M. Bollinger, *Geschichtliche Entwicklung des Homologiebegriffs*, Arch. Hist. Exact Sci. **9** (1972), 94–170.
[29] A. Borel and J.-P. Serre, *Le théorème de Riemann–Roch*, Bull. Soc. Math. France **86** (1958), 97–136.
[30] N. Bourbaki, *Éléments de Mathématique, Part I, Livre II, Algèbre ch. 2*, Hermann, Paris (1942).
[31] R. Brauer, *Untersuchungen über die arithmetischen Eigenschaften von Gruppen linearer Substitutionen. I*, Math. Z. **28** (1928), 677–696.
[32] R. Brauer, *Sur les invariants intégraux des variétés répresentatives des groupes de Lie simples clos*, C. R. Acad. Sci. Paris **201** (1935), 419–421.
[33] R. Brauer and E. Noether, *Über minimale Zerfällungskörper irreduzibler Darstellungen*, Sitzungsberichte Preußischen Akad. Wissen. (1927), 221–228.
[34] L. Breen, *Extensions du groupe additif*, Publ. I.H.E.S. **48** (1978), 39–125.
[35] D. Buchsbaum, *Exact categories and duality*, Trans. Amer. Math. Soc. **80** (1955), 1–34.
[36] E. Cartan, *Sur les nombres de Betti des espaces de groupes clos*, C. R. Acad. Sci. Paris **187** (1928), 196–198.
[37] E. Cartan, *Sur les invariants intégraux de certains espaces homogènes clos et les propriétés topologique de ces espaces*, Ann. Soc. Polon. Math. **8** (1929), 181–225.
[38] H. Cartan, *Séminaire Henri Cartan de Topologie Algébrique 1948/1949*, Secrétariat Mathématique, Paris (1955).
[39] H. Cartan, *Séminaire Henri Cartan de Topologie Algébrique 1950/1951, Cohomologie des Groupes, Suite Spectrale, Faisceaux*, Secrétariat Mathématique, Paris (1955).
[40] H. Cartan, *Séminaire Henri Cartan 1954/1955, Algèbre de Eilenberg–Mac Lane et Homotopie*, Secrétariat Mathématique, Paris (1955).
[41] H. Cartan and S. Eilenberg, *Homological Algebra*, Princeton Univ. Press, Princeton, NJ (1956).
[42] P. Cartier, *Dérivations dans les corps*, Exposé 13, Seminaire Cartan et Chevalley 1955/1956, Secrétariat Mathématique, Paris (1956).
[43] E. Čech, *Théorie générale de l'homologie dans un espace quelconque*, Fund. Math. **19** (1932), 149–183.

[44] E. Čech, *Les groupes de Betti d'un complex infini*, Fund. Math. **25** (1935), 33–44.
[45] C. Chevalley and S. Eilenberg, *Cohomology theory of Lie groups and Lie algebras*, Trans. Amer. Math. Soc. **63** (1948), 85–124.
[46] E. Cline, B. Parshall and L. Scott, *Derived categories and Morita theory*, J. Algebra **104** (1986), 397–409.
[47] A. Connes, *Cohomologie cyclique et foncteurs* Ext^n, C. R. Acad. Sci. Paris **296** (1983), 953–958.
[48] A. Connes, *Noncommutative differential geometry*, Publ. I.H.E.S. **62** (1985), 257–360.
[49] P. Deligne, *La conjecture de Weil I*, Publ. Math. I.H.E.S. **43** (1974), 273–307.
[50] G. de Rham, *Sur l'analysis situs des variétés à n dimensions*, J. Math. Pures Appl. **10** (1931), 115–200; C. R. Acad. Sci. Paris **188** (1929), 1651–1652.
[51] J. Dieudonné, *A History of Algebraic and Differential Topology 1900–1960*, Birkhäuser, Basel (1989).
[52] A. Dold, *Homology of symmetric products and other functors of complexes*, Ann. of Math. **68** (1958), 54–80.
[53] A. Dold and D. Puppe, *Homologie nicht-additiver Funktoren; Anwendungen*, Ann. Inst. Fourier (Grenoble) **11** (1961), 201–312; *Non-additive functors, their derived functors, and the suspension homomorphism*, Proc. Nat. Acad. Sci. **44** (1958), 1065–1068.
[54] B. Dundas and R. McCarthy, *Stable K-theory and topological Hochschild homology*, Ann. of Math. **140** (1994), 685–701.
[55] W. Dwyer, W.-C. Hsiang and R. Staffeldt, *Pseudo-isotopy and invariant theory*, Topology **19** (1980), 367–385.
[56] B. Eckmann, *Der Cohomologie-Ring einer beliebigen Gruppe*, Comment. Math. Helv. **18** (1945–1946), 232–282.
[57] S. Eilenberg, *Singular homology theory*, Ann. of Math. **45** (1944), 407–447.
[58] S. Eilenberg, *Topological methods in abstract algebra*, Bull. Amer. Math. Soc. **55** (1949).
[59] S. Eilenberg and S. Mac Lane, *Group extensions and homology*, Ann. of Math. **43** (1942), 757–831.
[60] S. Eilenberg and S. Mac Lane, *Relations between homology and homotopy groups*, Proc. Nat. Acad. Sci. **29** (1943), 155–158; *Relations between homology and homotopy groups of spaces*, Ann. of Math. **46** (1945), 480–509.
[61] S. Eilenberg and S. Mac Lane, *Natural isomorphisms in group theory*, Proc. Nat. Acad. Sci. **28** (1942); *General theory of natural equivalences*, Trans. Amer. Math. Soc. **58** (1945), 231–294.
[62] S. Eilenberg and S. Mac Lane, *Cohomology theory in abstract groups I*, Ann. of Math. **48** (1947), 51–78.
[63] S. Eilenberg and S. Mac Lane, *On the groups* $H(\Pi, n)$ *II*, Ann. of Math. **60** (1954), 49–139.
[64] S. Eilenberg and J. Moore, *Limits and spectral sequences*, Topology **1** (1962), 1–23.
[65] S. Eilenberg and J. Moore, *Homology and fibrations. I. Coalgebras, cotensor product and its derived functors*, Comment. Math. Helv. **40** (1966), 199–236.
[66] S. Eilenberg and N. Steenrod, *Axiomatic approach to homology theory*, Proc. Nat. Acad. Sci. USA **31** (1945), 117–120.
[67] S. Eilenberg and N. Steenrod, *Foundations of Algebraic Topology*, Princeton Univ. Press, Princeton, NJ (1952).
[68] S. Eilenberg and J. Zilber, *Semi-simplicial complexes and singular homology*, Ann. of Math. **51** (1950), 499–513.
[69] S. Eilenberg and J. Zilber, *On products of complexes*, Amer. J. Math. **75** (1953), 200–204.
[70] B. Feigin and B. Tsygan, *Cohomology of Lie algebras of generalized Jacobi matrices*, Funktional Anal. Appl. **17** (1983), 86–87; Functional Anal. Appl. **17** (1983), 153–155.
[71] H. Fitting, *Über den Zusammenhang zwischen dem Begriff der Gleichartigkeit zweier Ideale und dem Äquivalenzbegriff der Elementarteilertheorie*, Math. Ann. **112** (1936), 572–582.
[72] H. Freudenthal, *Der Einfluss der Fundamental Gruppe auf die Bettischen Gruppen*, Ann. of Math. **47** (46), 274–316.
[73] J.P. Freyd, *Abelian Categories*, Harper & Row, New York (1964).
[74] P. Gabriel, *Des catégories abéliennes*, Bull. Soc. Math. France **90** (1962), 323–428.
[75] S. Gersten, *On the functor* K_2, J. Algebra **17** (1971), 212–237.
[76] M. Gerstenhaber, *On the deformation of rings and algebras*, Ann. of Math. **79** (1964), 59–103; II, Ann. of Math. **84** (1966), 1–19.
[77] J. Giraud, *Analysis situs*, Séminaire Bourbaki 1962/1963, No. 256 (1964).
[78] J. Giraud, *Cohomologie nonabélienne*, C. R. Acad. Sci. Paris **260** (1965) 2666–2668.
[79] R. Godement, *Topologie algébrique et théorie des faisceaux*, Hermann, Paris (1958).
[80] T. Goodwillie, *Relative algebraic K-theory and cyclic homology*, Ann. of Math. **24** (1986), 347–402.

[81] A. Grothendieck, *Sur quelques points d'algèbre homologique*, Tôhoku Math. J. **9** (1957), 119–221.
[82] A. Grothendieck, *Fondements de la Géométrie Algébrique*, Extraits du Séminaire Bourbaki, 1957–1962, Secrétariat Mathématique, Paris (1962).
[83] A. Grothendieck and J. Dieudonné, *Éléments de Géométrie Algèbrique*, Part I: 4 (1960); Part II: 8 (1961); Part III: 11 (1961), 17 (1963); Part IV: 20 (1964), 24 (1965), 28 (1966), 32 (1967), Publ. I.H.E.S.
[84] A. Grothendieck, *Local Cohomology*, Lecture Notes vol. 41, Springer, Berlin (1967).
[85] M. Hall, *Group rings and extensions I*, Ann. of Math. **39** (1938), 220–234.
[86] M. Harada, *Note on the dimension of modules and algebras*, J. Inst. Poly. Osaka Univ. Ser. A **7** (1956), 17–27.
[87] D.K. Harrison, *Commutative algebras and cohomology*, Trans. Amer. Math. Soc. **104** (1962), 191–204.
[88] R. Hartshorne, *Residues and Duality*, Lecture Notes vol. 20, Springer, Berlin (1966).
[89] A. Heller, *Homological algebra in Abelian categories*, Ann. of Math. **68** (1958), 484–525.
[90] J. Herzog and E. Kunz, *Der kanonische Modul eines Cohen–Macauley-Rings*, Lecture Notes vol. 238, Springer, Berlin (1971).
[91] D. Hilbert, *Über die Theorie der algebraischen Formen*, Math. Ann. **36** (1890), 473–534.
[92] P. Hilton and U. Stammbach, *A Course in Homological Algebra*, Springer, Berlin (1971).
[93] G. Hochschild, *Semisimple algebras and generalized derivations*, Amer. J. Math. **64** (1942), 677–694.
[94] G. Hochschild, *On the cohomology groups of an associative algebra*, Ann. of Math. **46** (1945), 58–67.
[95] G. Hochschild, *Local class field theory*, Ann. of Math. **51** (1950), 331–347.
[96] G. Hochschild, *Relative homological algebra*, Trans. Amer. Math. Soc. **82** (1956), 246–269.
[97] G. Hochschild, B. Kostant and A. Rosenberg, *Differential forms on regular affine algebras*, Trans. Amer. Math. Soc. **102** (1962), 383–408.
[98] G. Hochschild and T. Nakayama, *Cohomology in class field theory*, Ann. of Math. **55** (1952), 348–366.
[99] G. Hochschild and J.-P. Serre, *Cohomology of group extensions*, Trans. Amer. Math. Soc. **74** (1953), 110–134; *Cohomology of Lie algebras*, Ann. of Math. **57** (1953), 591–603.
[100] M. Hochster and J. Roberts, *The purity of the Frobenius and local cohomology*, Adv. Math. **21** (1976), 117–172.
[101] O. Hölder, *Die Gruppen der Ordnungen p^3, pq^2, pqr, p^4*, Math. Ann. **43** (1893), 301–412.
[102] H. Hopf, *Über die Topologie der Gruppen-Mannigfaltigkeiten und ihre Verallgemeinerungen*, Ann. of Math. **42** (1941), 22–52.
[103] H. Hopf, *Fundamentalgruppe und zweite Bettische Gruppe*, Comment. Math. Helv. **14** (1941–1942), 257–309; announced, *Relations between the fundamental group and the second Betti group*, Lectures in Topology, Ann. Arbor (1942), 315–316.
[104] H. Hopf, *Über die Bettischen Gruppen, die zu einer beliebigen Gruppe gehören*, Comment. Math. Helv. **17** (1944–1945), 39–79.
[105] W. Hurewicz, *Beitrage zur Topologie der Deformationen IV. Asphärische Räume*, Proc. Akad. Amsterdam **39** (1936), 215–224.
[106] W. Hurewicz, *On duality theorems*, abstract 47-7-329, Bull. Amer. Math. Soc. **47** (1941), 562–563.
[107] N. Jacobson, *Abstract derivation and Lie algebras*, Trans. Amer. Math. Soc. **42** (1937), 206–224.
[108] M. Jibladze and T. Pirashvili, *Cohomology of algebraic theories*, J. Algebra **137** (1991), 253–296.
[109] D. Kan, *Abstract homotopy. III, IV*, Proc. Nat. Acad. Sci. **42** (1956), 419–421, 542–544.
[110] D. Kan, *On functors involving c.s.s. complexes*, Trans. Amer. Math. Soc. **87** (1958), 330–346.
[111] I. Kaplansky, *Modules over Dedekind rings and valuation rings*, Trans. Amer. Math. Soc. **72** (1952), 327–340.
[112] I. Kaplansky, *Projective modules*, Ann. of Math. **68** (1958), 372–377.
[113] I. Kaplansky, *Homological dimension of rings and modules*, University of Chicago, Dept. of Math. (1959), reprinted as part III of *Fields and Rings*, Univ. of Chicago Press (1969).
[114] Kawada and J. Tate, *On the Galois cohomology of unramified extensions of function fields in one variable*, Amer. J. Math. **77** (1955), 197–217.
[115] J. Kelley and E. Pritcher, *Exact homomorphism sequences in homology theory*, Ann. of Math. **48** (1947), 682–709.
[116] F. Keune, *Algèbre homotopique et K-théorie algébrique*, C. R. Acad. Sci. Paris **273** (1971), A592–A595; *Derived functors and algebraic K-theory*, Lecture Notes in Math. vol. 341, Springer, Berlin (1973), 166–178.
[117] J.-L. Koszul, *Sur les opérateurs de dérivation dans un anneau*, C. R. Acad. Sci. Paris **224** (1947), 217–219; *Sur l'homologie des espaces homogènes*, 477–479.

[118] J.-L. Koszul, *Homologie et cohomologie des algèbres de Lie*, Bull. Soc. Math. France **78** (1950), 65–127.
[119] H. Künneth, *Über die Bettischen Zahlen einer Produktmannigfaltigkeit*, Math. Ann. **90** (1923), 65–85; *Über die Torsionszahlen von Produktmannigfaltigkeiten*, Math. Ann. **91** (1924), 125–134.
[120] S. Lang and J. Tate, *Principal homogeneous spaces over Abelian varieties*, Amer. J. Math. **80** (1958), 659–684.
[121] S. Lefschetz, *On singular chains and cycles*, Bull. Amer. Math. Soc. **39** (1933), 124–129.
[122] J. Leray, *L'anneau d'homologie d'une représentation*, C. R. Acad. Sci. Paris **222** (1946), 1366–1368.
[123] J. Leray, *Stucture de l'anneau d'homologie d'une représentation*, C. R. Acad. Sci. Paris **222** (1946), 1419–1422.
[124] J. Leray, *L'anneau spectral et l'anneau filtré d'homologie d'un espace localement compact et d'une application continue*, J. Math. Pures Appl. **29** (1950), 1–139.
[125] S. Lichtenbaum and M. Schlessinger, *The cotangent complex of a morphism*, Trans. Amer. Math. Soc. **128** (1967), 41–70.
[126] J.-L. Loday, *Symboles en K-théorie algébrique supérieure*, C. R. Acad. Sci. Paris **292** (1981), 863–866.
[127] J.-L. Loday and D. Quillen, *Homologie cyclique et homologie de l'algèbre de Lie des matrices*, C. R. Acad. Sci. Paris **296** (1983), 295–297; *Cyclic homology and the Lie algebra homology of matrices*, Comment. Math. Helv. **59** (1984), 569–591.
[128] S. Lubkin, *Imbedding of Abelian categories*, Trans. Amer. Math. Soc. **97** (1960), 410–417.
[129] R. Lyndon, *The cohomology theory of group extensions*, Duke Math. J. **15** (1948), 271–292.
[130] S. Mac Lane, *Groups, categories and duality*, Proc. Nat. Acad. Sci. **34** (1948); *Duality for groups*, Bull. Amer. Math. Soc. **56** (1950), 485–516.
[131] S. Mac Lane, *Homologie des anneaux et des modules*, Colloque de Topologie Algèbrique, Georges Thon, Liège (1956), 55–80.
[132] S. Mac Lane, *Homology*, Springer, Berlin (1963).
[133] S. Mac Lane, *Group extensions for 45 years*, Math. Intelligencer **10** (1988), 29–35.
[134] S. Mac Lane, *The Applied Mathematics Group at Columbia in World War II*, A Century of Mathematics in America, Vol. III, P. Duren, ed., Amer. Math. Soc., Providence, RI (1989), 495–515.
[135] W. Massey, *Exact couples in algebraic topology*, Ann. of Math. **56** (1952), 363–396.
[136] W. Massey, *A history of cohomology theory*, History of Topology, I.M. James, ed., Elsevier, Amsterdam (1999), 579–603.
[137] E. Matlis, *Injective modules over Noetherian rings*, Pac. J. Math. **8** (1958), 511–528.
[138] J.P. May, *Simplicial Objects in Algebraic Topology*, Van Nostrand, New York (1967).
[139] J.P. May, *Stable algebraic topology*, History of Topology, I.M. James, ed., Elsevier, Amsterdam (1999), 665–723.
[140] L. Mayer, *Über Abstakte Topologie I und II*, Monatsh. Math. Phys. **36** (1929), 1–42, 219–258.
[141] J. Milnor, *On axiomatic homology theory*, Pacific J. Math. **12** (1962), 337–341.
[142] B. Mitchell, *Rings with several objects*, Amer. J. Math. **8** (1972), 1–161.
[143] M. Nagata, *A general theory of algebraic geometry over Dedekind domains II. Separably generated extensions and regular local rings*, Amer. J. Math. **80** (1958), 382–420.
[144] Y. Nakai, *On the theory of differentials in commutative rings*, J. Math. Soc. Japan **13** (1961), 63–84.
[145] E. Noether, *Ableitung der Elementarteilertheorie aus der Gruppentheorie*, Nachrichten der 27 Januar 1925, Jahresbericht Deutschen Math. Verein. (2. Abteilung) **34** (1926), 104.
[146] B. Osofsky, *Homological dimension and the continuum hypothesis*, Trans. Amer. Math. Soc. **132** (1968), 217–230.
[147] T. Pirashvili and F. Waldhausen, *Mac Lane homology and topological Hochschild homology*, J. Pure Appl. Algebra **82** (1992), 81–98.
[148] H. Poincaré, *Œvres*, Vol. 6, R. Garnier and J. Leray, eds, Gauthiers-Villars, Paris (1953).
[149] L. Pontrjagin, *Sur les nombres de Betti des groupes de Lie*, C. R. Acad. Sci. Paris **200** (1935), 1277–1280; *Homologies in compact Lie groups*, Mat. Sb. **6** (1939), 389–422.
[150] D. Puppe, *On the formal structure of stable homotopy theory*, Colloq. on Algebraic Topology, Aarhus University (1962), 65–71.
[151] D. Quillen, *Homotopical Algebra*, Lecture Notes vol. 43, Springer, Berlin (1967).
[152] D. Quillen, *On the (co)homology of commutative rings*, Proc. Sympos. Pure Math. vol. 17, Amer. Math. Soc., Providence, RI (1970), 65–87.
[153] D. Quillen, *Cohomology of groups*, Proc. I. C. M., Nice (1970/1971), 47–51.
[154] D. Quillen, *Projective modules over polynomial rings*, Invent. Math. **36** (1976).

[155] B. Riemann, *Gesammelte Mathematische Werke und Wissenschaftlicher Nachlass*, H. Weber, ed., Teubner, Leipzig (1876).

[156] G. Rinehart, *Differential forms on general commutative algebras*, Trans. Amer. Math. Soc. **108** (1963), 195–222.

[157] J.E. Roos, *Sur les foncteurs dérivés de lim*, C. R. Acad. Sci. Paris **252** (1961), 3702–3704.

[158] J.E. Roos, *Relations between the Poincaré–Betti series of loop spaces and of local rings*, Lecture Notes in Math. vol. 740, Springer, Berlin (1979), 285–322.

[159] J. Rotman, *Notes on Homological Algebra*, Van Nostrand, New York (1970); *An Introduction to Homological Algebra*, Academic Press, New York (1979).

[160] O. Schreier, *Über die Erweiterungen von Gruppen I*, Monatsh. Math. Phys. **34** (1926), 165–180.

[161] I. Schur, *Über die Darstellungen der endlichen Gruppen durch gegebene lineare Substitutionen*, J. Reine Angew. Math. **127** (1904), 20–50.

[162] J.-P. Serre, *Cohomologie des extensions de groupes*, C. R. Acad. Sci. Paris **231** (1950), 643–646.

[163] J.-P. Serre, *Homologie singulière des espaces fibrés*, Ann. of Math. **54** (1951), 425–505.

[164] J.-P. Serre, *Faisceaux algébrique cohérents*, Ann. of Math. **61** (1955), 197–278.

[165] J.-P. Serre, *Géométrie algébrique et géométrie analytique (GAGA)*, Ann. Inst. Fourier (Grenoble) **6** (1955–1956), 1–42.

[166] J.-P. Serre, *Sur la dimension homologique des anneaux et des modules noethériens*, Proc. Sympos. Algebraic Number Theory 1955, Science Council of Japan, Tokyo (1956), 175–189.

[167] J.-P. Serre, *Modules projectifs et espaces fibrés à fibre vectorielle*, Exposé 23, Seminaire P. Dubreil 1957/1958, Secrétariat Mathématique, Paris (1958).

[168] J.-P. Serre, *Cohomologie Galoisienne*, Lecture Notes in Math. vol. 5, Springer, Berlin (1964).

[169] J.-P. Serre, *Algèbre Locale Multiplicités*, Mimeographed Notes, I.H.E.S. (1957); Lecture Notes in Math. vol. 11, Springer, Berlin (1965).

[170] R. Sharp, *Local cohomology theory in commutative algebra*, Quart. J. Math. **21** (1970), 425–434.

[171] R. Sharp, *On Gorenstein modules over a complete Cohen–Macaulay local ring*, Quart. J. Math. **22** (1971), 425–434.

[172] U. Shukla, *Cohomologie des algèbres associatives*, Ann. Sci. École Norm. Sup. **78** (1961), 163–209.

[173] L. Smith, *Homological algebra and the Eilenberg–Moore spectral sequence*, Trans. Amer. Math. Soc. **129** (1967), 58–93.

[174] N. Spaltenstein, *Resolutions of unbounded complexes*, Compositio Math. **65** (1988), 121–154.

[175] N. Steenrod, *Universal homology groups*, Amer. J. Math. **58** (1936), 661–701.

[176] N. Steenrod, *Regular cycles of compact metric spaces*, Ann. of Math. **41** (1940), 833–851.

[177] A. Suslin, *Projective modules over polynomial rings are free*, Dokl. Akad. Nauk SSSR **229** (1976), 1063–1066; Soviet Math. Dokl. **17** (1976) 1160–1164.

[178] R. Swan, *Projective modules over finite groups*, Bull. Amer. Math. Soc. **65** (1959), 365–367.

[179] R. Swan, *Vector bundles and projective modules*, Trans. Amer. Math. Soc. **105** (1962), 264–277.

[180] R. Swan, *Nonabelian homological algebra and K-theory*, Proc. Sympos. Pure Math. vol. 17, Amer. Math. Soc., Providence, RI (1970), 88–123.

[181] R. Swan, *Some relations between higher K-functors*, J. Algebra **21** (1972), 113–136.

[182] J. Tate, *The higher dimensional cohomology groups of class field theory*, Ann. of Math. **56** (1952), 294–297.

[183] J. Tate, *The cohomology groups of algebraic number fields*, Proc. I.C.M., Amsterdam (1954), North-Holland (1954), 66–67.

[184] J. Tate, *Homology of Noetherian rings and local rings*, Illinois J. Math. **1** (1957), 14–27.

[185] B. Tsygan, *Homology of matrix Lie algebras over rings and the Hochschild homology*, Uspekhi Mat. Nauk **38** (1983), 217–218 (in Russian); Russian Math. Surveys **38** (1983), 198–199.

[186] O. Veblen and J. Alexander, *Manifolds of n dimensions*, Ann. of Math. **14** (1913), 163–178.

[187] J.-L. Verdier, *Catégories dérivées, état 0*, SGA $4\frac{1}{2}$, Lecture Notes in Math. vol. 569, Springer, Berlin (1977).

[188] F. Waldhausen, *Algebraic K-theory of spaces I*, Proc. Sympos. Pure Math. vol. 32, Amer. Math. Soc., Providence, RI (1978), 35–60.

[189] J.H.C. Whitehead, *On the decomposition of an infinitesimal group*, Proc. Cambridge Phil. Soc. **32** (1936), 229–237; *Certain equations in the algebra of a semi-simple infinitesimal group*, Quart. J. Math. **8** (1937), 220–237.

[190] H. Whitney, *On products in a complex*, Proc. Nat. Acad. Sci. **23** (1937), 285–291; Ann. of Math. **39** (1938), 397–432.

[191] H. Whitney, *Tensor products of Abelian groups*, Duke Math. J. **4** (1938), 495–528.

[192] N. Yoneda, *On the homology theory of modules*, J. Fac. Sci. Tokyo **7** (1954), 193–227.

CHAPTER 29

Topologists at Conferences

I.M. James
Mathematical Institute, 24–29 St. Giles, Oxford OX1 3LB, UK
E-mail: imj@maths.ox.ac.uk

Communication of the results of scientific research is usually achieved by publication in the appropriate scientific journals or by lectures on suitable occasions such as conferences. In the case of topology, journal publication has of course been extremely important but conferences have also played a vital role. Through the records of the international mathematical congresses one can trace the gradual acceptance of topology as a subject in its own right, until specialist conferences began to be organized. Nowadays, of course, such meetings are held very frequently but a century ago the situation was quite different.

The first International Congress of Mathematicians (see [1] for the proceedings and [12] for an account of congresses generally) was held at Zurich in 1897. Most of the participants came from nearby countries; there were no contributions from Great Britain or the United States. The two principal addresses were by Henri Poincaré "Sur les rapports de l'analyse pure et de la physique mathematique" and by Adolf Hurwitz "Über die Entwicklung der allgemeinen Theorie der analytischen Funktionen in neuerer Zeit"; the latter makes several references to analysis situs, the old term for topology; in particular he refers to the work of Jordan and Schoenflies. However the only lecture at the congress with a distinct topological flavour was that of Hermann Brunn, from Munich, entitled 'Über verknoten Kurven'. Others who one might expect to have been particularly interested in topology were Walter von Dyck from Munich, Felix Hausdorff from Leipzig, Arthur Schoenflies from Göttingen and, of course, Henri Poincaré from Paris.

Although the Zurich congress was the first of the regular series, there was an International Mathematical Congress of a rather different type at Chicago four years earlier (see [13] for the proceedings and [14, pp. 309–326], for the significance of this meeting for the development of mathematical research in North America). After the Zurich congress the next in the series was held in Paris in 1900 (see [2]). Elie Cartan and Henri Poincaré played a leading role in this, but neither chose to lecture on anything of a topological nature. Vito Volterra gave an interesting historical talk on "Betti, Brioschi, Casorati" which began:

> Dans l'automne de l'année 1858, trois jeunes géométres italiens partaient ensemble pour un voyage scientifique. Leur but était de visiter les Universités de France et

HISTORY OF TOPOLOGY
Edited by I.M. James

> d'Allemagne, d'entrer en rapport avec les savants les plus remarquables, d'en connaître les idées et les aspirations scientifiques et, en même temps, de repandre leurs travaux.

However, Volterra had little to say about Betti's interests in topology. It was at the Paris Congress that Hilbert presented his celebrated agenda of outstanding mathematical problems. Only one of the twenty-three could be described as topological, the famous fifth problem about Lie groups, finally settled by Montgomery and Zippin sixty years later.

From 1900, apart from interruptions due to the two world wars, congresses have been held regularly every four years. Thus the next one was held in 1904, with Heidelberg the venue (see [3]). More than half the participants were from Germany. Hausdorff and Schoenflies were present again, but one also notices the attendance of Max Dehn from Munster and Wilhelm Wirtinger from Vienna. Apparently there were no lectures of a topological nature.

It is in the proceedings of the 1908 congress in Rome (see [4]) that one begins to see signs that topology is gaining recognition. Poincaré himself, in the course of his address, declared:

> L'importance de l'analysis situs est énorme et je ne saurais trop y insister; le parti qu'en a tiré Riemann, l'un de ses principaux créateurs, suffirait à le demontrer. Il faut qu'on arrive à la construire complétement dans les espaces supérieurs; on aura alors un instrument qui permettra réellement de voir dans l'hyperspace et de suppléer à nos sens.

In the list of participants at the Rome Congress one notices not only the names of some of the participants at earlier congresses but also some new names, such as those of L.E.J. Brouwer, Camille Jordan and Heinrich Tietze. Brouwer gave two lectures, one on the 3-dimensional case of Hilbert's fifth problem, the other on set theory. But apparently these were the only lectures of a topological character.

Cambridge was selected as the venue for 1912, having been unsuccessful in the previous round. Due to a strong showing from the United Kingdom the number of participants was much greater than at previous congresses, over 700 altogether (see [5]). However, few of them were in any sense topologists. Brouwer and Tietze were present but not Dehn. Fréchet appears for the first time. Lectures of a topological character were given by Janiszewski and König.

Due to the first world war there was no congress in 1916, and so the next meeting was not until 1920, when Strasbourg was the venue. Participants included Antoine, Fréchet, Jordan, Lefschetz and Nielsen, with Elie Cartan playing a leading role (see [6]). No topology was featured in the formal programme.

These early congresses usually lasted up to a week. The number of participants and make-up by nationality varied a good deal according to the choice of venue. Apart from the formal addresses and lectures given to the whole congress the lecture programme, arranged in parallel sessions, could provide as many as a hundred shorter lectures. When topology featured on the programme, which was not often, it was regarded as a branch of geometry. It must be appreciated that the number of mathematicians with expertise in the subject was still quite small.

Attendance at the earlier congresses, which were all held in Europe, was dominated by the Europeans. Although the importance of North America for mathematical research was steadily increasing throughout the first half of the century, the Atlantic crossing was such a major undertaking that even the leading American mathematicians seldom put in an appearance. It was therefore of special significance that for 1924 the congress met in Toronto (see [8]). As was only to be expected, most of the participants were North Americans; as in 1920 mathematicians from the Central Powers were excluded. Fréchet was there and lectured three times. There was a talk on the four-colour problem by Errera. Elie Cartan lectured on 'La théorie des groupes et les recherches récentes de géométrie différentielle'. Following the Toronto meeting the prestigious Fields Medals were instituted.

For 1928 the congress returned to Italy and met in Bologna; mathematicians from Germany and Austria were no longer excluded. Among the participants with a special interest in topology one notices (see [6]) Alexandroff, both Elie and Henri Cartan, Čech, Fréchet, Heegard, Kuratowski, Menger, Newman, Reidemeister, Sierpinski, Veblen and Wirtinger. For the first time there was a session devoted to 'research of a topological character' at which the programme, presided over by the elder Cartan, was as follows:

P. Alexandroff, Das dimensionproblem und die ungeloesten Fragen allgemeiner Topologie.
B. Kerékjártó, On the general translation-theorem of Brouwer.
C. Kuratowski, Un système d'axiomes pour la topologie de la surface de la sphère.
K. Menger, Die Grundlagen der allgemeinen Kurventheorie.
W. Blaschke, Questioni topologiche di geometria differenziale.
L. Lusternik, Sur quelque méthodes topologiques dans la géométrie différentielle.
T. Bonnesen, Théorème de Brunn–Minkowski sur les corps convexes.
S. Cohn-Vossen, Der Index einer Nabel-punktes im Netze der Krummungslinien.
F. Gonseth et G. Juvet, Sur le problème des quatre couleurs.
K. Reidemeister, Fundamental gruppe und Ueberlagerung von Mannigfaltigkeiten.

Thus the 1928 congress at Bologna seems to have been something of a turning-point as regards recognition of the status of topology at international congresses. Four years later, at the 1932 congress (see [9]), the position of the subject seemed to be assured. Twenty-five years after the initial congress, the chosen venue was again Zurich. Topologists present included Alexander, Aleksandroff, Borsuk, Čech, Hopf, Hurewicz, Kuratowski, Menger, Morse, Newman, de Rham, Seifert, Threlfall, Tietze, Tucker, Ulam, Whitehead, and Wirtinger. Brouwer presided over one of the geometry sessions. Contrasting views of the subject were presented by Alexander, in the combinatorial tradition, and Menger, in the set-theoretic tradition. There were a number of other lectures of topological interest, including one by Čech on the higher-dimensional homotopy groups. The report of Čech's lecture is brief and uninformative, but those who were present seem to agree that its reception was such as to discourage him from pursuing the study of these invariants any further. Apparently it was thought that the commutative nature of the higher homotopy groups meant that they could not be any use, and in any case no-one knew how go about computing them.

The last congress before the second world war was held in Oslo in 1936 (see [10]). A fairly representative gathering of topologists took part, including Aleksandroff, Borsuk, Dehn, Ehresmann, Eilenberg, Freudenthal, Lefschetz, Heegard, Hurewicz, Morse, Nielsen, Seifert, Threlfall, Veblen and Whitehead, and this time topology was also better represented on the lecture programme. Elie Cartan and Nielsen gave plenary addresses. Among the other lectures one notices Marty 'Sur la théorie du groupe fondamental', Newman and Whitehead 'On the group of a certain linkage', Borsuk 'Über Addition der Abbildungsklassen', Pontrjagin 'Sur les transformations des sphères en sphères', and Hurewicz 'Lokaler Zusammenhang und stetige Abbildungen'.

After the war broke out communication between mathematicians in different countries became much more difficult, and this created problems. For example, Pontrjagin, in the lecture mentioned above, had asserted that the stable group of the 2-stem, in the homotopy groups of spheres, was trivial. The proof depended on what is now known as the Pontrjagin–Thom construction, but since details were not given it could not be checked. However, this key result was provisionally accepted and it was not until almost 10 years later that G.W. Whitehead conclusively demonstrated that the group in question was non-trivial, in fact of order two like the stable group of the 1-stem.

By the mid-thirties there were few European countries where topology was not being studied and the subject was also well established in the United States. The desirability of organizing a specialist conference to report on and discuss the latest research must have been obvious. The first truly international conference on topology was that organised in Moscow from September 4–10, 1935, entitled 'Première Réunion Topologique International'.

It was not confined to classical topology, in the sense used in this volume, but included general topology and adjacent areas of the subject. The proceedings, which were published in [16], give some idea of the research activity which was taking place in topology at this period:

J.W. Alexander, On the ring of a complex and the combinatory theory of integration. *
Paul Alexandroff, Einige Problemstellungen in der mengentheoretischen Topologie.
Garrett Birkhoff, Continuous groups and linear spaces.
Karol Borsuk, Über sphäroidale und H-sphäroidale Raume.
E. Čech, Accessibility and homology.
E. Čech, Betti groups with different coefficient groups. *
St. Cohn-Vossen, Topologische Fragen der Differentialgeometrie im Grossen.
D. van Dantzig, Neuere Ergebnisse der topologischen Algebra.
Hans Freudenthal, Entwicklungen von Raumen und Gruppen.
J.J. Gordon, On the intersection invariants of a complex and its residual space. *
Paul Heegard, Bemerkungen zum Vierfarbenproblem.
H. Hopf, Neue Untersuchungen über n-dimensionale Mannigfaltigkeiten. *
W. Hurewicz, Homotopie und Homologie.
E.R. van Kampen, On the structure of a compact group.
A. Kolmogoroff, Homologiering des Komplexes und des lokal-bikompakten Raumes.
Nicolas Kryloff et Nicolas Bogoliouboff, Les mesures invariants et transitives dans la mécanique nonlinéaire.
Casimir Kuratowski, Sur les ensembles projectifs.

Group photograph taken at the 1935 International Conference on Topology in Moscow. *Top Row*: 1. E. Čech; 2. H. Whitney; 3. K. Zarankiewicz; 4. A. Tucker; 5. S. Lefschetz; 6. H. Freudenthal; 7. F. Frankl; 8. J. Nielsen; 9. K. Borsuk; 10. ?; 11. J.D. Tamarkin; 12. ?; 13. V.V. Stepanoff; 14. E.R. van Kampen; 15. A. Tychonoff; *Bottom Row*: 16. C. Kuratowski; 17. J. Schauder; 18. St. Cohn-Vossen; 19. P. Heegaard; 20. J. Różańska; 21. J.W. Alexander; 22. H. Hopf; 23. P. Alexandroff; 24. ?.


Группа Первой Международной
Топологической Конференции,
Москва, 4-10 Сентября 1935 года

S. Lefschetz, Locally connected sets and their applications.
A. Markoff, Über die freie Äquivalenz der geschlossenen Zopfe. *
Stefan Mazurkiewicz, Über die Existenz unzerlegbares Kontinua. *
J. Nielsen, Topologische invarianten der Klassen von Flachenabbildungen. *
V. Niemytzki, Unstabile dynamische Systeme. *
J. von Neumann, The uniqueness of Haar's measure.
G. Nobeling, Beweisskizze fur die Triangulierung einer Mannigfaltigkeit und die sogenannte Hauptvermutung. *
L. Pontrjagin, Propriétés topologiques des groupes de Lie compacts. *
Georges de Rham, Sur les nouveaux invariants topologiques de M. Reidemeister.
Julia Rozanska, Über stetige Abbildungen eines Elementes.
J. Schauder, Einige Anwendungen der Topologie der Funktionalraume.
W. Sierpinski, Sur les images biunivoques et continues dans un sens.
W. Sierpinski, Sur les transformations des ensembles par les fonctions de Baire.
W. Sierpinski, Sur un ensemble projectif de classe 2 dans l'espace des ensembles fermés plans.
P.A. Smith, Transformations of period two.
M.H. Stone, Applications of Boolean algebras to topology.
A.W. Tucker, Cell spaces.
A. Tychonoff, Sur les points invariants des espaces bicompacts. *
André Weil, Démonstration topologique d'un théorème de Cartan.
André Weil, Les familles de courbes sur le tore.
Hassler Whitney, Differentiable manifolds in Euclidean space.
Hassler Whitney, Sphere-spaces.

Of the 34 lecturers who contributed to the proceedings (the contributions marked $*$ appear by title only, being published elsewhere), many lectured on topics which might not be classified as topological today. In fact, it was not until somewhat later that the tendency developed for the different branches of topology each to go their own way, in particular, general topology became increasingly separate from other kinds of topology. There were a few other participants who did not contribute to the proceedings, making about 40 in all.

Whitney gave a vivid account of the Moscow conference in an article [15] he wrote not long before his death, from which I quote:

> What was the main import of the conference? As I see it, it was threefold:
> (1) It marked the true birth of cohomology theory, along with the products among cocycles and cycles.
> (2) An item of application, vector fields on manifolds, was replaced by an expansive theory, of vector bundles.
> (3) The pair of seemingly diverse fields, homology and homotopy, took root and flourished together from then on.

He goes on to remark that in each of these major breakthroughs the first presenter turned out to be not alone; at least one other had been working up the same material. Thus (1) was presented by Kolmogoroff, not usually classed as a topologist, but when he had finished Alexander announced that he too had essentially the same definition and results. In the case of (2) Hopf presented the results of his student Stiefel concerning the existence of several independent vector fields in a manifold, which was just what Whitney himself had gone to Moscow to present.

As we have seen, Čech presented a definition of the higher homotopy groups at the 1932 congress. Apparently unaware of this, Hurewicz gave another version of the definition, with several simple but important applications, at the Moscow conference. In the subsequent discussion Alexander said that he had considered the idea many years earlier, but had rejected it since it was too simple in character and, hence, could not lead to deep results. Both Čech and van Danzig also said that they had considered introducing the same concept as Hurewicz, and if Dehn had been present he might have said the same. To further quote Whitney:

> Tucker spoke on cell spaces, a thesis written under Lefschetz's direction, which gave certain specifications about what can usefully be considered a 'complex'. Nobeling's talk was to present, in outline, the proof that all topological manifolds can be triangulated. According to von Neumann, Nobeling demonstrated amply that he had answers to every possible question that one might think of.

Unfortunately, Nobeling's argument contained an error, as van Kampen was soon to point out, although it not until 1969 that the Hauptvermutung was finally answered, by Kirby and Siebenmann, in the negative.

Only a month later there was a second international conference, this time at the University of Geneva, under the title 'Colloque sur quelques questions de Géométrie et de Topologie'. Some of those who lectured in Moscow also did so in Geneva. The programme was as follows.

E. Cartan, La topologie des espaces representatifs des groupes de Lie.
P. Alexandroff, title not available.
G. de Rham, Relations entre la topologie et la théorie des integrales multiples.
C. Kuratowski, La notion de connexité locale en topologie.
A. Weil, La mesure invariante dans les espaces homogènes clos.
W. Threlfall, Quelques progrès récents de la topologie algébrique.
E.G. Togliatti, Extension aux surfaces algébriques de la théorie des séries de groupes de points.
J. Nielsen, Topologie des transformations des surfaces.
B. Kaufmann, Topologie des surfaces closes et des variétés de Cantor.
B. de Kerekjarto, Sur la structure des transformations des surfaces en elles-mêmes.
C. Ehresmann, Les espaces localement homogènes.
H. Hopf, Quelques problèmes de la théorie des representations continues.
K. Menger, La géometrie métrique.
P. Finsler, Courbures supérieures dans les espaces généraux.
H. Seifert, La théorie des noeuds.
P. Heegard, Contribution à la théorie des 'graphes' de Tait.
G. Bouligand, Le role de la théorie des groupes en géométrie infinitésimale.

The Geneva meeting seems to have been rather overshadowed by the Moscow conference, at which America was more strongly represented. Nothing of this nature is on record for 1937 or 1938, and plans for another conference on the Moscow model to be held at Warsaw in 1939 were abandoned due to the deteriorating international situation. In America, however, a conference was organised at the University of Michigan in 1940 (see [18] for the proceedings) although by then the second world war was in progress and inevitably the international dimension was much reduced. Nevertheless the list of speakers and their titles provides some idea of research activity at this period. Note that general topology still remains in a close relationship with other kinds of topology.

Solomon Lefschetz, Abstract complexes.
R.L. Wilder, Uniform local connectedness.
N.E. Steenrod, Regular cycles of compact metric spaces.
Samuel Eilenberg, Extension and classification of continuous mappings.
Hassler Whitney, On the topology of differentiable manifolds.
Stewart S. Cairns, Triangulated manifolds and differentiable manifolds.
P.A. Smith, Periodic and nearly periodic transformations.
Leo Zippin, Transformation groups.
Saunders MacLane and V.W. Adkisson, Extensions of homeomorphisms on the sphere.
O.G. Harrold, Jr., The role of local separating points in certain problems of continuum structure.
L.W. Cohen, Uniformity in topological space.
E.W. Chittenden, On the reduction of topological functions.
Edward G. Begle, Homology local connectness.
Claude Chevalley, Two theorems on solvable topological groups.
Ralph H. Fox, Topological invariants of the Lusternik–Schnirelmann type.
O.H. Hamilton, Concerning the decomposition of continua.
Wilfred Kaplan, Differentiability of regular curve families on the sphere.
Erich Rothe, On topology in function spaces.
John W. Tukey, Compactness in general spaces.
E.R. van Kampen, Remark on the address of S.S. Cairns.
Heinz Hopf, Relations between the fundamental group and the second Betti group.

A meeting at Princeton in 1946 on 'The Problems of Mathematics' might also be mentioned at this point since it resulted in the publication in the Annals of Mathematics of a useful collection of unsolved problems, edited by Eilenberg [19]. As well as the classical unsolved problems, such as the Hauptvermutung, the collection includes problems about lens spaces, retracts and local connectedness, homotopy groups, homotopy classification, fibre bundles, homology theory and transformation groups.

In Europe, for some years after the end of the second world war, international conferences could only be organised with difficulty. However, the Rockefeller Foundation, working with the French body Centre National de la Recherche Scientifique, made it possible for several scientific colloquia to be held in Paris as early as the summer of 1947. One of these was in algebraic topology (see [17] at which the speakers were as follows.

Henri Cartan, Sur la notion de carapace en topologie algébrique.
Charles Ehresmann, Sur la théorie des espaces fibrés.
Hans Freudenthal, La géométrie énumérative.
Guy Hirsch, La géométrie projective et la topologie des espaces fibrés.
W.V.D. Hodge, The finite algebraic form of the theory of harmonic integrals.
Heinz Hopf, Sur les champs d'éléments de surface dans les variétés a 4 dimensions.
Jean Leray, L'homologie filtrée.
Henri Cartan and Jean Leray, Relations entre anneaux d'homologie et groupes de Poincaré.

Georges de Rham, Sur les conditions d'homéomorphie de deux rotations de la sphère à n dimensions, et sur les complexes avec automorphismes.
E. Stiefel, Sur les nombres de Betti des groupes de Lie clos.
J.H.C. Whitehead, On simply connected, 4-dimensional polyhedra.
Hassler Whitney, La topologie algébrique et la théorie de l'intégration.

There was another rather similar meeting in Brussels in 1950, at which the speakers included Cartan, Eckmann, Ehresmann, Hirsch, Hopf, Koszul and Leray. But perhaps the most important topology conference of this period was that held in conjunction with the International Mathematical Congress of 1950 (see [11]), the first after the end of the war, which was held at Harvard University and other institutions in the Boston area. Hassler Whitney was Chairman of the conference and in his introduction he said:

> The subject of algebraic topology and applications was chosen for one of the conferences of the Congress because of its great growth in recent years, and the increasingly large contact with other fields of mathematics, in geometry, algebra and analysis. The subject of general topology has moved considerably into the domain of analysis. It was with great regret that the field of point set theory had to be omitted altogether.

In fact, general topology increasingly went its own way. The textbook of Schubert was the last to follow the example of Aleksandroff and Hopf by combining algebraic topology with general topology. At conferences it was considered more useful to emphasize the links with other kinds of mathematics such as differential geometry.

The programme of the Harvard Conference in Topology was divided up as follows.

Homology and homotopy theory

W. Hurewicz, Homotopy and homology.
S. Eilenberg, Homotopy groups and algebraic homology theories.
J.H.C. Whitehead, Algebraic homotopy theory.
G.W. Whitehead, Homotopy groups of spheres.

Fibre bundles and obstructions

P. Olum, The theory of obstructions.
W.S. Massey, Homotopy groups of triads.
G.C. Hirsch, Homology invariants and fibre bundles.
E.H. Spanier, Homology theory of fiber bundles.

Differentiable manifolds

S.S. Chern, Differential geometry of fiber bundles.
C. Ehresmann, Sur les variétés presque complexes.
B. Eckmann, Complex-analytic manifolds.
C.B. Allendoerfer, Cohomology on real differentiable manifolds.

Topological groups

P.A. Smith, Some topological notions associated with a set of generators.
D. Montgomery, Properties of finite-dimensional groups.
K. Iwasawa, Some properties of (L)-groups.
A.M. Gleason, One parameter subgroups and Hilbert's fifth problem.
R.H. Fox, Recent developments in knot theory at Princeton.

In addition to these lectures there were a considerable number of lectures at the Congress itself which would have been of interest to topologists. For example, Hopf delivered one of the opening addresses, entitled 'Die n-dimensionalen Sphären und projektiven Räume in der Topologie'. Most of the participants were from North America although special efforts were made to encourage participants from elsewhere. Among those present with a particular interest in topology were:

> Cahit Arf, A.L. Blakers, Raoul Bott, D.G. Bourgin, L.E.J. Brouwer, S.S. Cairns, R.E. Chamberlin, James Dugundji, Beno Eckmann, Charles Ehresmann, Samuel Eilenberg, R.H. Fox, Alex Heller, Heinz Hopf, S.T. Hu, Witold Hurewicz, J.L. Kelley, Solomon Lefschetz, Jean Leray, C.B. de Lyra, Saunders MacLane, W.S. Massey, Karl Menger, J.W. Milnor, E.E. Moise, Deane Montgomery, J.C. Moore, Marston Morse, J.F. Nash, M.H.A. Newman, Paul Olum, R.S. Palais, Everett Pitcher, Moses Richardson, Hans Samelson, Paul Smith, E.H. Spanier, N.E. Steenrod, A.W. Tucker, S.M. Ulam, Oswald Veblen, A.D. Wallace, A.H. Wallace, Henry Wallman, G.W. Whitehead, J.H.C. Whitehead, Hassler Whitney, R.L.Wilder, Shaun Wylie, J.A. Zilber, and Leo Zippin.

As far as I am aware the first truly international conference on topology after the war was the one held at the National University of Mexico in 1956, entitled 'Symposium Internacional de Topología Algebraica'. In fact, although general topology was not included, the scope of the programme was broader than the title suggests. There was a large attendance of a strongly international character, where topologists of the older generation, such as Henri Cartan, Witold Hurewicz, Solomon Lefschetz and Henry Whitehead interacted with some of the rising stars of the new generation, such as Michael Atiyah, Raoul Bott, Jean-Pierre Serre and René Thom. If we compare the programme with that of the 1935 Moscow conference, hardly more than twenty years previously, it is obvious how much has changed.

Witold Hurewicz and Edward Fadell, On the structure of higher terms of the spectral sequence of a fibre space.
Henri Cartan and Samuel Eilenberg, Foundations of fibre bundles.
Jean-Pierre Serre, Sur la topologie des variétés algébriques en charactéristique p.
R. Thom, Les classes charactéristiques de Pontrjagin des variétés triangulées.
D.C. Spencer, A spectral resolution of complex structure.
M.F. Atiyah, Complex analytic connections in fibre bundles.
Raymond Raffin, Remarques sur certaines algèbres de Lie.
Shiing-Chen Chern, Geometry of submanifolds in a complex projective space.
Henri Cartan, Espaces fibrés analytiques.
John Milnor, On simply-connected 4-manifolds.
Friedrich Hirzebruch, Automorphe Formen und der Satz von Riemann–Roch.
W.S. Massey, Some higher order cohomology operations.
Emery Thomas, The generalized Pontrjagin cohomology operations.
Franklin P. Peterson, Functional higher order cohomology operations.
N.E. Steenrod, Cohomology operations.
José Adem, Operaciones cohomológicas de segundo orden asociadas con cuadrados de Steenrod.
I.M. James, On the homotopy groups of spheres.

Daniel M. Kan, The Hurewicz theorem.
John C. Moore, Semi-simplicial complexes and Postnikov systems.
J.H.C. Whitehead, Duality between CW-lattices.
E.H. Spanier, Duality and the suspension category.
P.J. Hilton, Homotopy theory of modules and duality.
R. Bott and H. Samelson, Applications of Morse theory to symmetric spaces.
Hassler Whitney, Singularities of mappings of Euclidean spaces.
B.A. Rattray, Generalizations of the Borsuk–Ulam theorem.
James Eells, Jr., On the geometry of function spaces.
Paul Dedecker, On the exact cohomology sequence of a space with coefficients in a non-Abelian sheaf.
I. Fary, Spectral sequences of certain maps.

Many exciting new results were announced at the symposium but the sensation of the meeting was the news that the young American mathematician John Milnor had shown that the 7-sphere admits more than one differential structure. Later he received a Fields medal for his work in this area. At the end of the meeting participants were shocked to hear that Witold Hurewicz had suffered a fatal accident in Yucatan; he was visiting one of the Mayan pyramids at Uxmal when he wandered off limits and fell to his death. The proceedings of the symposium [20] were dedicated to his memory.

Of course conferences were also being held with a more geometric emphasis. For example one took place in 1961 at the University of Georgia (USA) on the Topology of 3-manifolds with almost 40 participants, mostly from the USA. The lectures (see [24]) were grouped under the following headings: Decompositions and subsets of the plane, n-manifolds, Knot theory, The Poincaré conjecture, Periodic maps and isotopies, Applications. The Proceedings were dedicated to J.H.C. Whitehead, who had died suddenly the previous year and who would surely have played a leading role at the meeting had he lived.

In recent years the custom has grown up of using an appropriate birthday or other occasion to pay tribute to one of the leading researchers. Although former students and others who have been closely associated professionally with the person in question tend to predominate, these conferences can be quite large. In 1990, for example, such a meeting had been planned to mark the sixtieth birthday of the British homotopy theorist Frank Adams, but following his untimely death as the result of a motor accident it was transformed into a memorial meeting instead. Over 150 topologists, from almost twenty different countries, came together in Manchester for this symposium. The proceedings [21] include a survey of Adams' work as well as reports of recent research on topics in which he had been interested.

There are also some well-established *series* of conferences on topology. For example, a relatively small meeting, with an emphasis on younger topologists, takes place every autumn at the Mathematisches Forschungsinstitut in Oberwolfach. Another long-standing series, on a somewhat larger scale, is that of the Oxford Topology Symposia, where the emphasis is on homotopy theory. These meetings take place every four years, approximately, and although at first they were always held at the Mathematical Institute in Oxford, more recently other venues have been used, such as the Palazzo Feltrinelli on Lake Garda, by courtesy of the University of Milan. The large and enthusiastic attendance at these symposia is an indication of the continued vitality of the subject.

Bibliography

References to proceedings of the first eleven congresses

[1] *Verhandlungen des ersten internationalen Mathematiker-Kongress*, Rudio (ed.), Teubner, Leipzig (1898).
[2] *Compte Rendu du Deuxième Congrès International des Mathematiciens*, Gauthier-Villars, Paris (1902).
[3] *Verhandlungen des Dritten Internationalen Mathematiker Kongresses in Heidelberg*, Krazer (ed.), Teubner, Leipzig (1905).
[4] *Atti del IV Congresso Internazionale dei Matematici*, Castelnuovo (ed.), Academia dei Lincei, Rome (1909).
[5] *Proceedings of the Fifth International Congress of Mathematicians*, Hobson and Love (eds), Cambridge Univ. Press, Cambridge (1913).
[6] *Comptes Rendus du Congrès International des Mathematiciens*, Strasbourg 1920. Villat, Toulouse (1921).
[7] *Atti del Congresso Internazionale dei Matematici 1928*, Zanichelli, Bologna (1929).
[8] *Proceedings of the International Mathematical Congress*, Fields (ed.), University of Toronto, Toronto (1928).
[9] *Verhandlungen des Internationalen Mathematiker Kongresses Zurich*, Saxer (ed.), Fussli, Zurich/Leipzig (1932).
[10] *Comptes Rendus du Congrès International des Mathematiciens Oslo* (1936), Broggers, Oslo (1937).
[11] *Proceedings of the International Congress of Mathematicians*, Cambridge, MA, USA (1950), Amer. Math. Soc., Providence, RI (1952).

Other references

[12] D.J. Albers, G.L. Alexanderson and C. Reid, *International Mathematical Congresses: An Illustrated History 1893–1986*, Springer, New York (1987).
[13] *Mathematical Papers Read at the International Mathematical Congress*, Macmillan, New York (1896).
[14] K.H. Parshall and D.E. Rowe, *The Emergence of the American Mathematical Research Community, 1876–1900: J.J. Sylvester, Felix Klein and E.H. Moore*, Amer. Math. Soc. and London Math. Soc., Providence, RI (1991).
[15] H. Whitney, *Moscow 1935: Topology Moving toward America*, A Century of Mathematics in America, Duren, ed., Amer. Math. Soc., Providence, RI (1989).
[16] Matematicheskii Sbornik N.S. 1 (43) : 5 (1936).
[17] *Colloque International de Topologie Algébrique*, Paris (1947), Gauthier-Villars, Paris (1949).
[18] *Lectures in Topology*, Wilder and Ayres (eds), Univ. of Michigan Conf. of 1940, Univ. of Michigan Press, Ann Arbor (1941).
[19] S. Eilenberg, *On the problems of topology*, Ann. of Math. **50** (1949), 247–260.
[20] *Symposium Internacional de Topología Algebraica*, Universidad Nacional Autónoma de México and UNESCO (1958).
[21] *Adams Memorial Symposium on Algebraic Topology*, I, II, Ray and Walker (eds), London Math. Soc. Lecture Note Ser. vols 175, 176, Cambridge Univ. Press, Cambridge (1992).
[22] *Development of Science Publishing in Europe*, Meadows (ed.), Elsevier, Amsterdam (1980).
[23] I.M. James, *Topology: past, present and future*, Algebraic Topology, Carlsson, Cohen, Miller and Ravenel, eds, Lecture Notes in Math. vol. 1370, Springer, Berlin (1989).
[24] *Topology of 3-Manifolds and Related Topics*, Fort (ed.), Prentice-Hall, Englewood Cliffs, NJ (1962).

CHAPTER 30

Topologists in Hitler's Germany

S.L. Segal

Department of Mathematics, University of Rochester, Rochester, NY 14627, USA

E-mail: ssgl@troi.cc.rochester.edu

What happened to topology in Germany after January 30, 1933 when Adolph Hitler became chancellor are really two questions: what happened to topology, and what happened to topologists? Though "algebraic topology" goes back to the nineteenth century and has its beginnings in the work of G.F.B. Riemann, Felix Klein, Enrico Betti and, above all, Henri Poincaré; although L.E.J. Brouwer made some stunning advances in the early part of this century; nevertheless the 1920's saw a large discipline constructed on those beginnings. This was especially true in Germany, even though by 1911 important figures in topology like Max Dehn (David Hilbert's pupil), W. von Dyck (Klein's pupil and the inventor of the generator-relations description of groups), Heinrich Tietze (who though he had been Gustav von Escherich's student in Vienna, was influenced there by Wilhelm Wirtinger to an interest in topology) and Arthur Schoenflies (who had studied with E.E. Kummer but was strongly influenced by Klein) were already active. For example, it was von Dyck's work which inspired Poincaré to what we today call the Euler–Poincaré characteristic. Schoenflies, already in his forties when he turned his attention to topology, published several papers in the period 1903–1906. It was the errors and gaps in these which were an inspiration to L.E.J. Brouwer. Tietze's *Habilitationsschrift* in 1908 was one of the first significant studies of topological invariants. Dehn not only proved "Dehn's Lemma" in 1910 (though the proof was not correct), but together with Poul Heegaard, published one of the very first surveys of topology (the *Enzyklopädie* article of 1907–1910). In fact, it perhaps shows the difference between the topological sophistication of pre-World War I and the late twenties, that it was Hellmuth Kneser (in a footnote occasioned by his careful reading of the galley proofs of an article) who in 1928 pointed out that "Dehn's Lemma" had not been correctly proved. A correct proof seems to have waited until C. Papakyriakopoulos in 1952.

In Weimar Germany despite the initial exclusion, at French insistence, of the Germans from international gatherings and exchanges, topology flourished from these nascent beginnings. Among those stimulating such interest were also distinguished "nontopologists" like Erhard Schmidt. It was with Schmidt that Heinz Hopf wrote his dissertation (Ludwig Bieberbach was second reader) on the topology of manifolds (1925), which received the rarely awarded highest commendation for a dissertation. Hopf then traveled to Göttingen

HISTORY OF TOPOLOGY
Edited by I.M. James

for a year where he met Alexandroff, and "habilitated" in Berlin in 1926. In his evaluation of Hopf's dissertation, Schmidt spoke of topology as "the most difficult area of mathematics, and at the same time the one most rich with future possibilities". Before he left Berlin for Zürich in 1931, Hopf had two students, Hans Freudenthal and the differential geometer, Willi Rinow. While still at Breslau, Erhard Schmidt not only influenced Hopf, who followed him to Berlin, but apparently also Hellmuth Kneser, though the latter eventually went to Göttingen, wrote a dissertation on the quantum theory of the time under Hilbert's supervision, and then, seemingly deciding not to specialize, became one of the broadest mathematicians of his generation; but topology, the subject of his *Habilitationsschrift* was a constant interest. One final example of the topological scene in Germany in the pre-Hitler period deserves mention here. In 1930, Herbert Seifert received a Dr.rer.nat. with William Threlfall, got a Ph.D. in Leipzig, and returned to Dresden to "habilitate" in 1934. Seifert and Threlfall became known as "inseparable twins", frequently collaborating, especially on their famous textbook. Finally, it is worth mentioning that the topology of the time was far from removed from other mathematics. The most famous example is undoubtedly Emmy Noether's definition of homology groups which would seem to have first appeared in a paper of Hopf. Perhaps another example is that Hellmuth Kneser's first student who wrote a topological dissertation, and did some significant early work in topology, was the to-be-famous algebraist, Reinhold Baer. By January 30, 1933, Schoenflies had already died (in 1928), Heinz Hopf was in Zürich where he would remain throughout the Nazi period, and von Dyck, stricken with an illness which would be fatal, retired in 1933 (at age 77) and died the following year. Though remaining in Switzerland, Hopf did try to help colleagues in Germany who suffered Nazi persecution, such as the aerodynamicist, Ludwig Hopf, apparently a distant relative, who was tainted and dismissed as Jewish.

The roster of German university topologists when Hitler came to power thus ranged from Max Dehn (who was 54 and very well-established) to Herbert Seifert who was working towards his "*Habilitation*". They also included Hans Freudenthal, Hellmuth Kneser, Georg Feigl, Kurt Reidemeister, William Threlfall, Heinrich Tietze. To these should also be added the names of Hermann Künneth, who though he eventually became qualified as a university lecturer (at age 49), spent most of his career as a teacher at a Gymnasium in Erlangen, and Leopold Vietoris, an Austrian who was very active in the German Mathematical Society. There was also, of course, Felix Hausdorff, one of the greatest mathematicians of the first half of the twentieth century, whose *Grundzüge der Mengenlehre* completely surpassed Schoenflies' work (which with Brouwer's help, had appeared in a second edition in 1913). Hausdorff's book provided a set-theoretic foundation for topological ideas. Finally, mention must be made again of L.E.J. Brouwer in this context; for Brouwer, though Dutch, was an ardent Germanophile and "German nationalist," and played a significant role in the politics of German mathematics. In addition, while David Hilbert was never himself a topologist, and already showing signs of his age in 1933, some of his distinguished students who were not primarily topologists also stimulated topological activity; Erhard Schmidt is a good example. These are only some salient names; topology itself was thriving, and German mathematicians not primarily topologists (like Emmy Noether, Reinhold Baer, and Friedrich Levi) also made contributions. Germany was far from the only country in which topology was thriving in the twenties, but it certainly had more than its share of leaders.

Before turning to the fates of topologists under the Nazis, it is perhaps well to look at the subject itself. For some Nazis, at least for a while, had very definite ideas of what was and was not appropriate mathematics. This may sound astonishing, but mathematics and physics each saw specific different attempts to discern a truly Aryan subject from other varieties, and this was more than merely a question of the expulsion of Jews and communists. This article is not the place (if only for the reasons of space) to go into this in detail, but, briefly, the most prominent such Nazi movement saw "truly German" mathematics as intuitive and tied to nature, often geometrical, and certainly not axiomatic. Axiomatics, "logic chopping", too great abstraction, was Franco–Jewish. The details of how mathematicians like Richard Dedekind and David Hilbert were nevertheless qualified as "truly German" need not concern us here. Through no fault of his own, the deceased Felix Klein was apotheosized by this movement as its exemplar (rather like a Roman emperor divinized post-mortem). This movement was the *Deutsche Mathematik* associated principally with the names of Theodor Vahlen and Ludwig Bieberbach. Bieberbach was one of the leading mathematicians of his generation, as well as the most vocal protagonist of Nazi views with respect to mathematics. Furthermore, Bieberbach, though not a topologist, was associated with a number of the topological names previously mentioned. In 1910–1912, Bieberbach had partially solved "Hilbert's eighteenth problem", which partly stemmed from work of Schoenflies. In fact, this work formed his *Habilitationsschrift*. Though Bieberbach initially went with Ernst Zermelo to Zurich in 1910, Schoenflies, who was then a full professor in Königsberg, immediately arranged a position for him there, and this was the German university at which he "habilitated", also in 1910. In 1914, Schoenflies became the first full professor at the newly established university in Frankfurt (lectures began October 23, 1914), and he helped arrange for Bieberbach to be his full professorial colleague. When in 1921 Bieberbach accepted a call to Berlin, as is clear from the Bieberbach–Blaschke correspondence, he was actively involved in choosing his Frankfurt successor who turned out to be Max Dehn. Finally, at least from 1928 onward, and to some extent earlier, Bieberbach and L.E.J. Brouwer collaborated in conservative German ultra-nationalist politics affecting mathematics – in 1928, the issue was German attendance at the international congress in Bologna.

And Brouwer? By 1925, Brouwer was no longer working in topology: yet topologists of the caliber of Paul Alexandroff, Karl Menger, Leopold Vietoris and Hans Freudenthal stayed with him. Brouwer's interest in the foundations of mathematics dated back to his 1907 dissertation; but it was not until the publication in 1923 of his famous paper on the excluded middle that his mathematical work became almost exclusively in intuitionism. Topologists seemingly went to Brouwer because he was a great man in their field and one can always learn from a great man. However, Brouwer was more interested in promoting intuitionism and in Berlin he found an apparently willing audience. According to Hans Freudenthal, as early as 1923, the young Karl (Charles) Löwner gave a Berlin course in calculus on an intutionistic basis, and in 1926–1927, Brouwer gave there a very well received series of lectures on intuitionism. Some even talked of a mathematical "putsch". Freudenthal, for example, was so attached by Brouwer's lectures that, on obtaining his doctorate in 1930, he decided to go to the Netherlands to be with Brouwer, eventually becoming a professor at Utrecht. In 1926, Bieberbach gave an unpublished public lecture (a copy is in my possession) in which be castigated Hilbert, saw Brouwer's intuitionism as the coming mathematical philosophy, and Klein's views as an early predecessor of Brouwer's.

But what of topology for Nazi ideologists? Was it a "truly German" subject, being geometrical, or was it not so, being highly abstract? This is not as silly a question as it at first seems, since, for example, Bieberbach stigmatized both Cantorian set theory and Measure Theory as "non-German". Some ideologically Nazi mathematicians even looked suspiciously at abstract algebra as being not "as German" as, say, probability theory or geometric subjects.

Did topology have difficulty getting published? What happened to topologists under the Nazis? The political opinions of topologists such as those listed earlier, at the beginning of 1933 ran the gamut from Hellmuth Kneser, who was a very conservative nationalist, even a bit of a reactionary, to Kurt Reidemeister who was opposed to the Nazis on philosophical, intellectual and political grounds. Most German academics during the Weimar period, whatever their ethnicity, seem to have been conservative nationalists; mathematicians as a group do not seem to have been an exception, nor do topologists among mathematicians. Kneser, for example, though initially a supporter of Nazi ideals, seems to have been guided by a solid set of conservative and humanistic principles to an eventual rejection (by then necessarily tacit) of all the Nazis stood for. Of the individuals mentioned above, those with Jewish ansectors, of course, suffered severely under the Nazis. Max Dehn and Hans Freudenthal have separate articles in this volume devoted to them. Briefly, it should be said here that Freudenthal, because he had a non-Jewish wife, was saved from immediate deportation when the Germans occupied Holland during World War II. However, he spent six weeks in prison during 1942 and in 1944 was taken to a labor camp from which he later escaped and went into hiding for the duration of the war. Max Dehn was dismissed in 1935. He had not been immediately dismissed under the Nazi law of April 7, 1933, which called for the dismissal of civil servants (hence university professors) who were Jewish or unable fully to support the new government, because he fell under one of the exceptions clauses – he had been a civil servant prior to 1914. However, in 1935, he was suddenly dismissed as "supernumerary" (another clause in the law), in a second wave of dismissals. Carl Ludwig Siegel, in his story of Frankfurt mathematics, suggest that this was an act of revenge by Theodor Vahlen – in 1905 Dehn had savagely reviewed a geometry book by Vahlen. More likely is that this was simply the Nazi bureaucracy doing its job, since on January 21, 1935, a new law allowed for dismissal of professors when it was "in the national interest". This allowed the dismissal of the previously protected; in addition, Frankfurt had a reputation as a "Jewish university" and, for a while, the Nazis thought of closing it completely. What the immediate stimulus was for Dehn's dismissal seems unknown. Though dismissed, Dehn stayed in Germany and in November 1938, was arrested during the *Kristallnacht* pogrom. However, there being no room in the jail, he was temporarily released. To prevent his rearrest, the Dehns fled Frankfurt the next day. This began a long and arduous voyage of escape described in detail in John Stillwell's article. First via Hamburg to Copenhagen in early 1939 and then Trondheim, Norway. The Germans invaded Norway on April 19, 1940: no arrests immediately followed; enabling the Dehns to flee Norway in October of that year. With the war on, they thought the Atlantic waters more dangerous than going eastward, and so travelled to America via Stockholm, Moscow, Siberia (a ten-day rail journey with temperatures as low as −50°C.); from Vladivostok to Japan, and then San Francisco; ending finally at Idaho State University in Pocatello, Idaho. The Dehns moved three more times within the US, to the Illinois Institute of Technology in 1942, St. Johns in Annapolis the following year, and finally to Black Mountain College in North Carolina, a now defunct experimental college. One can wonder why a man of Dehn's distinction did not find more prestigious positions, but it must be remembered that despite the efforts of the Emergency Committee set up to help displaced scholars (whose "assistant secretary" was the young Edward R. Murrow), the United States did not unanimously welcome those refugee scholars with open arms. A similar point might be made about Great

Britain. Prominent mathematicians like Oswald Veblen and G.H. Hardy might strive to place their displaced colleagues; yet as Nathan Reingold has documented for mathematicians, the reception of the foreigners in the US was often frosty. In addition, Dehn was over 60 years old, and presumably not accustomed to American students.

On October 27, 1934, Karl Menger wrote Oswald Veblen (in English):

> *What I could not write you from Vienna is a description of the situation there. You know how fond I am of Vienna .... But the moment has come when I am forced to say: I hardly can stand it longer. First of all the situation at the university is as unpleasant as possible. Whereas I still don't believe that Austria has more than 45% Nazis, the percentage at the university is certainly 75% and among the mathematicians I have to do with, except, of course, some pupils of mine, not far from 100%.*

This was written after the failed July 1934 Nazi *putsch* in Austria, during which the Austrian chancellor, Engelbert Dollfuss, was killed. Menger in fact had been in Vienna since 1927, when, after two years with Brouwer, he succeeded to Kurt Reidemeister's chair, Reidemeister having gone to Königsberg. Menger also speaks of how an assistant of his, whom he regarded as "one of his best friends", went home to Germany for a visit, and, to Menger's amazement, returned full of enthusiasm for Hitler. In 1937 Menger took a job at Notre Dame, officially on leave from Vienna. With Hitler's *Anschluss* of Austria (March 12–13, 1938), Menger resigned his professorship in Vienna; in 1942 he moved to Chicago and the Illinois Institute of Technology, where he stayed for the rest of his life.

While Felix Hausdorff was not, properly speaking, an algebraic topologist, his work seems fundamental to the appropriate establishment of topology as a mathematical subdiscipline. Some people are surprised to learn that Hausdorff's early work was astronomical in character, and, of course, he made significant contributions to areas of mathematics other than set theory. He also (like later, Hans Freudenthal) was devoted to literary production and published poems and other literary works often with a Nietzschean flavor under the pseudonym Paul Mongré. What deserves brief mention here is his tragic fate. On November 7, 1934, Hausdorff took the new civil service oath sworn to Adolf Hitler. All civil servants still in office (including Jews) were required to take this oath; however, on March 5, 1935, he was dismissed under the same new law which affected Max Dehn. Though not religious, Hausdorff had never denied his Jewish origins, nor had he ever been baptized, though his wife had long before converted to the Lutheran religion. He stayed in Germany as conditions for "non-Aryans" became progressively worse. The only Bonn mathematician who maintained contact with Hausdorff after his forced emeritization was Erich Bessel-Hagen, a student of Carathéodory, also a classical scholar, and very interested in the history of mathematics as well. Bessel-Hagen, an analyst and historian, who was also lame, does not seem to have gotten on well with at least some topologists, as he is somewhat cruelly caricatured on p. 151 of Bela Kerékjártó's 1923 topology text. The Hausdorffs were threatened several times with internment and deportation. This managed to be staved off until an order finally came in mid January, 1942 that they would be interned in Endenich, a suburb of Bonn. This was in fact preparatory to deportation to Theresienstadt. On Sunday, January 25, 1942, Hausdorff wrote his last letter to his friend the Jewish lawyer, Hans Wollstein (who would himself be deported several months later and die in Auschwitz). There is not space here to quote this, except to remark that typical of Hausdorff, is the poignant pun it contains. Wollstein had tried to convince the Hausdorffs that being interned in Endenich was perhaps bearable. Hausdorff writes:

"... auch Endenich
ist noch vielleicht das Ende nich."

That is, "also Endenich is still perhaps not the end", and leaves the "t" off the German word "nicht" to indicate expressly that he is punning.

After writing the letter, which announced his suicide, Hausdorff, his wife, and her sister (also a Lutheran convert) took lethal doses of barbiturates. In 1948, Hindenburgstrasse, on which the Hausdorffs had lived, was renamed Hausdorffstrasse.

Most topologists, however, did not have to leave Germany or suffer tragedy. Furthermore, despite its somewhat uncertain status from a purely ideological Nazi point of view, algebraic topology did get published during the Hitler years. A survey (carried out by Beata Smarcynzska at the University of Rochester) has been made of the three major German mathematics journals: *Mathematische Annalen*, the *Journal für die Reine und Angewandte Mathematik ("Crelle")* and *Mathematische Zeitschrift* for the years 1933–1944, as well as of *Deutsche Mathematik* for the years 1936–1944. *Deutsche Mathematik* was the journal, first appearing in 1936, which was founded by Bieberbach originally to promote a "truly Aryan" mathematics. After the first two volumes, however, except for occasional excrescences, it settled down to being just another mathematics journal. While crude classification of mathematics papers is sometimes very difficult, nevertheless it is apparent that the three leading journals steadily published topological papers; this was especially true of *Mathematische Annalen*. In fact, their content hardly seems overtly tainted by the political atmosphere in which they were appearing. This was at least partly because the three chief editors: Erich Hecke (*Annalen*), Konrad Knopp (*Zeitschrift*), and Helmut Hasse ("*Crelle*"), whatever their different political views and differing mathematical specialties, were all intent on keeping mathematics free from political interference. On the other hand, Bieberbach's *Deutsche Mathematik* published exactly one topological article in those years, a paper on curves by one Erich George, about whom I know nothing more, except that apparently he did not have a university career. On the other hand, *Deutsche Mathematik* published an overwhelming amount of nontopological geometry, even slightly more pages than in analysis. Nontopological geometry appeared in the other three journals as well, and indeed more than the newer subject of topology, but far less than algebra or analysis. *Prima facie* journal content alone would seem to indicate that some mathematics was more suitable to Nazi views than other mathematics. This is not to demean *Deutsche Mathematik*. In addition to the famous papers of Oswald Teichmüller, it published other quite creditable mathematics, especially from volume 3 onward.

If most topologists stayed in Germany during the Hitler years and were also enabled to publish, the atmosphere in which they worked is of more than passing interest. I should like to tell three brief stories involving topologists which reveal that atmosphere and only one of which has received mention previously in the literature. Kurt Reidemeister was at first more interested in philosophy than mathematics. Born in 1893, as a nineteen-year-old he heard Edmund Husserl lecture in Freiburg. His studies there and elsewhere were interrupted by World War I, and after that war (in which he rose to lieutenant), he returned to university in Göttingen where he qualified simultaneously as a secondary school teacher in mathematics, philosophy, physics, chemistry, and geology! In 1920, he followed Erich Hecke to Hamburg, completing a dissertation in algebraic number theory in less than an additional year. At Hamburg, Wilhelm Blaschke turned Reidemeister's attention towards geometry, and he was collaborator on Blaschke's Differential Geometry. While a brilliant mathematics student, he continued to pursue his philosophical interests, lectured on Oswald Spengler's "Decline of the West", and wrote stories and poems. Thus he had a breadth of interests analogous to that already mentioned for Hausdorff and Freudenthal. Although not yet "habilitated", two years after following Hecke to Hamburg, he received a call to a professorship in Vienna. Not only did he pursue mathematics in Vienna, but philosophy as well, becoming part of the famous "Vienna Circle" founded by Moritz Schlick. When Reidemeister moved to Königsberg, his successor at Vienna, Karl Menger, engaged in similar philosophical pursuits as well. It has while he was at Königsberg that Reidemeister's well-know books on combinatorial topology and knot theory appeared. In January 1933, shortly before Hitler's becoming chancellor, Nazi students at Königsberg fomented a disturbance against the university *Rektor*. Apparently, Reidemeister devoted a whole mathematical lecture to explaining why such student behavior was irrational and totally unsupportable. As a result, he was dismissed shortly after January 30 at a time when three "non-Aryan" colleagues in mathematics, Gabor Szegö, Richard Brauer, and Werner Rogosinski, were all left in office (until after the law of April 7 of that year, "reforming the civil service"). Wilhelm Blaschke, who had been Reidemeister's mentor, circulated a petition for Reidemeister's retention and attempted to find him another job. In autumn 1934, when Hasse went to Göttingen, Reidemeister succeeded him at Marburg. In 1946, Reidemeister wrote a lengthy manuscript in German on "The Freedom of Science" (a copy is in the Veblen papers at the Library of Congress). Here he comments (my translation):

> "... *The validity of science for the general public of the German state was destroyed in May 1933, the administration of institutions became drawn along in sympathy; however, the inner scientific public remained, seen as a whole, intact, and a struggle took place around it, which was carried through with great tenacity and up until the last hours of the Hitler state. The success of this struggle, the wisdom and bravery, which stood the test here, is that in which the German universities preserved their inner essence* (Kern) *and saved the possibility of their flourishing once more when the time of fearful destruction was over.*

A second incident involved William Threlfall and Herbert Seifert. Its story is told in the Hecke correspondence preserved at his death, and maintained by Rotraut Stanik at the university at Hamburg. If shows the triviality which might excite concern in the Nazi atmosphere. The principals are Threlfall (his first name was actually William – his mother was English); Erich Hecke (who rejected the Nazis, a fact, says Bruno Schoeneberg in a memorial notice, known even to the German authorities) and Wilhelm Blaschke. The behavior of Blaschke, a great mathematician and very cosmopolitan man, during the Nazi years, can perhaps best be described as cynical opportunism. He was a Nazi fellow-traveller both for personal self-aggrandizement and, as he believed, to the benefit of his department in Hamburg. For example, he propagandized for the Nazis without necessarily believing the National Socialist doctrine. In some ways he was the traditional German professor who, in Max Weber's words, "sings the song of him whose bread I eat".

After the success of Seifert and Threlfall's famous textbook, they decided to write a monograph on "*Global Calculus of Variations: Morse Theory*". This was accepted by Blaschke as a publication in a series of "Hamburg monographs" (it was to be #24). The epigraph was in Latin and the opening words of Kepler's *Introduction* to the *Astronomia Nova*, running about 20 short lines. The first sentence begins with how most difficult it is "today" (*hodie*) to write mathematical books. On May 22, 1938, Blaschke wrote Threlfall that the epigraph could lead to misunderstandings and so, without asking the authors, had instructed the publisher to suppress it. Threlfall wrote Hecke, Blaschke's co-editor of the series, saying he would not give in on this matter and that any reference to contemporary times was only to the fact that mathematicians should not be too seriously upset if a textbook had appeared in which an octahedron had appeared as a cube *en point*, since it was as difficult in 1609 as in 1938 to write mathematical books which were simultaneously rigorous and intelligible, and at all epochs, the people who knew how things were, were few, and it took effort to uphold the level of their science.

The letter suggests various alternatives, all of them maintaining the epigraph if Threlfall is to be associated with the book. It ends, however, as follows:

> *What concerns me is that I would die of home sickness if I had to leave Germany, something which can easily happen if Blaschke withdraws his protection from me.* [Hand von mir zurückzieht] *Granted that has no influence on my position in the struggle over the epigraph.*

This ominous-sounding last paragraph indicates the power which by then Blaschke had amassed in academic matters.

Though Threlfall had written Hecke on May 30, when Hecke had just returned from a trip to the US, the accumulated work during Hecke's absence prevented him from answering for two weeks. On the very day he was answering Threlfall, Threlfall was writing Blaschke again. A few days earlier, Blaschke had written him a letter which apparently warned Threlfall about possible consequences if he persisted in the Kepler epigraph. Threlfall replies (emphasis in original):

> [My opinion] *is that the epigraph indeed can be pernicious* to my person – *who is able to predict such a thing today! – and I thank you once again that you have made me aware of this possibility and warned me in a friendly manner. However one may also interpret the epigraph, it will only extend both internally and in foreign countries to the honor and profit of* German Science. *Never yet have I allowed personal considerations to be decisive in questions of science and will not do so now.*

Hecke to Threlfall on the same day (June 15) says:

> *If one is very careful and anxious – or must be so, then perhaps one can have second thoughts about some places in the epigraph. (I personally have none.)*

He then suggests that striking the word "*hodie*" would remove all dangers and he has suggested this to Blaschke who then told him that he had just written Threlfall. Blaschke also said to Hecke that he did not believe Threlfall would be so stubborn as previously (but as seen above, Blaschke was wrong). Hecke also worries about the suggestion at the end of Threlfall's letter, and warns him that Blaschke is a man of unaccountable impulses, who is more influenced by a momentary frame of mind than rational deliberations, and that:

> *Since he has made application to be a member of the* [*National Socialist*] *party, and probably will shortly be one, now does everything in order not to excite any offense and to rise as quickly as possible.*

A few days later Hecke reemphasized that he had nothing against printing the citation from Kepler.

Threlfall's reply was grateful for Hecke's support, but deplored his suggestion of striking "*hodie*" to satisfy Blaschke. Things dragged on with further correspondence, also with the publisher, Teubner, but finally on November 28, 1938 (six months after Blaschke's original complaint), Threlfall could announce to Hecke that Blaschke had finally given in, and that he hoped Blaschke would now leave him in peace. However, a letter from Blaschke to Seifert "foaming with rage [*wutschnaubend*]" which ended the correspondence, did not make him sanguine. He finds Blaschke completely ununderstandable and entertains the idea that perhaps Blaschke is against him because "... I agreed to his pressing request to tell him my real judgment on his lecture in Holland". He comments that he had previously rejected all warnings about Blaschke; until he himself had had bad experiences.

The book appeared shortly after this letter, with the complete epigraph from Kepler, including "*hodie*".

This all seems so absurdly trivial. But it was not in 1938 in Germany. A few final contextual comments are in order. At the time Threlfall was awaiting official appointment as Carl Ludwig Siegel's successor in Frankfurt (Siegel had gone to Göttingen). This did eventually come through, effective for 1938, though by late November of that year, it still had not. Blaschke and Hecke not only had quite different political views, but they were personal enemies as well, and never spoke about anything but mathematical matters. Nevertheless, they managed to work together to lead the Hamburg Mathematics department (with first Johannes Radon, and then, most significantly, Emil Artin, as a third) for twenty-five years. In Threlfall's letters, the close to Hecke is always "Most sincerely" or an equivalent expression; to Blaschke, it is always the officially asked for "Heil Hitler". Generally, one only failed to use "Heil Hitler" to people whose opinion one was sure of. Hecke's letters to Threlfall (and others) have similar conventional closes. When World War II broke out, Seifert managed to get transferred (officially on leave from Heidelberg) to a research position at the air force research installation in Braunschweig, from winter semester 1939/40 through winter semester 1944/45. This position removed him from being subject to the military draft, vitally important since he was only 32 in 1939. After the war, he returned to Heidelberg as a full professor and rebuilt mathematics there. In 1946, Threlfall, then at Frankfurt, joined him in Heidelberg, but died suddenly in 1949 at the age of 61.

The third and last "atmospheric incident" involving topologists to be mentioned here illustrates the direct pressure put on traditional academic values by the Nazi government. It also shows how a new young faculty member could become a political pawn through simply trying to do what he thought was necessary in those times to obtain a position.

During the Nazi period in Germany, the full professors of mathematics at the University of Munich were (with year of appointment) Oskar Perron (1923), Constantin Carathéodory (1924), Heinrich Tietze (1925). There had been a fourth, the early analyst of several complex variables, Friedrich Hartogs, appointed in 1927 and compulsorily dismissed as Jewish in 1935. Hartogs committed suicide in 1943. None of these three full professors had any sympathy for the Nazis, though Carathéodory, born of Greek parents in Berlin, his father being a diplomat in the service of Turkey, kept his opinions rather to himself (during World War II, Greece was an enemy state of Germany, despite its best efforts to stay neutral, and Turkey was neutral). Tietze and Perron, however, did not fail to be outspoken about academic matters and the attempted intrusion of political chicanery into mathematics. On the other hand, the academic administration at Munich in those days was another matter entirely. Munich, for the Nazis, the "chief city of the movement", was important to them as a showplace in academic as well as other matters. When the famous mathematical physicist, Arnold Sommerfeld, retired, the issue of who should replace him became a lengthy and interesting story. The upshot, though, was Wilhelm Müller, who was essentially forced on the faculty by the German education ministry. Müller's expertise was in classical mechanics, and at the time was in the mathematics department at Aachen (where he had been Theodor von Kármán's successor). He was also an ardent and vigorous proponent of "Aryan physics". Unqualified as a physicist to succeed Sommerfeld, his political *bona fides* carried him through (as they no doubt had already in Aachen), and he eventually became *Dekan* (roughly Dean) for natural and mathematical sciences.

One of the things Tietze and Perron struggled against was the imposition on the mathematics and natural science faculty by Müller of politically respectable incompetents; this went on on more than one occasion. A rather interesting case was that of Eduard May. May was born in 1905 in Mainz, and in 1923 matriculated at Frankfurt in the natural science faculty, studied various sciences and mathematics, but soon gave up on mathematics because he says in his *curriculum vitae* of November 21, 1941, he "was not able to follow the leading Frankfurt mathematicians of the time, Dehn and Hellinger" (for the Nazis, Dehn and Hellinger were both Jews). May later concentrated his attention on botany and especially zoology, and received a doctorate in 1928 with a dissertation on mollusks. He failed to get a start on a university career (times *were* hard) and went to work for a chemical factory as a scientific advisor. In 1931, he married and moved to Göttingen to be nearer the firm employing him, at the same time, taking classes at the university and using its library. In particular, not only did this result in zoological publications, but he educated himself in mathematics, physics, and philosophy, becoming convinced he says of the necessity of a thoroughgoing study of the region between philosophy and natural science. He won an essay prize in 1934 which brought him to the attention of a Nazi general science journal. He published several articles therein, remarking that "with respect to the reform and foundation of science in the Indogermanic spirit" the journal followed "the same tendencies as I do". Various vicissitudes along the way, including more philosophical essay prizes, brought him to the attention of "scholars in Munich", and he moved to Munich. One of these essays was a monograph published in early 1941 entitled "At the Abyss of Relativism". At Munich he pursued study of the "major philosophical and epistemological problems with especial attention to the racial and national [*völkisch*] point of view".

One month after May submitted this c.v., Wilhelm Müller "warmly approved using" "At the Abyss of Relativism" as the paper for his "Habilitation" as a faculty member in Munich. "Habilitation" was the way one started on a German academic career: it consisted of an original essay in a subject matter, an oral examination on that essay, and usually then a sample lecture. All candidates for "Habilitation" necessarily already had a Ph.D. Müller's statement explaining the importance of May's work contains fulminations against "relativism" and "empiricism" citing May's work as a turning point "in the scientific thinking of our time". A second reader, the philosopher Dirlmeyer, also a Nazi party member, chimed in as to the importance of May's work, not failing to mention "the coming European cultural tasks of Germany". However, the acceptance of this book for May's original essay for "Habilitation" depended upon the faculty of natural science and mathematics. Perron and Tietze led the opposition to its acceptance. On January 23, 1942, Müller wrote a riposte to their opposition saying it was to be expected, however, in his opinion, represented a completely false understanding "of the ideological struggle of the present which burned even more strongly than ever exactly in the area of natural science". Their opposition proved to him how important it was to have May as a teacher. Exactly one month later, the Munich physical chemist, K. Clusius, also explicitly joined the opposition to May. At the same time, however, Müller scheduled the oral examination. Clusius complained that not all senior faculty members had yet given their judgment of the essay as they were supposed to prior to the scheduling of the oral. Tietze and Perron complained that they had investigated and discovered that several of their colleagues had also rejected the essay, however, some had not even received it (!). Müller put off May's oral from February 27 to March 13, 1942. It seems clear that, as more faculty members were opting for rejection, the habilitation procedures were accelerated to prevent that. Even the additional two weeks seemed hardly adequate time (since not all had received the essay). Actually, of those faculty giving an opinion on May's essay, ten rejected it (including two Nazi party members), four abstained (including three Nazi party members), and six voted for its acceptance (all party members). Present at the oral examination which Müller nevertheless pushed through, was the Nazi *Rektor* of the University, Walter Wüst, an "Indogermanic" scholar, also a highly placed SS man, who sometime earlier had been interested in the "Jewish influence in mathematics". Wüst's opinion was that because May's doctorate was in natural science, he had to become accepted in the natural science and mathematics faculty. Wüst's presence at the examination was certainly unusual, nevertheless, not only the likes of Tietze and Perron fought the habilitation at the oral examination, but even one of the Nazi party members who voted for the essay's rejection. Nevertheless, Müller pushed things through, though the discussion seems to have been lively. The faculty as a whole discussed the matter on April 29, accepted May, and his trial lectures were scheduled. The local leader of the *Dozentenschaft* (the Nazi organization for university teachers) in approving the scheduling spoke explicitly of the "inimical behavior of Messrs. Tietze and Perron" on April 29.

May held three lectures on June 1, 4, 5, 1942 whose subject was "Description and criticism of the logical and epistemological bases of modern theoretical physics". Apparently the faculty vote was fourteen to four *against* May's receiving a teaching position (the fourteen including four party members). Nevertheless, Müller pushed him through, full of praise of his abilities. Consequently, May was given the right to teach "History and Methodology of Natural Science" in the natural science and mathematics faculty, provided, of course, the necessary proofs that he and his wife were both Aryan were provided. This was done on June 9, and though they apparently were slightly incomplete, May taught at Munich during 1942–1944. May was immune from the military draft because of chronic otitis in both ears and in 1943, received a contract for entomological research on insecticides, this being civilian service in the interest of the war's prosecution. This was under the auspices of the police and the *Waffen-SS*, and May's laboratory began work in June 1944. After the war, May admitted that the only reason he was admitted to an oral examination was because of the *Führerprinzip*, and explicitly mentions Perron and Tietze as the people who got a majority of the faculty to vote against his being so admitted. After the war also (Clusius was a university administrator at the time), the faculty refused to permit May to teach any longer in Munich because he had been forced on them as an exponent of national socialism. Ironically, though the leaders of the Nazi element in the 1942 faculty thought of May "as an old national socialist" and Clusius thought he must be, in fact apparently he was never a member of the party, nor even an applicant for party membership. It should be made clear that the faculty's post-war decision made no judgment of May's actual abilities, he was dismissed because Nazi officials had forced him on the faculty. May eventually ended up teaching history of science at the Free University in Berlin but not until 1951, and died in 1956 shortly after his 51st birthday.

May seems to have been no Nazi, but merely an opportunist trying to make the best academic career he could under the conditions of the Third Reich. Tietze and Perron (they often acted together in such matters) would fight against such pseudo-science infiltrating the scientific faculty whenever they could. May's was not the only such case, but it is the one for which I have the best documentation.

What permitted such "insubordination" among full professors? Another example of Tietze's open attitude occurred in late September 1938. Asked in a questionnaire about his military connections, he replied sarcastically:

> *In September 1937, I received an appointment as substitute air raid warden. I possess no auto, horse, available land, gasoline station or canned goods factory. For a more precise answer, I lack explanation of the new to me and difficult to understand word "Bedarfsträger" (typographical error?). [The neologism* Bedarfsträger *roughly translates as "consumer".]*

This earned him the following report from the university *Rektor*:

> *Since this remark represents a gross impropriety both in content and tone, I reject it most sharply and make you aware that a repetition of such childishness could be accompanied by the most unpleasant consequences for you.*

To show open disagreement with Nazi officialdom was never without peril, but actually the mechanisms of the state cared little about academe, provided Jews and political opponents were eliminated (Max Zorn, for example, had to emigrate because he was a youthful communist). Nazi ideologues could (and did) complain of the nonreception of their message by many mathematicians. Yet a number of prominent German mathematicians were Nazi fellow-travelers of one sort or another, ranging from ideologues like Bieberbach and Teichmüller to cynics like Blaschke.

This was even true for some non-Germans. For example, L.E.J. Brouwer was offered a professorship by the Nazis, which he seems to have actually seriously considered before rejecting it (he had earlier rejected a position proferred by the Weimar government). Brouwer also wrote an encomium for Vahlen's 70th birthday (1939). Nor was he the only non-German topologist who seems to have had similar Nazi leanings: the Norwegian, Poul Heegaard (Dehn's collaborator), was another.

Significant algebraic topology was certainly done during the Nazi regime, even though it would probably not have been the mathematics advocated by the likes of Bieberbach or Vahlen. (In any case, Bieberbach was essentially a complex analyst with strong interests in certain algebraic structures, and Vahlen a number theorist and geometer turned applied mathematician.) Algebraic topology and most mathematics *qua* mathematics had no difficulties under the Nazis because the mechanisms of the state did not care about academic disciplines. They did care that Jews were expelled or transported and serious open resistance was suppressed – that is, they cared about who were mathematicians, not what was mathematics. Somewhat surprisingly, not even applied mathematics was fostered or even prevented from decay in a Germany either rearming or at war. The reactions of Nazi officialdom were always uncertain; it took courage for Tietze to act as he did in both the cited instances as well as others. While there were academic Nazi true believers like Müller and Bieberbach and many others who would agitate for a sort of national Aryanism in science, that was not important to the political powers. What counted for them was the assimilation of all German institutions, and the elimination of Jews and other undesirables. The Nazi ideology among academics did great service in producing this; but after about 1936, their activities became less and less important to the powers that were, though, of course, as true believers, they would carry on as before.

Bibliography

Artzy, R. (1973), *Kurt Reidemeister, 13.10.1893–8.7.1971*, Jahresbericht der DMV **74**, 96–104.

Bessel-Hagen, E., *Nachlass*, as in University Archive in Bonn.

Beyerchen, A.D. (1977), *Scientists Under Hitler*, Yale Univ. Press.

Bieberbach Correspondence deposited by Niels Jacob in Niedersächsische Staats-und Universitätsbibliothek, Göttingen.

Biermann, K. (1988), *Die Mathematik und Ihre Dozenten*, Akademie-Verlag, Berlin.

Brouwer, L.E.J. (1976), *Collected Works*, Vol. 2., North-Holland, Amsterdam and Oxford; (life by Hans Freudenthal and Arend Heyting, pp. x–xv.)

Dierksmann, M. (1967), *Felix Hausdorff Ein Lebensbild*, Jahresbericht der DMV **69**, 51–54.

Dieudonné, J.A. (1989), *History of Algebraic and Differential Topology, 1900–1960*, Birkhäuser.

Fischer, H.J. (1984), *Von der Wissenschaft zum Sicherheitsdienst*, Quellenstudien der Zeitgeschichtlichen Forschungsstelle Ingolstadt (Band 3).

Freudenthal, H. (1987), *Berlin 1923–1930*, Walter de Gruyter, Berlin.

Hecke, E., *Hecke correspondence in the care of Frau Rotraut Stanik at the University of Hamburg*, unpublished.

Henn, H.-W. and Puppe, D. (1990), *Algebraische Topologie*, in *Ein Jahrhundert Mathematik 1890–1990*, Friedr. Vieweg u. Sohn, Braunschweig/Wiesbaden.

Kass, S. (1996), *Karl Menger*, Notices Amer. Math. Soc. 558–561.

May, E., *Personalakten in the archive of the University of Munich.*

Menger K., *My Memories of L.E.J. Brouwer*, K. Menger, Selected Papers in Logic and Foundations, Didactics, Economics, D. Reidel, Dordrecht, Holland, 237–258.

Neuenschwander, E. (January 1992), *Felix Hausdorff's Letzte Lebensjahre nach Dokumenten aus dem Bessel-Hagen-Nachlass*, Technische Hochschule Darmstadt, Preprint #1446.

Pinl, M., *Kollegen in einer dunklen Zeit*, Jahresbericht der DMV **71** (1969) 167–228 (Teil I); **72** (1971) 165–189 (Teil II); **73** (1972) 153–208 (Teil III).

Pinl, M. and Furtmüller, L. (1973), *Mathematicians under Hitler*, Leo Baeck Yearbook XVIII, 129–182.

Reingold, N. (1981), *Refugee mathematicians in the United States of America 1933–1941: reception and reaction*, Annals of Science **38**, 313–338; also reprinted in *A Century of Mathematics in America*, Part I (edited by P. Duren, R. Askey and U. Merzbach), Amer. Math. Soc., Providence, RI, 1988, 175–200.

Schappacher, N. (with M. Kneser) (1990), *Fachverband-Institut-Staat* in *Ein Jahrhundert Mathematik 1890–1990*, Friedr. Vieweg u. Sohn, Braunschweig/Wiesbaden.

Scharlau, W. (1990), *Mathematische Institute in Deutschland 1800–1945*, Friedr. Vieweg u. Sohn, Braunschweig/Wiesbaden.

Shils, E. (1974) (ed. and trans.), *Max Weber on Universities*, Chicago.

Siegel, C.L. (1966), *Zur Geschichte des Frankfurter mathematischen Seminars*, Gesammelte Abhandlungen, Vol. III, Springer, Berlin, 462–474.

Tietze, H., *Personalakten in archive of the University of Munich.*

Toepell, M. (1996), *Mathematiker und Mathematik an der Universität München*, Munich (#19 of the series Algorismus).

Veblen, O., *Correspondence and Papers held at the Library of Congress*, Washington, DC (donated by his widow).

This Page Intentionally Left Blank

CHAPTER 31

The Japanese School of Topology

Mamoru Mimura

Department of Mathematics, Faculty of Science, Okayama University, Okayama 700-8530, Japan
E-mail: mimura@math.okayama-u.ac.jp

Sources

The present manuscript is essentially based on the following articles written in Japanese, in particular on the first one [C], which was compiled on the occasion of the Centenary of the Mathematical Society of Japan in 1977 by a committee of this Society, presided over by Atuo Komatu. The compilation lasted several years and the first edition of the book is published in two volumes, Vol. I (1983) and Vol. II (1984). Chronologically, this book was arranged according to the Imperial eras, namely Meiji 1868–1912, Taisho 1912–1925, and Showa 1925–1989. (Since 1989 we have been in the Heisei era. We shall also use occasionally in our text these names of Imperial eras.) The contents of the book are limited to the period before around 1970. Volume I contains the following four chapters: 1. Before Meiji, 2. The first half of Meiji (1868–1890), 3. The second half of Meiji (1890–1910), 4. Taisho (1910–1925). Volume II contains two chapters: 5. The first half of Showa (1925–1945), 6. The second half of Showa (1945–1970). Each chapter describes (i) the historical background, (ii) the establishment of institutions and societies and their activities, (iii) publications, (iv) lives of leading scholars, and (v) the development of each branch of mathematics. The article [MM] is essentially a continuation of the topology section of [C], that is, developments after 1970. The book [NK] is a volume concerning the mathematical sciences out of a 25 volume history of the development of science and engineering in Japan, and [N] treats the development of geometry in Japan. The articles [K1, K2, T, F] are personal essays which appeared in the monthly journal "Sugaku Seminar" ("Mathematics Seminar").

[C] *The Centennial History of Mathematics in Japan*, Vols I, II, Committee of the Centennial History of Mathematics in Japan, Iwanami (1984).

[MM] Y. Matsumoto and S. Morita, *Topology of manifolds: Current mathematics in Japan*, Sugaku, Iwanami-Shoten **25** (1973), 64–67.

[NK] Nippon Kagaku Gizyutusi Taikei, Mathematical Sciences **12**, Daiichi Hohki Shuppan (1969).

[N] H. Noguchi, *The World of Geometry*, Nihon Hyoron-sha (1972).

[K1] A. Komatu, *The Department of Mathematics at the University of Tokyo*, 40 *years ago*, Sugaku Seminar (September 1969), 35–37.

[K2] A. Komatu, *On topology: the development of its idea*, Sugaku Seminar (August 1979), 2–7.

[T] H. Terasaka, *Memories of K. Nakamura*, Sugaku Seminar (March 1987), 36–37.

[F] H. Fukaishi, *Another topology in Japan*, Sugaku Seminar (July 1987), 34–35.

HISTORY OF TOPOLOGY
Edited by I.M. James

We have not included a section on the development of general topology in Japan after World War II, since there is a good article for this:

> J. Nagata, *The Flowering of General Topology in Japan*, Handbook of the History of General Topology, Vol. 1, Kluwer Academic Publishers, Dordrecht (1997), 181–241.

1. Introduction

The Tokyo Mathematical Society was founded in 1877, which was an epoch-making event for the mathematical world in Japan together with the adoption of western style calculation in the school system and the foundation of the University of Tokyo in the same year. In June 1884 the name of the society was changed to the Tokyo Physico-Mathematical Society. In May 1919 it was further changed to the Physico-Mathematical Society of Japan, but this society was disbanded in December 1945 to form two new societies: the Mathematical Society of Japan and the Physical Society of Japan. In June 1946 the Mathematical Society of Japan was founded.

During this period 1877–1945 six further (ex-Imperial) universities with departments of mathematics were founded:

> Kyoto 1897, Tôhoku (Sendai) 1907, Hokkaido (Sapporo) 1930, Osaka 1931, Kyushu (Fukuoka) 1939, Nagoya 1942;

in addition two other universities (called Higher Normal Schools at that time, later renamed Bunrika Daigaku) with departments of mathematics were founded in Tokyo and Hiroshima in 1929.

Thus the University of Tokyo is the oldest, and the teaching staff of other universities were recruited initially from among the graduates of the University of Tokyo. (Incidentally, the name of this university has changed several times: Imperial University, Tokyo Imperial University, etc.) But soon each of these universities created its own school with a distinct character. One might say that these nine universities have been the centres of development of mathematics in Japan. They had their own journals of mathematics to publish the works of their members and others:

> Journal of the Faculty of Science, Imperial University of Tokyo, Section I;
> Memoirs of the College of Science, Kyoto Imperial University, Series A;
> The Science Reports of the Tôhoku Imperial University, First Series (there was another journal published privately by T. Hayashi, "The Tôhoku Mathematical Journal");
> Journal of the Faculty of Science, Hokkaido Imperial University, Series I;
> Collected Papers from the Faculty of Sciences, Osaka Imperial University;
> Memoirs of the Faculty of Science, Kyushu Imperial University, Series A;
> Collected Papers from the Mathematical Institute, Faculty of Science, Nagoya Imperial University;
> Science Reports of the Tokyo Bunrika Daigaku;
> Journal of Science of Hiroshima University, Series A.

Kunugui, Terasaka and Komatu graduated from Tokyo in 1926, 1928 and 1932, respectively. Kunugui joined Hokkaido after having studied in France; Terasaka and Komatu joined Osaka.

Soon after its establishment in 1931, Osaka became the centre of the study of modern mathematics in Japan. (Kunugui also joined Osaka later on, in 1949). The 1930's were the time when topology in its widest sense began to be actively studied in Japan. At the University of Tokyo Iyanaga became a member of the teaching staff in 1935, after coming back from Europe. Although the field of his own research was arithmetic, he taught geometry and topology as well, and had students such as Abe, Kodaira, and (later,) Yoneda, Tamura and Hattori.

2. The early days

Probably the first paper published in Japan related to topology is [92] in which Takeo Wada (1882–1944) defines a simple curve in the n-dimensional Euclidean space from the set theoretical viewpoint and shows that it is a Jordan curve. Wada graduated from Kyoto University and became assistant professor in 1908. He visited the USA, France and Germany from 1917 to 1920; after coming back he was promoted to professor. Wada's main work was in analysis but he was also interested in topology.

C. Jordan constructed a curve as a simple example to express curves analytically in his book

C. Jordan, *Cours d'Analyse I* (1893).

Later in 1903 F. Riesz named the simple curve the Jordan curve, which became famous together with the recognition of the importance of the Jordan theorem. Wada's paper appeared around this time (the beginning of the 20-th century) when there was much interest in the theory of curves.

The basic concepts of topology were being developed by Poincaré, Brouwer and others during this period. As for books in the Taisho era, there were only those of Kerékjárto and Veblen:

S. Kerékjártó, *Vorlesungen über Topologie*, Springer (1923);

O. Veblen, *Analysis Situs*, Colloquium Publ., Amer. Math. Soc. (1922).

It is remarkable that the topology of curves and surfaces had been studied independently and successfully in Japan during this time.

Kunizô Yoneyama (1877–1968), who, though older than Wada, came later to Kyoto University after having taught in middle school, did research in topology under Wada's influence and extended Wada's ideas about curves to higher dimensions in his study [98] of the concept of curvilinear solid surfaces. Furthermore, Yoneyama published in 1917–1920 a lengthy work [99] of more than 300 pages which may be the first treatment of general topology in Japan. This work aims to classify continua (connected perfect sets) in n-dimensional space, and many results about indecomposable continua were obtained in this series of papers. In fact, Rosenthal presented some of Yoneyama's results in his appendix of the "Enzyklopädie". In 1910 L.E.J. Brouwer constructed an example of an indecomposable continuum by dividing the plane into three regions so that these three regions have a common boundary, in

L.E.J. Brouwer, *Zur Analysis Situs*, Math. Ann. **68** (1910), 422–434.

In 1912 Janiszewski introduced the notion of the irreducibility between two points in a continuum after studying Brouwer's result. Around the same time, Yoneyama [99], saying that he was taught by Wada, showed how to construct such an indecomposable continuum by an easier method than that of Brouwer. This is known as the "lakes of Wada". Wada's idea was beautiful and it was praised as "poetic" on p. 143 of the book

J. Hocking and G. Young, *Topology*, Addison-Wesley (1961)

and "particularly attractive" in the commentary on the complete works of Brouwer.

It seems that no more papers on topology appeared after that in the Taisho era, while in the Showa era, the study was carried on by Kôshirô Nakamura, Hidetaka Terasaka and Atuo Komatu. In particular, it could be said that the achievements of Takeo Wada and Kunizô Yoneyama in the Taisho era were taken over by Hidetaka Terasaka.

The idea that topology supports the basic structure of mathematics along with algebra had been widely promoted during this period, and a new word "*isosugaku*" (topology) appeared in Japan. The field included the theory of topological spaces, topological algebra, and analysis by means of topological and algebraic methods (at that time this was called *topological analysis*).

The concept of a topological space was defined by using neighbourhoods in the book

F. Hausdorff, *Grundzüge der Mengenlehre*, Teubner (1914).

Kuratowski gave a version using closure in his paper

C. Kuratowski, *Topologie I*, Fund. Math. **3** (1922); Warszawa (1933)

and described the theory systematically in his book of 1933. Fréchet defined various abstract spaces in his book

M. Fréchet, *Les Espaces Abstraits*, Gauthier-Villars (1928)

and emphasized their importance in the study of analysis. As for algebraic topology, an advanced book

O. Veblen, *Analysis Situs*, Colloquium Publ., Amer. Math. Soc. (1922)

using algebraic methods was published in 1922, and in the next year, a purely geometrical book was published:

S.B. von Kerékjártó, *Vorlesungen über Topologie*, Springer (1923).

We should mention another topologist working during this period; Keitarô Haratomi (1895–1968) of the Toyama High School, who published two papers [9, 10] when there were no Japanese mathematicians working in the theory of general topology other than K. Kunugui who was still working under M. Fréchet in France. (In fact, it was in 1933 that Kunugui wrote his book [38].) Haratomi published a few more papers in the Proceedings of the Physico-Mathematical Society of Japan, the Tôhoku Mathematical Journal and the Japanese Journal of Mathematics. It is remarkable that he was not a university graduate, but learned mathematics all by himself.

3. The period 1925–1945

The book by Seifert and Threlfall,

H. Seifert and W. Threlfall, *Lehrbuch der Topologie*, Teubner (1934)

in which topology was dealt with algebraically, even going as far as manifolds, was published in 1934. The influential book by Alexandroff and Hopf

P. Alexandroff and H. Hopf, *Topologie I*, Springer (1935)

which described both general and algebraic topology was published in 1935.

In 1929 Kôshirô Nakamura (1901–1985), then at Tokyo Bunrika University, visited Berlin University and studied algebraic topology under H. Hopf. Then he visited Switzerland for one year, following Hopf who had moved to Zurich. Thus, fortunately for him and for Japanese topologists later, he encountered a new branch of mathematics in its development. He came back to Japan in 1932 and published two books [52, 53] to introduce algebraic topology to Japan. This was when A. Komatu was in his 3-rd year at the University of Tokyo. (Nakamura later became interested in the history of mathematics.)

Kinjirô Kunugui (1903–1975) went to France in 1928 to study in Strasbourg, and then in Paris under Fréchet, and he obtained the degree of national doctor in France with the paper *Sur la théorie des nombres de dimensions, Thèse* (1930) which discussed the relation between the dimensional type of Fréchet and the dimensional number of Menger. That is, he introduced the concept of "the dimensional class", and obtained a relation with the dimensional type of Fréchet, and by finding a relation with the theory of Menger he succeeded in relating the two theories of the dimension. In another paper [37] Kunugui found a condition that an analogy of Baire's theorem holds in the U-space of Fréchet. After coming back to Japan, he published the book [38]. Subsequently, he changed his area of study to the descriptive set theory, and then, to the theory of functions.

Weil defined uniform spaces in his book

A. Weil, *Sur les Espaces á Structure Uniforme et sur la Topologie Générale*, Actualités Sci. Ind. (1938).

Kiiti Morita (1915–1995) discussed in 1940 [49] the dimension of a uniform space and a compact space. This is the first paper in Japan on uniform spaces. After World War II, under the influence of the fundamental work of Hurewicz and Wallman

W. Hurewicz and H. Wallman, *Dimension Theory*, Princeton Univ. Press (1941)

Kiiti Morita, Jun-iti Nagata (1925–), Keiō Nagami (1925–) and others made major contributions to the development of dimension theory.

Descriptive set theory mainly studies subsets of the real continuum which can be characterized descriptively. This theory had been started in 1916 by Luzin and Suslin, and was concerned with effectively constructed sets, such as Borel sets and analytic sets. Results in this field were obtained in Japan only in the first half era of Showa. Many studies were done by Kinjirô Kunugui, Motokiti Kondô (1906–1980), Takeshi Inagaki (1911–1989) et al., mainly at Hokkaido University.

One important concept in this field is "uniformisation". The uniformisation problems had been considered since around 1930. It was known that Borel sets are not necessarily uniformised by Borel sets and that analytic sets are not necessarily uniformised by analytic sets or by complementary analytic sets. Kondô made a significant advance by proving the theorem that a complementary analytic set E can be uniformised by a complementary analytic set U in his study [34]. This result has an important meaning in modern mathematical logic in the wide sense, and even now it is often quoted and still generalised. On the other hand, in 1940 Kunugui obtained the result [39] that if E is a Borel set and if all the cross-sections are compact, then the projection $p(E)$ of E is a Borel set, and that E is uniformised by some Borel set U. Inagaki made a deep study [16] in 1937 on the constituents of the sheaves for determining a zero analytic set. After World War II, Yoemon Sampei, Tosiyuki Tugué et al., followed in this direction.

Concerning "the metrisation problem", which asks for conditions under which a topological space is expressible as a metric space, Motokiti Kondô dealt in 1933 [33] with the case of a topological space on which a topological group acts. Using this result, Shizuo Kakutani (1911–) showed in 1936 that the condition for a topological group to have a one-sided invariant metric is simply that it satisfies the first countability axiom [17].

Shizuo Kakutani, born in 1911, graduated from Tôhoku University in 1934 and became an assistant at the newly founded mathematics department of Osaka University, where he was under the general guidance of T. Shimizu. His 1936 paper in Japanese Journal of Mathematics on Riemann surfaces, which was later to become the main part of his doctoral dissertation, caught the attention of H. Weyl at the Institute for Advanced Study, and he was invited to Princeton in 1940 for two years. There in 1941 he generalised in [18] the Brouwer theorem that a continuous map from the n-dimensional disk into itself has always a fixed point, to a form which can be applied to an existence theorem in the game theory of Von Neumann. This result, called "the fixed point theorem of Kakutani", is not difficult to prove, though it required a skillful formulation. But war broke out between Japan and the USA in December 1941 and he was obliged to return to Japan the following summer; he joined Osaka University with the rank of assistant professor. (He returned to Princeton in 1948 and later moved to Yale.)

H. Terasaka (1904–1994) graduated from the University of Tokyo in 1928 as a student of S. Nakagawa, and went to Germany and Austria in 1933 as a scholarship fellow of the Ministry of Education of the Japanese Government. He spent two years in Vienna, where he attended the seminar of K. Menger, becoming acquainted there with H. Tietze and H. Seifert. He came back to Japan in 1935 to join the newly founded Osaka University as assistant professor. The following year he was promoted to professor of Osaka University, where he gave lectures on projective geometry, topology and lattice theory.

In the 1930's, Terasaka published many papers using the purely geometric methods of the book by Kerékjártó. In the 1930 paper [78] he used the concept of critical domain to give a simple proof of "the translation theorem of the plane" due to Brouwer. This is the theorem that a topological transformation of the plane is a topologically parallel transformation if it preserves direction and if it has no fixed point. In the following year, he considered the problem of dividing a Riemann surface into sheets and, correcting defects in the paper by Radoitchitch, gave an example of a Riemann surface which cannot be divided into sheets even if it has no boundary [79]. After the consideration of transformations of plane $\mathbb{R}^2$ in 1930, he dealt with the case of $\mathbb{R}^n$ in his work [80] of 1938 and showed that there is an essential difference between the cases $n = 2$ and $n \geqslant 3$. If f is a topological orientation-preserving transformation of $\mathbb{R}^n$ without fixed points, then, when $n = 2$, the sequence $x, f(x), f^2(x), \ldots$ for any $x \in \mathbb{R}^n$ diverges according to the theorem of Brouwer. Montgomery proposed the problem of determining whether the same holds in the case $n = 3$. Terasaka proved that if $n \geqslant 3$, then there are x and f such that $\ldots, f^{-1}(x), x, f(x), f^2(x), \ldots$ remains bounded. In this paper, he discussed systems of curves on $\mathbb{R}^n$, and proved the above results. In another paper of the same year, he discussed *Erreichbarkeit* of a 0-dimensional set on $\mathbb{R}^2$. In his paper [81] of the following year, by examining topological properties of a system of curves generalising a system of geodesics on a surface homeomorphic to a sphere, he showed that the irreducible continua of Wada–Brouwer could be constructed by linking finitely many smooth curves among them, that is, half circles with radii between 1 and 2. By applying to general lattices Kuratowski's method of defining a topological space by using the closure operator, Terasaka introduced and developed the theory of topological lattices [82]. One episode concerning this paper illustrates high standards of his research. During the early stages of World War II he submitted the paper to Fundamenta Mathematica; it was accepted but was burnt to ashes in the German invasion of Poland. With the revival of Poland, Fundamenta Mathematica decided to republish the issue which had been destroyed and asked Terasaka to submit his paper again. He declined to do so, answering that the paper was no longer up to date. Thus the volume had only the title of the paper, not the content. Around this time a geometrical study of surfaces with boundary was made by Ken-iti Koseki (1917–1980) of Kyoto University, but this work was done independently of that of Terasaka.

Now we turn to work in the field of algebraic topology. Atuo Komatu (1909–1995) entered the University of Tokyo in 1929. In his 3-rd year he attended the seminar of S. Nakagawa, who was in charge of geometry, choosing problems in *Raumformen* as a topic. At the seminar he read papers such as

H. Hopf, *Die Curvature integra Clifford–Kleinscher Raumformen*, Nachr. Ges. Wiss. Göttingen, Math.-Phys. Klasse (1925), 131–141;

P. Koebe, *Riemannsche Mannigfaltigkeiten und nichteuklidische Raumformen I*, Sitzungs Bericht Preuss. Akad. Wiss. (1927), 164–196;

H. Hopf and W. Rinow, *Über den Begriff der vollständigen differentialgeometrischen Fläche*, Comment. Math. Helvetici **3** (1931), 209–225.

It was around this time that K. Nakamura came back from Europe; he influenced Komatu further in the study of algebraic topology. Komatu entered the graduate school in 1932. He joined Osaka University in 1934 and was promoted to assistant professor in 1938.

In his paper [28] of 1936, Komatu proved "the duality of a covering", using the lemma that the automorphism ring of an Abelian group and its character group are anti-isomorphic as rings. (It was proved in the paper [2] by Makoto Abe in 1940 that this lemma can be generalised to the case of a locally compact Abelian group.) Komatu constructed in his paper [29] of 1937 the anti-homomorphism of Hopf and an isomorphism between Gordon's ring and Alexander's ring. He did this by using U-cycles and O-cycles, which are well known these days, but were new concepts at that time. This was done independently of Freudenthal who announced similar results in the same year. Furthermore in a joint paper [32] with Ryozi Sakata (Shizuma), [1] he solved a problem proposed by Borsuk using a generalisation of the Hopf theorem and the addition theorem of homotopy groups. This paper was important for the techniques used in the proof, such as the addition property of homotopy, and such methods were systematised soon afterwards by Eilenberg. Sakata published a paper [63] on mappings from a compact metric space to a sphere in 1938. In 1941, Komatu developed "a transformation theory of complexes" in his paper [30], and developed "an obstruction theory" by using local coefficients under some algebraic conditions. This became his Ph.D thesis under the supervision of S. Nakagawa, where he discussed the problem of stability presented by Panwitz. By examining axioms for topological spaces in terms of neighbourhoods, convergence, closure, etc., Atuo Komatu investigated properties such as the Hausdorff condition. He also considered the weak topology, and he announced his results in the book [31].

The studies of Atuo Komatu and people around him at Osaka University during this period were the source of later developments, but, in concert with it, there was the work of Makoto Abe (1914–1945) and Kunihiko Kodaira (1915–1997), both of whom were students of S. Iyanaga at the University of Tokyo. As noted in one of Abe's papers, Abe and Kodaira have the following achievements. In his paper [26] of 1939, Kodaira proved that Hopf's extension theorem can be generalised to a compact metric space by using Kuratowski's maps. In the following year, Abe proved the same result [2] by using Freudenthal's

[1] Ryozi Sakata (1916–1985) was adopted by the Shizuma family in 1942, after which he used the name Ryozi Shizuma.

method which compares a compact metric space to a sequence of polyhedra. Besides, in a joint paper [27] in 1940, they proved that a compact, separable, connected Abelian group is homeomorphic to its one-dimensional Betti group with coefficients in the real field mod 1. Having also proved a relation which could be described as a form of the Künneth theorem about Betti groups of a product complex, Abe published it [3] in Japanese. He died in the chaos after the end of World War II.

Comparing the tendencies and achievements in this research field at the University of Tokyo and Osaka University around that time, it could be said that the University of Tokyo was more algebraic, and Osaka University was more geometric. Later in 1949, Komatu moved from Osaka University to Osaka City University, and then to Kyoto University, and the study of topology which he was leading developed more and more. After World War II, in the field of topology, Komatu taught Tatsuzi Kudo, Hiroshi Uehara, Minoru Nakaoka, Hirosi Toda, and others in the algebraic direction, while Hidetaka Terasaka taught Tatsuo Homma, Shin'ichi Kinoshita, and others in a more geometric direction.

4. The Topology Colloquium and its bulletin

One of the activities of the Physico-Mathematical Society of Japan before the end of World War II was the organisation of the topology colloquium. The first meeting of the colloquium was held in April 1936 at the Faculty of Science of the University of Tokyo. It was held on the occasion of the annual meeting of the Physico-Mathematical Society of Japan with the intention of developing further the prosperity of topology. The planning was accomplished by Motokiti Kondô of Hokkaido University, Atuo Komatu of Osaka University and Shizuo Kakutani of Osaka University, all in their twenties, and it was held with the approval of many universities and the Physico-Mathematical Society of Japan.

The chairman of the first meeting was Mitio Nagumo of Osaka University and there were two talks:

Kinjirô Kunugui: *On a problem in point-set theory*;

Kôshirô Nakamura: *Algebraic methods in topology*.

There were about 80 participants, which was remarkable for that time.

The second meeting of the colloquium was held in July 1937, also on the occasion of the annual meeting of the Physico-Mathematical Society of Japan at the Faculty of Science of Hokkaido University. The chairman was Kondô and again there were two talks:

Fumitomo Maeda: *The eigenvalue problem in Hilbert space*;

Atuo Komatu: *Über die Homotopiegruppen*.

There were also colloquium talks and a problem session led by Kondô, Kakutani and Nakamura.

The third meeting of the colloquium was held in April 1938 at the Faculty of Science of the University of Tokyo. The attendance was about 50. The chairman was Komatu and the talks were as follows:

Motokiti Kondô: *On various kinds of point-sets having a remarkable property*;

Hidetaka Terasaka: *The formulation of topology and its application*.

After the talks, Kodaira, Komatu and Kakutani reported on

A problem on cellular space,

A problem on the spherical decomposition,

An impossibility problem on representation of Boolean rings, respectively. At this meeting Tatsujirô Shimizu proposed the publication of a bulletin of the colloquium, which was supported by everyone present. As a result, it became necessary to equip the colloquium formally with regulations, and its name was changed to Isosugaku Danwakai (the Topology Colloquium). It was at this time that the name "Isosugaku" (topology, in Japanese) was chosen for this branch of mathematics. The following people were appointed to the committees:

Editorial committee: Kôshirô Nakamura, Hidetaka Terasaka, Fumitomo Maeda, Kinjirô Kunugui (chief), Shin-ichi Izumi, Shôkichi Iyanaga;

Staff committee: Motokiti Kondô, Atuo Komatu, Shizuo Kakutani.

Kondô at Hokkaido University was in charge of the clerical work such as accounting, while Komatu and Kakutani at Osaka University were in charge of editing work.

The Topology Colloquium gradually became prosperous and Hiroshi Okamura (1905–1948) of Kyoto University and Masuo Hukuhara of Kyushu University joined the editorial committee. The colloquium consisted of 2 or 3 special lectures and several 15 minute lectures, and continued until the 8-th Topology Colloquium in October of 1942. The titles and speakers of the special lectures were as follows:

4-th Colloquium (April 1939; at Kyoto University)
 Shizuo Kakutani: *Markov chains and the ergodic theorem*;
 Shôkichi Iyanaga: *The foundation of general topology*;
 Ryoji Shizuma (Sakata): *Continuous mappings of polyhedra.*
5-th Colloquium (April 1940; at Tokyo Bunri University)
 Shin-ichi Izumi: *Integration theory*;
 Takeshi Inagaki: *The problem of Suslin*;
 Makoto Abe: *Betti groups of a product space.*
6-th Colloquium (April 1941; at Hiroshima Bunri University)
 Kôsaku Yosida: *On normed rings*;
 Masuo Hukuhara: *The existence theorem for fixed-points and its application.*
7-th Colloquium (April 1942; at the University of Tokyo)
 Tadao Tannaka: *Introduction to the Morse theory*;
 Hidegorô Nakano: *Continuous linear functionals on partially ordered modules.*
8-th Colloquium (October 1942; at the University of Tokyo)
 Yukiyosi Kawada: *The lattice-theoretic probability theory*;
 Shizuo Kakutani: *The measure algebra.*

On the occasion of the birth of the Mathematical Society of Japan after World War II the Topology Colloquium was disbanded and replaced by the topology branch of the Mathematical Society of Japan.

Isosugaku (Topology) was the name of the bulletin of the Topology Colloquium and the first number of the first volume was published in October 1938; the editor and publisher was the Topology Colloquium. The title "Isosugaku (Topology)" was temporarily used in the first number and remained afterwards. Let us quote from the explanation of terminology in the first volume.

> ... The terminology "topology" today has various meanings, so when we use the word we have to clarify the meaning by adding adjectives after "topology" such as topology in the wide sense or topology in the narrow sense. This is inconvenient. In our country

> some people use the word "isokikagaku" as the Japanese translation of topology in the narrow sense, but by choosing the word topology (= isosugaku) which has a similar form to this word, we want to accept it as the Japanese translation in the wide sense....

And to show what was included in "isosugaku", the names of the specific fields were given explicitly.

> ... In this bulletin, the terms used were "topology in general", "algebraic topology", "set theory", "theory of abstract spaces", "topological algebra", "topological analysis", but there is no significant reason for that classification....

After that there were explanations of the contents of the above six branches, but let us just quote here the explanation of "set theory".

> ... Purely descriptive set theory and theory of real variable functions directly related to it are central. Analytical set theory, of course, belongs to it. Others are Borel sets, the theory of projective sets, the theory of a family of sets in general, the problem of Baire's function and of implicit functions, various problems on the Kontinuum Hypothese....

One can see from this that a fairly wide range of set theory was considered as topology. Also, in the first issue, the main purpose to publish "Isosugaku (Topology)" was discussed as follows.

> ... In fact, one of the characteristics of mathematics in the 20-th century is to make use of topological methods satisfactorily. ... There are many mathematicians throughout the fields of mathematics who are interested in topology, and who are making tireless efforts to develop mathematics, but the incompleteness of the research network of these people has brought up many obstacles for research.... Now, based on the main purpose of its foundation, we rename the colloquium the Topology Colloquium, and we publish the bulletin "Isosugaku (Topology)" in cooperation with the mathematical departments of universities in Japan and the publisher Teikoku-Shoin....

This bulletin was published twice a year (each volume had 70–80 pages) with large size print. The contents consisted of special lectures at the Topology Colloquium mentioned above and some articles such as the following:

Atuo Komatu: *On 0-Zyklus* (1939);
Motokiti Kondô: *On the parameter representation of sets* (1940);
Shin-ichi Izumi : *On the mean motion* (1940);
Junshirô Higuchi: *On the existence problem of collectives* (1941);
Kôsaku Yosida: *On representation of vector lattice* (1941);
Hidegorô Nakano: *On the abstract spectral theory* (1942);
Shizuo Kakutani: *The lattice and ring of a Banach space* (1943).

These articles were useful for interchanging information between many mathematicians whose interests were different. The themes taken up were also gradually approaching the major problems of that time.

"Isosugaku (Topology)" continued to be published for about 5 years up to number 1 of volume 5, but publication was suspended in February 1943 due to World War II. However, it had greatly encouraged mutual understanding amongst Japanese mathematicians.

5. Developments after the end of World War II

5.1. *Algebraic topology*

In Japan it was about 1950 when the confusion after World War II finally abated and the study of topology became active. Below we survey the history up to the early 1970's of the development and the results of algebraic topology.

The basis for the development of topology, especially algebraic topology, in Japan was constructed by Atuo Komatu. During the war, Komatu continued to study topology at Osaka University together with Ryozi Shizuma (Sakata), but in 1949 he became professor of Osaka City University, and organised a research group in topology by inviting Tatsuzi Kudo (1919–), Minoru Nakaoka (1925–), Katsuhiko Mizuno (1926–), Ichiro Yokota (1926–), Hirosi Toda (1928–), et al. from Osaka University. Hiroshi Uehara (1923–) and Nobuo Shimada (1925–) (supervised by Shizuma at Nagoya University) participated in Komatu's group by obtaining the (Hideki) Yukawa Scholarship which was established at Osaka University at that time. New journals from abroad were available only in the library of American Culture Center then, but many papers on topology appeared in those journals. They had different aspects from those published before World War II and were worthy of the name algebraic topology. Komatu encouraged the members of the group not to adhere to the old references but to read these new papers, from which they developed their research activities starting as early as 1950. The results of this group before about 1953 can be described as follows.

The central subjects of interest to topologists all over the world at that time were the study of the homology of fibre bundles, and the homotopy classification of continuous maps, especially homotopy groups of spheres. The research of Komatu's group was also on these topics; in particular, Kudo studied the Leray spectral sequence in his own way. Obstruction theory for the extension of a continuous map was the research object of Komatu's research during World War II, and it was pursued by Uehara, Shimada and Nakaoka. They studied applications of cohomology operations to the homotopy classification problem and generalised Pontrjagin's theorem and Steenrod's theorem; this attracted the attention of Eilenberg and others during 1951–1952 (see [66, 67] and [55]). As for homotopy groups of spheres, important works at that time, extending the results obtained by their forerunners before World War II, were done by American mathematicians such as G.W. Whitehead, and also H. Cartan and J.-P. Serre in France were engaging in the study based on the application of cohomology of Eilenberg–MacLane spaces. Toda published more detailed results [83] in 1952 which went further. This drew international attention to Japanese achievements in topology.

Around 1950, in addition to the topologists at Osaka City University and Nagoya University, Kiyoshi Aoki (1913–) and Hidekazu Wada (1924–) and others of Tôhoku University were studying topology, as well as other mathematicians in the University of Tokyo such as Nobuo Yoneda (1930–1996).

With the aim of fostering interchanges between these people, Komatu proposed the idea of a symposium for topology, and (he was a very good organiser) the first one was held at the College of Arts and Sciences of the University of Tokyo in 1951. This symposium, the Zen-Nippon (all Japan) Topology Symposium, which encompassed every area of topology, proceeded to Tôhoku University in 1952, and then Osaka City University in 1953, and it has been held once a year until the present day, making a great contribution to the development of topology in Japan. At that time there seemed to be no mathematical symposia other than this one, although these days symposia are held as a matter of daily occurrence.

In the summer of 1953 S. Eilenberg, who was interested in the Japanese school of topology, paid a visit to Japan. He brought with him the galley proofs of "Homological Algebra" which was a joint work with H. Cartan, thereby introducing it to Japan, and this caused great excitement. It could be said that this stimulated Yoneda to study the product on Ext* [97].

In 1954 Toda was invited to CNRS in France, where he worked hard on the homotopy groups of spheres and Lie groups by using his construction called the "Toda bracket", and received high praise from Cartan and others. In May, 1957 Komatu transferred to Kyoto University along with Toda. But even after that, until about 1965, seminars were still being held in Osaka City University, and many mathematicians in the Kansai district (the western part of Japan) participated in them. In the seminars, Nakaoka studied the relation between Smith's special cohomology theory and Steenrod operations, and the cohomology groups of cyclic products and symmetric products; Yokota gave a wonderful cellular decomposition of the classical groups [96]; Mizuno studied the obstruction theory from the viewpoint of Postnikov decompositions. Masahiro Sugawara (1928–) participated in this seminar from time to time, although he was at Okayama University at that time. He discussed a condition for a space to be an H-space [70]; he also showed in [71] that a classifying space can be constructed for some kind of H-spaces. By this time, Kudo and Uehara had transferred to Kyushu University, and Shōrō Araki (1930–), a graduate of Nagoya University, joined them too. They considered the spectral sequence of a fibre space. Kudo proved a transgression theorem [35] and Araki introduced in [5] squaring operations of Steenrod in the spectral sequence. Moreover, Kudo and Araki wrote a joint paper [36] in which they proved that, for the homology of some kinds of spaces, homology operations corresponding to squaring operations for cohomology can be constructed. This was later generalised by Dyer–Lashof, and now it plays an important role in the theory of infinite loop spaces. Other studies around that time were those by Ken-ichi Shiraiwa (1928–) of Nagoya University who discussed the homotopy types of $(n-1)$-connected $(n+3)$-dimensional complexes; work on the minimal complex of a fibre space by Tokusi Nakamura (1930–) of the University of Tokyo; the paper [95] by Tsuneyo Yamanoshita (1929–) of Tsuda Women's University who calculated unstable homotopy groups of spheres; a work by Haruo Suzuki (1931–) on Eilenberg–MacLane invariants of loop spaces, and so on.

Nakaoka and Araki (from 1958), and Toda (from 1959) visited the Institute for Advanced Study in Princeton for two years, where Nakaoka obtained a stability theorem for the homology groups of a symmetric group, and determined the homology groups of the infinite symmetric group [56]. Araki studied differential Hopf algebras and their application to the cohomology of compact exceptional Lie groups [6]. Toda completed a book [84] in which he determined the homotopy groups $\pi_{n+k}(S^n)$ of spheres for $k \leqslant 19$; he then spent one year at Northwestern University at the invitation of Yamabe, who however had passed away before he came over. (Toda's book is often quoted in subsequent papers on topology.) During that time, just before Toda went to the USA, I.M. James visited Japan, giving lectures at the major universities and providing a good stimulus to the Japanese topologists. In Japan, Nakamura at the University of Tokyo studied a relation between the constructive definition and the formal definition of cohomology operations [54], also Akio Hattori (1929–) of the University of Tokyo studied the spectral sequence of de Rham cohomology of fibre spaces. Shimada and Yamanoshita independently proved the triviality of mod p Hopf invariants [68], and Sugawara who transferred to Kyoto University studied the homotopy commutativity of groups and loop spaces [72].

By 1960 results started to appear by people who had recently obtained their master's degree. Many works on homotopy theory were announced during the following five years: a work [64] on homotopy types of certain kinds of complexes by Seiya Sasao (1933–), a graduate of the University of Tokyo; a work [62] on Toda brackets and its application by Kunio Ôguchi (1933–), also a graduate of the University of Tokyo; a work [61] on the so-called soft homotopy by Yasutoshi Nomura (1932–), a graduate of Nagoya University; works [41, 42] on homotopy groups of unitary groups by Hiromichi Matsunaga (1935–), a graduate of Kyushu University; works [43, 44] on homotopy groups of spheres and classical groups by Mamoru Mimura (1938–), a graduate of Kyoto University, and so on. During this period, Araki, who had transferred to Osaka City University, studied homology of symmetric spaces intensively. Mimura went to France as a Scholarship Fellow of the French Government in 1964 for two years; during that period he visited the UK at the invitation of J.F. Adams, M.G. Barratt and I.M. James, and also met M. Mahowald and F.P. Peterson in West Germany; he was influenced mathematically by these people rather than by the French topologists.

Around 1960, M.F. Atiyah and F. Hirzebruch created K-theory to prove the Riemann–Roch theorem on differentiable manifolds, and since then K-theory and other generalised cohomology theories were studied actively and widely throughout the world. From about 1965, a series of results in this area were announced in Japan. Typical of these were the achievements of Hattori and Araki–Toda. The work of Hattori [11, 12] was the affirmative solution of the Atiyah–Hirzebruch conjecture on the exactness of the Riemann–Roch relations among the Chern numbers, which was proved independently by Stong around the same time; thereafter it was referred to as "the theorem of Hattori–Stong". In contrast to the computational proof by Stong, Hattori succeeded by reducing the statement to a proposition concerning the Hurewicz homomorphism in K-theory, a result which exerted much influence on the development of general cohomology theory. Araki and Toda [100] worked on a thorough study of multiplications on general cohomology with mod p coefficients, which is considered to be one of the basic references in general cohomology. Among other research in these fields published by 1970 were Araki's work [7] giving another proof without using the classification of simple Lie groups of Hodgkin's theorem on the structure of the K-theory of Lie groups, a study by Haruo Suzuki of Kyushu University and Teiichi Kobayashi (1936–) of Kyoto University on the possibility of embedding projective spaces or lens spaces into Euclidean space as an application of K-theory, and many works done by young mathematicians in Kyoto University and Osaka City University. Around 1966 in Kyoto University a research group of algebraic topology was organised by Toda and his pupil Mimura, and they studied actively the homotopy groups of compact Lie groups [47]. The young and promising Goro Nishida (1942–) and Akihiro Tsuchiya (1942–) dealt with the homology of infinite loop spaces; the former [58] discovered relations, which came to be called the "Nishida formulae", between Kudo–Araki–Dyer–Lashof homology operations and classical Steenrod cohomology operations, and the latter [87–90] analysed the characteristic classes of spherical fibrations and PL micro-bundles by applying Dyer–Lashof homology operations, which was a generalisation of the result of Kudo–Araki.

Mimura, Nishida and Toda developed in [46] the localisation theory of CW-complexes in a quite different manner from that of D. Sullivan; that is, they defined the localisation of

CW-complexes by extending the ideas of [48, 45]. Around the same time Toda wrote two remarkable papers [85, 86], the first of which was later to be extended by Nishida to prove the nilpotency of elements of the stable homotopy groups of spheres.

By that time Nakaoka had transferred to Osaka University, Sugawara to Hiroshima University, Suzuki to Hokkaido University, and Yokota to Shinshu University, where they headed research groups of topology.

5.2. *Differential topology*

The modern theory of the topology of manifolds started with Whitney's 1936 paper "*Differentiable manifolds*", which formulated definitively the concept of manifold in the present form. However its real development began after Thom's cobordism theory in 1954, a product of the algebraic topology developed after World War II, and Milnor's work in 1956 showing that the 7-dimensional sphere admits distinct differentiable structures. It was around 1960 that the term "differential topology" started to be used.

Research in this field started in Japan about ten years later than that in algebraic topology. Ryozi Shizuma, who transferred to Nagoya University from Osaka University during World War II, gradually changed his interest to topology of manifolds from the homotopical study of fibre bundles which he had been studying since 1950. Under his influence Nobuo Shimada and Masahisa Adachi (1931–1993) emerged from Nagoya University as the influential mathematicians in this field. In 1957 Shimada published a paper [65] applying the methods of Milnor to the 15-dimensional sphere, showing that it admits distinct differentiable structures. Adachi calculated [4] cobordism classes of higher dimension than Thom's calculations produce. Shizuma himself published an interesting paper [69] in 1958 on the existence of closed geodesics on a manifold.

The University of Tokyo got a late start, but from the middle of 1950's there appeared graduate students who started to study differential topology. Ichiro Tamura (1926–1991) of the University of Tokyo wrote a paper [74] showing that Pontrjagin classes are not homotopy invariants, and also in a paper [75] of the following year he showed by generalising Milnor's invariant that there exist some 2-connected 7-dimensional rational spheres and 6-connected 15-dimensional rational spheres which admit distinct differentiable structures. Furthermore, Tamura generalised characteristic classes to manifolds with singularities. Tamura published a paper [76] in 1961 showing that there exists an 8-dimensional manifold which cannot have a differentiable structure. This was proved independently by Eells, Kuiper and Wall almost at the same time. In Tôhoku University, Haruo Suzuki tried in a paper [73] in 1958 to generalise Thom's result realising characteristic classes by submanifolds.

It was in 1960 that the solution by Smale of the higher-dimensional Poincaré conjecture was published, and around the same time Hiroshi Yamasuge of Osaka City University reached the solution of the Poincaré conjecture for the 5-dimensional case independently, starting from the partition of differentiable manifolds by Morse functions. It was regrettable that he died young in November 1960; his result [95] was published in 1961 after his death. His work is conspicuous by its originality.

In addition to the above mentioned papers, there are some other papers published in 1961/1962, such as one on singular sets of mappings by Yoshihiro Saito (1930–1997) of Osaka City University, one on Pontrjagin classes by Yoshihiro Shikata (1936–) of Osaka

University, one on the index of homogeneous spaces by Hattori of the University of Tokyo, and one on approximations of mappings between differentiable manifolds by Kôzi Shiga (1930–) of Tokyo Institute of Technology, and so on.

Most of these mathematicians were around 30 years old, and many of them began the study of topology by themselves. Their number was so limited that they could not organise a school, but their researches were so advanced at a time of worldwide rapid developments in differential topology that it drew international attention even in the early 1960's. From around 1965 some of the above mathematicians started to supervise graduate students in major universities, and hence there appeared the first group of mathematicians who had received their education in this field from the beginning, such as Hajime Sato (1944–), Katsuo Kawakubo (1942–) and Yukio Matsumoto (1944–). By the late 1960's research in Japan had become substantial.

Differential topology began to have a close relation with the combinatorial theory of manifolds from about 1962 or 1963. In this latter field, in addition to a paper on the introduction of a prebundle in 1967 by Mitsuyoshi Kato (1942–) of Tokyo Metropolitan University (see the section on Combinatorial Topology), some papers on higher-dimensional knots published in 1969 are worthy of note.

In 1967 the construction theory for manifolds and its application to the Hauptvermutung was published by Sullivan, and furthermore a complete solution of the Hauptvermutung was published by Kirby and Siebenmann in 1969. These results had much influence on differential topology.

From the middle of the 1960's differential topology became connected with transformation group theory, the theory of dynamical systems, the structure theory of foliations, the theory of singularities and so on, in each of which many results were obtained. Among the papers published in Japan in 1968–1970 were those on the generalisation of the Borsuk–Ulam theorem by Minoru Nakaoka [57] in transformation group theory, on free and semi-free actions of homotopy spheres by Kawakubo [22], on cobordism groups of semi-free actions of S^1 and S^3 by Fuichi Uchida (1938–) [91], and on the Brieskorn algebraic variety and the differential topological generalisation of group actions on it by Tamura [77] (this result furnished with a basic method in constructing codimension 1 foliations on an odd dimensional sphere). Until 1970 research in Japan in fields such as transformation groups, dynamical systems and foliation structures was limited, but after a few years, it developed rapidly, and the results were presented at the International Conference on Manifolds and Related Topics in Topology in 1973 in Tokyo of which we shall speak later.

5.3. *Combinatorial topology*

Combinatorial topology is topology with a strong geometrical aspect, also called PL (piecewise linear) topology or geometric topology; it is a subject which is concerned with the structure of a complex (a polyhedron) consisting of simplexes of simple figures or of a manifold. The relation with differential topology is very important and, as the difference between combinatorial and algebraic topology was not so clear at the beginning of the history of topology, algebraic methods were a driving force. Low-dimensional topology in most cases is studied by specialists in combinatorial topology.

Active research was done in the 1950's by Hidetaka Terasaka of Osaka University, and in the 1960's by Hiroshi Noguchi (1925–) of Waseda University, Tatsuo Homma (1926–) of Yokohama City University and Fujitsugu Hosokawa (1930–) of Kobe University. These were the leaders, who taught many young mathematicians who are active nowadays. Visits to Japan by people like Moise, Fox, Cairns, Bing, Harold and visits to the USA by Noguchi, Shin'ichi Kinoshita (1925–), Homma, Kunio Murasugi (1929–), Hosokawa, Junzo Tao (1929–) made possible a close research cooperation and the exchange of information between Japan and the USA. The Japanese academic world was greatly stimulated by the Hauptvermutung for 3-dimensional manifolds and the affirmative solution of the triangulation problem by Moise (1952), Dehn's lemma and the proof of the sphere theorem by Papakyriakopoulos (1957), the unknotting theorem by Zeeman, the proof of the Poincaré conjecture for dimension $\geqslant 5$ (1963), the theorem on the Hauptvermutung for higher dimensions by Kirby and Siebenmann and others, and the development of the triangulation problem (1969).

Right after World War II, Terasaka and Kinoshita in Osaka and Noguchi, Homma and Murasugi in Tokyo were working by themselves without good contacts, but research made good progress when Homma and Noguchi paid a visit to Osaka University. Around 1950 Terasaka, Kinoshita and Homma were studying homeomorphisms of $\mathbb{R}^2$ and of $\mathbb{R}^3$, which bore fruit as work on quasi-parallel translations of $\mathbb{R}^3$ by Kinoshita [24]. In this research the importance of how to embed a closed curve (knot) and a surface (2-dimensional manifold) in $\mathbb{R}^3$ attracted attention. Homma proved Dehn's lemma, i.e. whether a knot ($S^1 \subset \mathbb{R}^3$) is truly unknotted is determined by whether the knot group $\pi_1(\mathbb{R}^3 - S^1)$ is isomorphic to $\mathbb{Z}$, almost at the same time as and independently of Papakyriakopoulos [13]. Terasaka, Kinoshita and others also undertook research in knot theory. One of the typical results at that time was the generalisation of composition of knots by Kinoshita–Terasaka [25]. It can be said that research in knot theory in Japan truly started around this time.

Murasugi has always been an authority in research on knots by algebraic methods. His research standard was high, in both quality and quantity. Let us mention a few typical examples: a work [50] on the genus of the alternating knot and on Alexander's polynomial, and a work [51] on the matrix of links defined by Murasugi and the signatures of links, which has many applications. Alexander's polynomial is a strong tool for research on knots and links together with the knot group, and by defining the ∇-polynomial (Hosokawa polynomial) using Alexander's polynomial, Hosokawa proved [15] interesting results such as the relation with a covering space with links as a bifurcation set. Japanese mathematicians played the leading role in research on ribbon knots.

In generalising knots in higher dimension, only the cases of codimension 1 or 2 are interesting, by Zeeman's result mentioned above. The case of codimension 1 is called the Schoenflies conjecture, which has not been solved completely. For the case of codimension 2, almost all Japanese mathematicians who studied combinatorial topology such as Terasaka, Takeshi Yajima (1914–), Noguchi, Kinoshita, Hosokawa and Takaaki Yanagawa (1935–) took some interest in it, and valuable research has resulted. Yajima [93] considered a certain kind of 2-dimensional sphere in $\mathbb{R}^4$, which was called later the ribbon 2-knot. One of the reasons why higher-dimensional knots give rise to such great interest is that understanding this problem is indispensable in the construction and embedding of higher-dimensional manifolds. Thanks to the efforts of Noguchi and Mitsuyoshi Kato, who was a student of Noguchi at Waseda University, research in this direction is well developed, and it will be mentioned again later.

In the late 1960's, work on the Hauptvermutung and the triangulation problem made great progress. The aim was to make clear the relation between combinatorial manifolds and general manifolds, and Kato and Yukio Matsumoto pushed the study forward. Homma [14] proved that a homeomorphism of n-dimensional manifolds can be approximately modified to be PL up to a k-dimensional submanifold if $2k + 2 \leqslant n$ but it was later proved by Miller and others that this theorem holds for codimension $\geqslant 3$ (that is, $n - k \geqslant 3$).

The problems of smoothing or locally flat embedding are important in differential topology, and were studied by Noguchi and Kinoshita. Assuming the Schoenflies conjecture on the smoothing of an n-manifold in $\mathbb{R}^{n+1}$, Noguchi showed in [59] that it can be approximated by a differentiable manifold. Noguchi's work [60] that an embedding of a manifold of codimension 2 can be approximated by a locally flat embedding if the Whitehead class is 0 was highly acclaimed, and this research was followed up by Kato and others. Defining a prebundle by introducing a bundle structure on a combinatorial manifold, Kato [19] obtained many results beginning with a locally flat embedding. This paper was from a different and independent viewpoint, although it partly overlapped with the work on block bundles by Rourke and Sanderson. It exerted much influence on topology and became a driving force in the development of combinatorial topology in Japan in the 1970's. Then Kato obtained a counterexample in the (relative) Hauptvermutung for regular neighbourhoods, by studying the action of the Whitehead group on regular neighbourhoods in a sphere constructed using higher dimensional knots [20, 21]. Under the stimulus of the "drama" of the Hauptvermutung, Y. Matsumoto [40] extended the result of Sullivan on the Hauptvermutung for simply connected manifolds to the case where the fundamental group is the infinite cyclic group.

The International Conference on Manifolds and Related Topics in Topology was held in Tokyo during the period April 10–17, 1973 by the Mathematical Society of Japan, under the co-sponsorship of the International Mathematical Union and the support of the Ministry of Education and the Science Council of Japan. The organising committee of the conference consisted of Y. Akizuki, S. Araki, M.F. Atiyah, A. Hattori, H. Hironaka, T. Homma, M. Hukuhara, S. Iitaka, S. Iyanaga, Y. Kawada (chairman), K. Kodaira (chairman of the program committee), A. Komatu, J. Milnor, M. Nakaoka, K. Ono, V. Poénaru, S. Sasaki, N. Shimada, T. Shioda, K. Shoda, I. Tamura, H. Toda and K. Yosida. M.F. Atiyah and V. Poénaru acted as representatives of the IMU on the organising committee. Financial support for the conference was provided by the IMU and by donations from Japanese companies.

The topics of the conference were differentiable manifolds, PL manifolds, certain aspects of complex manifolds and of algebraic varieties, and related topics in topology such as homotopy, generalised cohomology, singularities, foliations, and dynamical systems.

The proceedings [12] were published in 1975.

Acknowledgement

The author wishes to thank John Hubbuck and Martin Guest for reading the manuscript, H. Fukaishi for collecting materials and Professor S. Iyanaga for giving him much useful advice and information. Indeed, it was Professor Iyanaga's letter to the editor of this book that prompted the writing of this manuscript.

Bibliography

[1] M. Abe, *Über die Methode der Polyederentwicklung der Kompakten und ihre Anwendungen auf die Abbildungstheorie*, Compositio Math. **7** (1939), 185–193.

[2] M. Abe, *Über Automorphismen der lokal-kompakten Abelschen Gruppen*, Proc. Imp. Acad. Japan **16** (1940), 59–62.

[3] M. Abe, Isosugaku **3** (1940), 58–62. ("Isosugaku Kenkyu" (1950), 32–42 (in Japanese).

[4] M. Adachi, *On the groups of cobordism Ω^k*, Nagoya Math. J. **13** (1958), 135–156.

[5] S. Araki, *Steenrod reduced powers in the spectral sequences associated with a fibering*, Mem. Fac. Sci. Kyushu Univ. **11** (1957), 15–64.

[6] S. Araki, *Differential Hopf algebras and the cohomology* mod 3 *of the compact exceptional groups E_7 and E_8*, Ann. of Math. **73** (1961), 404–436.

[7] S. Araki, *Hopf structures attached to K-theory: Hodgkin's theorem*, Ann. of Math. **85** (1967), 508–525.

[8] S. Araki and H. Toda, *Multiplicative structures in* mod q *cohomology theories*, I, II, Osaka J. Math. **2** (1965), 71–115; **3** (1966), 81–120.

[9] K. Haratomi, *Über die höherstufige Separabilität und Kompaktheit I, II*, Japanese J. Math. **8** (1931), 113–142; **9** (1932), 1–18.

[10] K. Haratomi, *On a topological problem*, Japanese J. Math. **9** (1932), 103–110.

[11] A. Hattori, *Integral characteristic numbers for weakly almost complex manifolds*, Topology **5** (1966), 259–280.

[12] A. Hattori (ed.), *Manifolds Tokyo 1973*, Univ. of Tokyo Press, Tokyo (1975).

[13] T. Homma, *On Dehn's lemma for S^3*, Yokohama Math. J. **5** (1957), 223–244.

[14] T. Homma, *On the imbedding of polyhedra in manifolds*, Yokohama Math. J. **10** (1962), 5–10.

[15] F. Hosokawa, *On ∇-polynomials of links*, Osaka Math. J. **10** (1958), 273–282.

[16] T. Inagaki, *Sur les ensembles analytiques nuls*, J. Fac. Sci. Hokkaido Imp. Univ. (1) **6** (1937), 175–216.

[17] S. Kakutani, *Über die Metrisation der topologischen Gruppen*, Proc. Imp. Acad. Japan **12** (1936), 82–84.

[18] S. Kakutani, *A generalization of Brouwer's fixed point theorem*, Duke Math. J. **8** (1941), 457–459.

[19] M. Kato, *Combinatorial prebundles I, II*, Osaka J. Math. **4** (1967), 289–303, 305–311.

[20] M. Kato, *Regular neighborhoods are not topologically invariant*, Bull. Amer. Math. Soc. **74** (1968), 988–991.

[21] M. Kato, *Geometric operations of Whitehead groups*, J. Math. Soc. Japan **21** (1969), 523–542.

[22] K. Kawakubo, *Free and semi-free differentiable actions on homotopy spheres*, Proc. Japan Acad. **45** (1969), 651–655.

[23] K. Kawakubo and F. Uchida, *On the index of a semi-free S^1-action*, J. Math. Soc. Japan **23** (1971), 351–355.

[24] S. Kinoshita, *On quasi-translations in* 3*-space*, Fund. Math. **56** (1964), 69–79.

[25] S. Kinoshita and H. Terasaka, *On unions of knots*, Osaka Math. J. **9** (1957), 131–153.

[26] K. Kodaira, *Die Kuratowskische Abbildung und der Hopfsche Erweiterungssatz*, Compositio Math. **7** (1939), 177–184.

[27] K. Kodaira and M. Abe, *Über zusammenhängende kompakte Abelsche Gruppen*, Proc. Imp. Acad. Japan **16** (1940), 167–172.

[28] A. Komatu, *Über die Dualitätssätze der Überdeckungen*, Japanese J. Math. **13** (1936), 493–500.

[29] A. Komatu, *Über die Ringdualität eines Kompaktums*, Tôhoku Math. J. **43** (1937), 414–420.

[30] A. Komatu, *Zur Topologie der Abbildungen von Komplexen*, Japanese J. Math. **17** (1941), 201–228.

[31] A. Komatu, *The Theory of Topological Spaces*, Iwanami (1947) (in Japanese).

[32] A. Komatu and R. Sakata (Shizuma), *Einige Sätze über Abbildungen auf die Sphäre*, Japanese J. Math. **16** (1940), 163–167.

[33] M. Kondô, *A problem of the metrisation in Hausdorff's topological spaces*, Tôhoku Math. J. **37** (1933), 383–391.

[34] M. Kondô, *Sur l'uniformisation des complémentaires analytiques et les ensembles projectifs de la seconde classe*, Japanese J. Math. **15** (1939), 197–230.

[35] T. Kudo, *A transgression theorem*, Mem. Fac. Sci. Kyushu Univ. **9** (1956), 79–81.

[36] T. Kudo and S. Araki, *Topology of H_n-spaces and H-squaring operations*, Mem. Fac. Sci. Kyushu Univ. **10** (1956), 85–120.

[37] K. Kunugui, *Sur un théorème de Baire généralisé dans la théorie des espaces abstraits*, Fund. Math. **21** (1933), 244–249.

[38] K. Kunugui, *The Theory of Abstract Spaces*, Iwanami (1933) (in Japanese).
[39] K. Kunugui, *Contribution à la théorie des ensembles Boreliens et analytiques III*, J. Fac. Sci. Hokkaido Imp. Univ. (1) **8** (1940), 79–108.
[40] Y. Matsumoto, *Hauptvermutung for* $\pi_1 = \mathbb{Z}$, J. Fac. Sci. Univ. Tokyo **16** (1969), 165–177.
[41] H. Matsunaga, *The homotopy groups* $\pi_{2n+1}(U(n))$ *for* $i = 3, 4$ *and* 5, Mem. Fac. Sci. Kyushu Univ. **15** (1961), 72–81.
[42] H. Matsunaga, *Applications of functional cohomology operations to the calculus of* $\pi_{2n+i}(U(n))$ *for* $i = 6$ *and* 7, $n \geqslant 4$, Mem. Fac. Sci. Kyushu Univ. **17** (1963), 29–62.
[43] M. Mimura, *On the generalized Hopf homomorphisms and the higher composition I, II*, J. Math. Kyoto Univ. **4** (1964), 171–190, 301–326.
[44] M. Mimura, *The homotopy of Lie groups of low rank*, J. Math. Kyoto Univ. **6** (1967), 131–176.
[45] M. Mimura, R.C. O'Neill and H. Toda, *On p-equivalence in the sense of Serre*, Japanese J. Math. **40** (1971), 1–10.
[46] M. Mimura, G. Nishida and H. Toda, *Localization of CW-complexes and its applications*, J. Math. Soc. Japan **23** (1971), 593–624.
[47] M. Mimura and H. Toda, *Cohomology operations and the homotopy of compact Lie groups I*, Topology **9** (1970), 317–336.
[48] M. Mimura and H. Toda, *p-equivalences and p-universal spaces*, Comment. Math. Helv. **46** (1971), 87–97.
[49] K. Morita, *On uniform spaces and the dimension of compact spaces*, Proc. Phys. Math. Soc. Japan **22** (1940), 967–977.
[50] K. Murasugi, *On the genus of the alternating knot I, II*, J. Math. Soc. Japan **10** (1958), 94–105, 235–248.
[51] K. Murasugi, *On the signature of links*, Topology **9** (1970), 283–298.
[52] K. Nakamura, *Topology*, Iwanami (1933) (in Japanese).
[53] K. Nakamura, *A Survey of Topology*, Kyoritsu-Shuppan (1935) (in Japanese).
[54] T. Nakamura, *Equivalence between two definitions of the cohomology operations*, Sci. Papers Colloq. Gen. Ed. Univ. Tokyo **9** (1959), 1–16.
[55] M. Nakaoka, *Classification of mappings of a complex into a special kind of complex*, J. Inst. Polytech., Osaka City Univ. **3** (1952), 101–143.
[56] M. Nakaoka, *Homology of the infinite symmetric group*, Ann. of Math. **73** (1961), 229–257.
[57] M. Nakaoka, *Generalizations of Borsuk–Ulam theorem*, Osaka J. Math. **7** (1970), 423–441.
[58] G. Nishida, *Cohomology operations in iterated loop spaces*, Proc. Japan Acad. **44** (1968), 104–109.
[59] H. Noguchi, *The smoothing of combinatorial n-manifolds in* $(n + 1)$*-space*, Ann. of Math. **72** (1960), 201–215.
[60] H. Noguchi, *Obstructions to locally flat embeddings of combinatorial manifolds*, Topology **5** (1966), 203–213.
[61] Y. Nomura, *On mapping sequences*, Nagoya Math. J. **17** (1960), 115–145.
[62] K. Ôguchi, *Generators of 2-primary components of homotopy groups of spheres, unitary groups and symplectic groups*, J. Fac. Sci. Univ. Tokyo **11** (1964), 65–111.
[63] R. Sakata (Shizuma), *Über Abbildungen der Kompakten auf die Sphäre*, Proc. Imp. Acad. Japan **14** (1938), 301–303.
[64] S. Sasao, *Homology 4-spheres with boundary*, Topology **7** (1968), 417–427.
[65] N. Shimada, *Differentiable structures on the 15-sphere and Pontrjagin classes of certain manifolds*, Nagoya Math. J. **12** (1957), 59–69.
[66] N. Shimada and H. Uehara, *On a homotopy classification of mappings of an* $(n + 1)$*-dimensional complex into an arcwise connected topological space which is aspherical in dimensions less than n* $(n > 2)$, Nagoya Math. J. **3** (1951), 67–72.
[67] N. Shimada and H. Uehara, *Classification of mappings of an* $(n + 2)$*-complex into an* $(n - 1)$*-connected space with vanishing* $(n + 1)$*-st homotopy group*, Nagoya Math. J. **4** (1952), 43–50.
[68] N. Shimada and T. Yamanoshita, *On triviality of the mod p Hopf invariant*, Japanese J. Math. **31** (1961), 1–25.
[69] R. Shizuma, *Über geschlossene Geodätische auf geschlossenen Mannigfaltigkeiten*, Nagoya Math. J. **13** (1958), 104–114.
[70] M. Sugawara, *On a condition that a space is an H-space*, Math. J. Okayama Univ. **6** (1957), 109–129.
[71] M. Sugawara, *A condition that a space is group-like*, Math. J. Okayama Univ. **7** (1957), 123–149.

[72] M. Sugawara, *On the homotopy-commutativity of groups and loop spaces*, Mem. Colloq. Sci. Kyoto Univ. **33** (1960/1961), 257–269.
[73] H. Suzuki, *On the realization of the Stiefel–Whitney characteristic classes by submanifolds*, Tôhoku Math. J. **10** (1958), 91–115.
[74] I. Tamura, *On Pontrjagin classes and homotopy types of manifolds*, J. Math. Soc. Japan **9** (1957), 250–262.
[75] I. Tamura, *Homeomorphy classification of total spaces of sphere bundles over spheres*, J. Math. Soc. Japan **10** (1958), 29–43.
[76] I. Tamura, *8-manifolds admitting no differentiable structure*, J. Math. Soc. Japan **13** (1961), 377–382.
[77] I. Tamura, *On the classification of sufficiently connected manifolds*, J. Math. Soc. Japan **20** (1968), 371–389.
[78] H. Terasaka, *Ein Beweis des Brouwerschen ebenen Translationssatzes*, Japanese J. Math. **7** (1930), 61–69.
[79] H. Terasaka, *On the division of Riemann surface into sheets*, Japanese J. Math. **8** (1931), 309–326.
[80] H. Terasaka, *Topologische Abbildungen und Kurvensysteme in* $\mathbb{R}^n$, Japanese J. Math. **14** (1938), 1–13.
[81] H. Terasaka, *Zweiparametrige reguläre Kurvensysteme auf der 2-Sphäre*, Japanese J. Math. **15** (1939), 57–103.
[82] H. Terasaka, *Die Theorie der topologische Verbände*, Fund. Math. **33** (1939); and Collected Papers from the Faculty of Sci. Osaka Imp. Univ. **8** (1940), 1–33.
[83] H. Toda, *Generalized Whitehead products and homotopy groups of spheres*, J. Inst. Polytech. Osaka City Univ. **3** (1952), 43–82.
[84] H. Toda, *Composition Methods in Homotopy Groups of Spheres*, Ann. of Math. Stud. vol. 49, Princeton Univ. Press, Princeton, NJ (1962).
[85] H. Toda, *Extended p-th powers of complexes and applications to homotopy theory*, Proc. Japan Acad. **44** (1968), 198–203.
[86] H. Toda, *On spectra realizing exterior parts of the Steenrod algebra*, Topology **10** (1971), 53–65.
[87] A. Tsuchiya, *Characteristic classes for spherical fibre spaces*, Proc. Japan Acad. **44** (1968), 617–622.
[88] A. Tsuchiya, *Characteristic classes for spherical fiber spaces*, Nagoya Math. J. **43** (1971), 1–39.
[89] A. Tsuchiya, *Characteristic classes for* PL *micro bundles*, Bull. Amer. Math. Soc. **77** (1971), 531–534.
[90] A. Tsuchiya, *Characteristic classes for* PL *micro bundles*, Nagoya Math. J. **43** (1971), 169–198.
[91] F. Uchida, *Cobordism groups of semi-free* S^1 *and* S^3*-actions*, Osaka J. Math. **7** (1970), 345–351.
[92] T. Wada, *The conception of a curve*, Mem. Colloq. Sci. Engrg. Kyoto Imp. Univ. **3** (1911/1912), 265–275.
[93] T. Yajima, *On simply knotted spheres in* $\mathbb{R}^4$, Osaka J. Math. **1** (1964), 133–152.
[94] T. Yamanoshita, *On the homotopy groups of spheres*, Japanese J. Math. **27** (1957), 1–53.
[95] H. Yamasuge, *On Poincaré conjecture for* M^5, J. Math. Osaka City Univ. **12** (1961), 1–17.
[96] I. Yokota, *On the cellular decompositions of unitary groups*, J. Inst. Polytech. Osaka City Univ. **7** (1956), 39–49.
[97] N. Yoneda, *On the homology theory of modules*, J. Fac. Sci. Univ. Tokyo **7** (1954), 193–227.
[98] K. Yoneyama, *The conception of a curve, a surface and a solid*, Mem. Colloq. Sci. Engrg. Kyoto Univ. **5** (1912/1913), 261–269.
[99] K. Yoneyama, *Theory of continuous set of points I–III*, Tôhoku Math. J. **12** (1917), 43–158; **13** (1918), 33–158; **18** (1920), 134–186, 205–255.
[100] *Special Issue on Algebraic Topology*, Sugaku **10** (1958), Mathematics, Math. Soc. of Japan (in Japanese).

CHAPTER 32

Some Topologists

I.M. James

Mathematical Institute, 24–29 St. Giles, Oxford OX1 3LB, UK
E-mail: imj@maths.ox.ac.uk

To include even short biographies of all the topologists who have been mentioned in this book would be impracticable. To try and select the most important would seem presumptuous. After careful consideration I decided, reluctantly, to exclude living persons. Even after this the numbers are too great so I chose about thirty who are certainly not minor figures and whose lives are not only interesting in themselves but in some way illustrative of the period and part of the world in which they lived.

It seemed appropriate to arrange for a separate article on the Japanese school, because of its distinctive character, and not to arrange for individual biographies. Unfortunately I was unable to do the same for the Russian school. The idea of organizing all the material under schools seemed attractive in some ways but difficult to carry out. The life of someone like Hurewicz, who spent his early years in Poland, began his academic career in Austria, continued it in the Netherlands, and ended up in the United States is such that one might hesitate to class him as an American, although as a matter of fact he took out American nationality, since from a cultural point of view he remained so much a European and it was in Europe that he did his most important work.

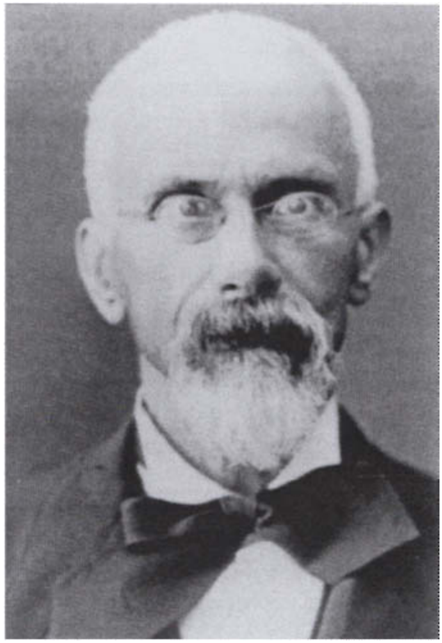
The present chapter contains short biographies of about twenty individuals which I have compiled myself, drawing mainly on published sources. The remaining chapters in the book consist of rather longer biographies of a number of individuals written by people with special knowledge of the life of the subject. These more extended biographies often contain information not published hitherto.

Short biographies

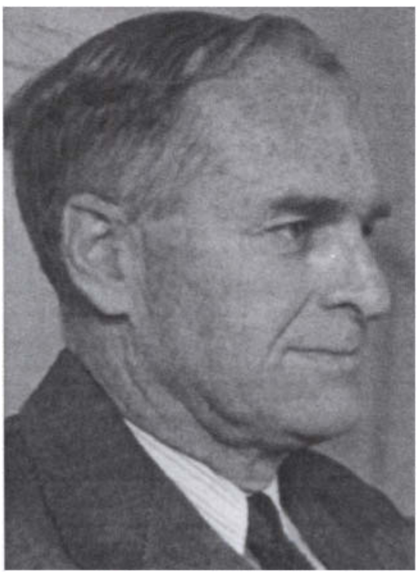
To arrange the short biographies in alphabetical order, although convenient for reference, would tend to mask the often significant relationships between individuals. After trying various alternatives I have ended up by arranging them chronologically according to date of birth. The information has been collected from a variety of sources, including the standard reference books such as the Dictionary of Scientific Biography, and the obituary articles and other memoirs which have been published, usually by learned societies of which the

HISTORY OF TOPOLOGY
Edited by I.M. James

Auguste Möbius (1790–1868)



Enrico Betti (1823–1892)



James Alexander (1888–1971)

Solomon Lefschetz (1884–1972)

subject was a member. The principal sources used are listed at the end of each biography. However, such accounts are often of a rather formal character and where possible this has been supplemented by further information from other sources. Since research contributions are generally described elsewhere in this volume, no more than brief indications are given here.

Topology first began to develop in the German-speaking part of Europe. The universities of Berlin, Göttingen, Leipzig, Munich and Vienna were particularly important, but others, such as Königsberg and Prague, played a significant role. The terminology used to describe academic posts, qualifications etc. varies from country to country. Not without some reluctance I have succumbed to the practice of translating the terms ausserordentliche professor and ordentliche professor, and their equivalents, by the terms associate professor and full professor, although the correspondence is only very approximate. However, I have retained the term habilitation for the distinctive qualification, usually obtained several years after the Ph.D., which is a requirement for becoming a university teacher under the German system.

There is a problem in deciding just where to begin. Leibniz, Euler and Gauss were precursors, but so much has been written about them already that to have included short biographies here would seem rather pointless. It is hardly possible to claim Gauss himself as a topologist but there can be little doubt that he inspired Listing and Möbius. Listing is the subject of an extended biography in the next section. He is less well-known than Möbius although in some respects his work is more important in the history of topology. I begin the present section with a note on the life [3] of his contemporary Möbius.

August Ferdinand MÖBIUS was born 17 November 1790 at Schulpforta, near Naumburg (Germany). He was the only child of Johann Heinrich Möbius, a dancing teacher in Schulpforta until his death in 1793, and the former Johanne (née Keil), a descendant of Martin Luther. His father's unmarried brother succeeded him as dancing teacher and as provider for the family until his own death in 1804. Möbius was educated at home until his thirteenth year, by which time he was already displaying an interest in mathematics. In 1803 he began formal education in his native town, where he studied mathematics under Johann Gottlieb Schmidt. The young man entered Leipzig University in 1809 to study law, but his early love for mathematics soon prevailed, leading him to take courses in mathematics, physics and astronomy instead.

In May 1813, shortly before the battle of Leipzig, Möbius went to Göttingen, where he spent two semesters studying theoretical astronomy under Gauss. He then proceeded to Halle for further studies in mathematics with Pfaff. The following year he returned to Leipzig as professor of astronomy, which he remained for the rest of his life. In 1820 Möbius' mother, who had come to live with him, died. Shortly afterwards he married Dorothea (née Rothe), whose subsequent blindness did not prevent her from raising a daughter, Emilie, and two sons, both of whom became distinguished literary scholars.

Möbius rarely travelled, and in general his life centered around his study, his observatory, and his family. His writings were fully developed and original, but he was not widely read in the mathematical literature of his day, and consequently found at times that others had previously discovered ideas presented in his writings. Moreover, his investigations were frequently aimed at developing simpler and more effective methods for treating existing subjects. Like his contemporaries Gauss and Hamilton, Möbius was employed as an astronomer but made his most important contributions in mathematics.

So far as topology is concerned, Möbius is now best remembered for his discovery of the one-sided surface called the Möbius strip, which may be formed by glueing together the ends of a rectangular strip of paper after giving it a half twist. The Paris Academy had offered a prize for research on the geometric theory of polyhedra, and in 1858 Möbius began to prepare an essay on the subject. Although this was unsuccessful, possibly because his French was not of the best, Möbius published much of the material later, including a discussion of the famous strip. It seems that Listing discovered the properties of the strip independently at about the same time. Ten years later, on 26 September 1868, Möbius died, shortly after celebrating fifty years of teaching at Leipzig; his wife had died nine years earlier.

Much has been written about Riemann, whose ideas have proved so important for modern mathematics. Since a full-scale biography [18] has recently appeared, only an outline of his life is given here.

(Georg Friedrich) Bernhard RIEMANN was born 17 September 1826 at Breselenz, near Dannenberg, in Lower Saxony, the second son of a Protestant minister, Friedrich Bernhard Riemann, and Charlotte (née Ebell). The children received their elementary education from their father, who was later assisted by a local teacher. Riemann showed remarkable skill in arithmetic at an early age. From Easter 1840 he attended the Lyceum in Hanover, where he lived with his grandmother. When she died two years later, he entered the Johanneum at Luneburg. He was a good student and keenly interested in mathematics beyond the level offered at the school.

In the spring term of 1846 Riemann enrolled at Göttingen University to study theology and philology, but he also attended mathematics lectures and finally received his father's permission to devote himself wholly to mathematics. At that time, however, Göttingen offered a rather poor mathematical education; even Gauss taught only elementary courses. In the spring term of 1847 Riemann migrated to Berlin University, where a host of students flocked around Jacobi, Dirichlet and Steiner. He became acquainted with Jacobi and Dirichlet, the latter exerting the greatest influence on him. When Riemann returned to Göttingen in the spring term of 1849, the situation there had changed as a result of the return of the physicist W.E. Weber. For three terms Riemann attended courses and seminars in physics, philosophy and education. In November 1851 he submitted his thesis on complex function theory, including the idea of what are now called Riemann surfaces, and defended it a month later, thereby earning his Ph.D.

Riemann then prepared for his Habilitation as a Privatdozent, which took him two and a half years. At the end of 1853 he submitted his Schrift on Fourier series and a list, of three possible subjects for his Vortrag. Against Riemann's expectation Gauss chose the third: 'Über die Hypothesen, welche der Geometrie zu Grunde liegen'. It was thus through Gauss's acumen that the splendid idea of this paper was saved for posterity. Both papers were posthumously published in 1867, and in the twentieth century the second became one of the great classics of mathematics. Its reading on 10 June 1854 was a historic occasion: young, timid Riemann lecturing to the aged, legendary Gauss, who would not live beyond the next spring, on consequences of ideas the old man must have recognized as his own and which he had long secretly cultivated. Weber recounts how perplexed Gauss was, and how with unusual emotion he praised Riemann's profundity on their way home.

At that time Riemann was also working as an assistant, probably unpaid, to H. Weber. His first course as a lecturer was on partial differential equations with applications to physics. Further courses in 1855/56, in which he expounded his now famous theory of abelian functions, were attended by C.A. Bjerknes, Dedekind, and Ernst Schering; the theory itself, one of the most notable masterworks of mathematics, was published in 1857. Meanwhile he had also published a paper on hypergeometric series.

When Gauss died early in 1855, his chair went to Dirichlet. Attempts to make Riemann even an associate professor did not succeed until 1857. He only became a full professor after Dirichlet's death in 1859. On 3 June 1862 Riemann married Elise (née Koch) of Korchow, Mecklenburg-Schwerin. The next month he suffered an attack of pleurisy; in spite of periodic recoveries he was a dying man for the remaining four years of his life. His premature death from consumption is usually ascribed to the illness of 1862, but numerous early complaints about bad health and the early deaths of his mother, his brother and his three sisters make it probable that he had been long been a sufferer from tuberculosis. To try and alleviate his condition by moving to a better climate, Riemann took leave of absence from Göttingen and found financial support for a stay in Italy. The winter of 1862/63 was spent in Sicily; in the spring he travelled through Italy, seeing the sights and visiting Italian mathematicians, in particular Betti, whom he had got to know when the latter was at Göttingen.

By June Riemann was back in Germany but his health deteriorated so rapidly that in August 1864 he returned to northern Italy where he stayed until October 1865. He spent that winter in Göttingen, then set out for Italy again in June 1866. On 16 June he had reached Selasca on Lake Maggiore. The day before his death he was lying under a fig tree with a view of the landscape and working on the great paper on natural philosophy which he left incomplete. He was buried in the cemetery of Biganzole.

In the second half of the nineteenth century interest in topology began to spread outside the German-speaking area of Europe. Cayley was reporting on Listing's work to the London Mathematical Society as early as 1869, and Clifford was developing some very significant ideas a few years later. In the case of Italy Betti is an isolated example. In the case of France, Vandermonde must be regarded as a precursor and it is not until we reach Jordan [19] that we find someone who seemed deeply interested in topological questions.

Camille JORDAN was born 5 January 1838 in Lyons, France. One of his granduncles (also named Camille Jordan) was a fairly well-known politician who took part in many events from the French Revolution in 1789 to the beginning of the Bourbon restoration; a cousin, Alexis Jordan, is known in botany as the discoverer of 'smaller species' which still bear his name. Jordan's father, an engineer, was a graduate of the Ecole Polytechnique; his mother was a sister of the painter Pierre Puvis de Chavannes. A brilliant student, Jordan followed the usual career path of French mathematicians from Cauchy to Poincaré: at seventeen he entered the Ecole Polytechnique and was an engineer (at least nominally) until 1885. That profession left him ample time for mathematical research, and most of his 120 papers were written before he retired as an engineer. From 1873 until his retirement in 1912 he taught at both the Ecole Polytechnique and the Collège de France. He was elected a member of the Academy of Sciences in 1881. His most famous contribution to topology was to realize that the observation that a simple closed curve in the plane decomposes the plane into two regions is a result which is capable of being proved and to conceive of such a proof for the first time. His classification of the free loops on a surface, another great achievement, is described elsewhere in this volume. He died in Paris on 22 January 1921.

Although mathematics was flourishing in France throughout the nineteenth century, there were no French topologists apart from Jordan until Poincaré turned his attention to the subject towards the end of his extraordinarily productive career. Single-handed he virtually created the subject in its modern form. Much has been published about his life and work, beginning with [11]. The following short biography gives just the bare outline but there is a separate article on Poincaré's work elsewhere in this volume.

(Jules) Henri POINCARÉ was born 29 April 1854 in Nancy (Lorraine). Poincaré's mathematical ability became apparent while he was still a student at the Lycée. He won first prizes in the concours general (a competition between students from all the lycees of France) and in 1873 entered the Ecole Polytechnique at the top of his class; his professor at Nancy is said to have described him as a 'monster of mathematics'. After graduation he followed courses in engineering at the Ecole des Mines and briefly worked as an engineer while writing his thesis for the doctorate in mathematics which he obtained in 1879. Shortly afterwards he started teaching at the University of Caen, and in 1881 he became a professor at the University of Paris, where he taught until his untimely death. In the same year he married Louise (née Poulain), who bore him a son and three daughters. At the early age of thirty-three he was elected to the Académie des Sciences and in 1908 to the Académie Francaise. He was also the recipient of innumerable prizes and honours both in France and elsewhere. He died in Paris 17 July 1912.

Poincaré left no French school of topology behind him; that developed much later. Dehn and Heegard promoted his work for the German-speaking world in their Enzyklopädie article of 1907. It was also about this time that Vienna began to be recognized as an important centre for topology, under the leadership of Wirtinger [16].

Wilhelm WIRTINGER was born 19 July 1865 in Ybbs, on the Danube in lower Austria, the son of a medical practioner also noted for his research. In his schoolyears he had the opportunity to visit several of the great Benedictine foundations in that part of Austria, and take advantage of their excellent libraries. He entered the University of Vienna in 1884 and in due course was encouraged to visit first Berlin, where he heard Weierstrass, Kronecker and Fuchs, and then Göttingen, where he met Felix Klein, a close friend in later years.

In 1890 Wirtinger married Amalia (née Feyertag), the same year as he took his Habilitation. She bore him three sons, two of whom died in the first world war, and two daughters. He became interested in topology through function theory, where his early work led to his appointment as associate professor at the University of Innsbruck (1895). Eight years later he was appointed to a full professorship at the University of Vienna, where he remained until his retirement in 1935. Although he published comparatively little himself he attracted to Vienna others who were influential in the development of topology in the early twentieth century, such as Tietze and Newman. Among other honours he was awarded the Sylvester medal of the Royal Society of London and was elected to membership of several academies. After retirement he returned to his birthplace of Ybbs, where he died on 16 January 1945.

At this period the German-speaking universities of Central Europe were closely related, and movement between them was encouraged for students and quite normal at faculty level. Vienna-trained topologists, such as Tietze [27], were to be found in a number of universities in Germany and elsewhere. Like Veblen in the United States Tietze was trying to place Poincaré's work on a firmer foundation and to develop it further.

Heinrich (Franz Friedrich) TIETZE was born 31 August 1880 in Schleinz near Vienna. He was the son of Emil Tietze, director of the Geological Institute at the University of Vienna, and of Rosa, daughter of the geologist Franz Ritter von Hauer. He began to study mathematics at Vienna in 1898. Following the advice of his friends Ehrenfest and Herglotz, he moved to Munich for a year in 1902. After returning to Vienna, he worked on his doctoral thesis on functional equations under Escherich, being awarded his doctorate in 1904.

Meanwhile Wirtinger had moved from Innsbruck to Vienna and it was through Wirtinger's lectures on algebraic functions and their integrals that Tietze became interested in topological problems, thereafter the focus of his most important mathematical work. He received his Habilitation at Vienna (1908) with a Schrift 'On topological invariants of multidimensional manifolds', inspired by the work of Poincaré. Two years later he was appointed associate professor of mathematics at the technical college of Brunn (nowadays Brno), being promoted to full professor in 1913.

Tietze's academic career was interrupted by the outbreak of the first world war, when he was drafted into the Austrian army. He returned to Brunn after the war, but in 1919 accepted a full professorship at the University of Erlangen. While at Erlangen he wrote his three-part 'Beitrage zur allgemeinen Topologie'. In 1925 he moved to Munich, where he remained for the rest of his life. Most of his 120 publications were produced there.

Tietze's best-known result is his extension theorem of 1914, but his research in knot theory and other areas of topology is also important. For example, he was the first to show that the first homology group of a space was the abelianization of its fundamental group. With his friend Vietoris he wrote the article on topology for the 1930 edition of the Enzyklopädie, replacing that of Dehn and Heegard. As well as giving an account of what had been achieved in the intervening period the article dealt with the relationship between combinatorial topology and set-theoretic topology and incidentally helped to standardize some of the terminology.

In 1929 Tietze was elected a member of the Bavarian Academy of Sciences. He was also a corresponding member of the Austrian Academy of Sciences and was awarded the Bavarian Verdienstorden. After his retirement in 1950 he continued with research until shortly before his death on 17 February 1964.

In the first quarter of the twentieth century Vienna was exceptionally strong in topology. However, across the Atlantic another place was beginning to establish a reputation in the subject, and by about 1930 Princeton had eclipsed Vienna and become supreme for many kinds of mathematics, especially topology. For what is to follow it is necessary understand something about the way in which a rather unimportant American college developed first into a leading international university [1], at least where mathematics is concerned, and then in partnership with the Institute for Advanced Study [4], consolidated that leadership.

In 1896 the College of New Jersey changed its name to Princeton University, reflecting its ambitions for graduate education and research. When Woodrow Wilson was called to the Princeton presidency in 1903, his first priority was to match the quality of the educational programme to the upgraded status of the university. At Wilson's instigation, the preceptorial system was introduced in 1905 to provide smaller classes and more personalized instruction. Efforts to achieve this in the case of mathematics were placed in the hands of Henry Burchard Fine, the senior mathematics professor, who had studied in Leipzig with Felix Klein and in Berlin with Leopold Kronecker. Those appointed initially included Eisenhart and Veblen, each of whom played a major role in turning Princeton into a world centre for mathematical research and education following the untimely death of Fine in 1928. By then many first-class appointments had been made, for example, Alexander and Lefschetz in topology, but the commanding position of the university was confirmed when a splendid new departmental building, based on Veblen's farsighted ideas of what would be needed, was constructed in memory of Fine and named after him.

By the time the department moved into Fine Hall in 1931 another initiative was under way. This was an institute for advanced research, not part of the University although located at Princeton. Mathematics was the first mission of the Institute for Advanced Study, and Veblen the first Professor, soon to be joined by others, including Alexander. For the first few years they worked in Fine Hall, until Fuld Hall was ready to provide office accommodation and other facilities both for the permanent members of the Institute and for numerous visitors. The atmosphere was very different from Fine Hall, in particular there were no students to be catered for. However, the combination of these two institutions enabled Princeton to become an unrivalled centre for research in the subject.

It was Veblen who introduced Poincaré's ideas to the English-speaking world in his 'Analysis Situs' of 1922. Topology soon caught on in America. Biographies of five American topologists will be given in this section, all associated with Princeton. Of these, Lefschetz [15, 21, 22, 24, 29] and Steenrod [31] were University-based, Morse [6, 28] and Whitney [8, 32] were Institute-based, while Alexander [10, 23] was first at the University and then at the Institute. I begin with Lefschetz and Alexander.

Solomon LEFSCHETZ was born 3 September 1884 in Moscow (Russia), the son of Alexander Lefschetz, an importer, and his wife Verba, who were Turkish citizens. Young Solomon's father's business interests required him to spend much time away in Persia, and he decided to settle his family in Paris, where his children were brought up from a very early age. The boy's first language was French, but he became fluent in Russian and other languages in later years.

There is little on record about the future mathematician's early years in Paris; the first event of note is the award of the degree of 'Ingenieur des arts et manufactures' in 1905, after he had spent 3 years at the Ecole Centrale in Paris, where the professors included Appell and Picard. In November of that year he emigrated to the United States, and after a short apprenticeship became an engineer at the Westinghouse Electric and Manufacturing Co. of Pittsburgh. He was with this firm from 1907 to 1910, but then a promising career in industry was abruptly terminated by an accident at work in which Lefschetz lost both his hands and forearms.

After a period in hospital, he faced up to the fact that his career as an engineer was finished. He decided to change over to pure mathematics, and with this in view he became a graduate student at Clark University in Worcester, Massachusetts, where he took his doctorate (1911) in just one year with a thesis on algebraic geometry. He then occupied a series of positions of increasing seniority first at the University of Nebraska (1911/13) and then at the University of Kansas (1913/25). It was during those 14 years in the prairies that he came to terms with his disability, rebuilding his self-confidence and laying the foundations of a new career. He became an American citizen on 17 June 1912, and in the following year married Alice (née Hayes), who had been a fellow mathematics student at Clark. She helped him overcome his initial despair and face up to life. Later, when sometimes his exuberance burst all bounds, she was equally successful at calming him down.

The major part of Lefschetz' massive contribution to algebraic geometry was completed before he left Kansas. As he has said in [21], his mathematical isolation was complete, and this circumstance was most valuable in that it enabled him to develop his ideas in complete mathematical calm, applying topological methods to the theory of algebraic surfaces. As he put it: 'the harpoon of algebraic topology was planted in the body of the whale of algebraic geometry'. It is not too much to say he arrived in the prairies unknown, and left 14 years later recognized as one of the most outstanding geometers of the day. In 1919 was awarded the Prix Bordin by the Paris Academy and in 1923 the Bocher Memorial Prize of the American Mathematical Society.

In 1924 Lefschetz spent a year visiting Princeton, at the end of which he was appointed to a permanent post as associate professor. Three years later he was promoted to full professor, and in 1932 he succeeded Veblen as Henry B. Fine Research Professor, an office he held until his retirement in 1953. The move to Princeton was a turning-point in Lefschetz' life. The isolation of Kansas was over and he found himself in close contact with the wide circle of mathematicians at Princeton, not only the permanent staff but also the many distinguished visitors who spent periods of leave there, and the able graduate students. It also made it easier for him to travel and visit other universities. That he was able to take full advantage of these opportunities in spite of his physical handicap was due to his indomitable courage.

When Lefschetz moved to Princeton his research interests became centred on algebraic topology. Among the numerous distinguished mathematicians who were around Princeton when he arrived were Veblen and Alexander. The interests of the latter were very close to those of Lefschetz, and although they never wrote a paper together they frequently discussed such matters of common interest as fixed-point theory and duality in topology. Lefschetz was a great admirer of Alexander, and in later years was greatly saddened when Alexander gradually withdrew from contact with mathematicians and became a recluse.

Lefschetz' main contribution to mathematics during the thirties lay in his powerful influence on others. He worked very hard to keep himself informed on what his students and associates were doing, and was a vigorous critic of anything he did not approve of. He employed equally drastic methods in his capacity as editor of the Annals of Mathematics over a period of 25 years. No leniency was shown towards any paper which was submitted to the journal which was not up to his standards. He tended to make up his mind in a flash and anyone who disagreed with his judgement had to work very hard to make him change his mind. By these methods he made the Annals one of the top journals in the world, and he and his colleagues made Princeton a world centre for mathematics. In the course of this vigorous programme he made very few enemies: it was felt that there was no personal animosity in his bark, and no self-seeking: he just wanted to serve mathematics as best he could.

During the second world war, when he was acting as a consultant to the US Navy, Lefschetz came across Russian work on nonlinear oscillations and stability. He immediately recognized the importance of the work of Poincaré and Liapunov on the geometrical theory of differential equations, and saw that the subject had been 'too long neglected' in the United States. After the war, with the support of the Office of Naval Research, he organized a differential equations project at Princeton University, which became the leading centre of research in ordinary differential equations in the United States.

In a related development Lefschetz built up the Research Institute for Advanced Studies in Baltimore into an outstanding example of support for basic research by industry. In 1964 part of the Institute moved to Brown University, at Providence, Rhode Island, and became the Lefschetz Centre for Dynamical Systems. For 6 years he commuted weekly by plane from Princeton to Providence, where he lectured, discussed research, and spread his wit, enthusiasm and love of life and mathematics. But this was not all: during the same period he was also making frequent visits to the National University of Mexico, in Mexico City, where his enthusiasm, drive and organizing ability contributed greatly to the establishment of a lively school of mathematics.

Lefschetz' contribution to mathematics was recognized by his election to the National Academy of Sciences of Washington in 1925, and to the Presidency of the American Mathematical Society ten years later. He received honours from numerous universities and learned societies. He died after a short illness in Princeton on 5 October 1972.

James (Waddell) ALEXANDER was born 19 September 1888 in Seabright, New Jersey. His father, John W. Alexander, was a noted American painter of the last century; his mother was an active patron of the arts. Alexander himself received his early education in France and at the Browning school in New York. After a distinguished undergraduate career at Princeton University he graduated in 1910, and received his doctorate five years later. He remained at Princeton as an assistant in mathematics until 1917. In that year he married Natalie (née Levitzkaya), a Russian who he had met in Italy, and he volunteered for military service in the first world war. Attached to the technical staff of the Ordnance Department, he was stationed in Washington and later in France.

After the war Alexander returned to Princeton, where he was appointed full professor in 1928. Five years later he moved to the newly-founded Institute for Advanced Study. Alexander was fortunate in his scientific development in that he came under the guidance of Veblen and, no doubt under Veblen's perceptive influence, directed practically all of his own scientific endeavours towards the still young science of combinatorial topology, very largely created by Poincaré.

What Poincaré contributed to the subject was immense but not always supported by a strong logical base. The first contributions of Alexander (in collaboration with Veblen) was to provide the subject with a reasonable element of logic. This led to the famous paper establishing the topological invariance of the Betti numbers. Although he only dealt with the three-dimensional case, the argument he gave is valid in all dimensions. When Poincaré was studying the homology of manifolds he at first thought that these invariants might be sufficient to classify manifolds of a given dimension up to homeomorphism. When he discovered that 3-dimensional manifolds exist which have the same homology as the 3-sphere but different fundamental groups he conjectured that homology and the fundamental group might be sufficient. In 1919 Alexander found a family of 3-manifolds, the lens spaces, which provided counter-examples.

This was soon followed by the famous Alexander duality theorem of 1920, later extended in various directions, notably into the Pontrjagin duality theorem. It is of great importance, not only in itself but also because there are contained within it certain anomalies whose resolutions were major influences in developing homology theories with different coefficients and in developing cohomology theories.

In the 1920's Alexander was becoming increasingly interested in knot theory, and in 1925 he made a fundamental discovery. To a given knot diagram he associated a matrix of polynomials and showed that the equivalence class of this matrix (equivalence having a slightly more extended meaning from the normal one) is an invariant of the knot type. From this matrix equivalence class he extracted by essentially classical means a sequence of polynomials, particularly one now called the Alexander polynomial, which is such a sensitive invariant that it readily distinguishes most of the knots found in the knot tables compiled in the last century. Other polynomial invariants of knots were discovered not long ago.

In the mid-thirties Alexander played an important role in the development of the idea of the cohomology ring structure. Although he continued to pioneer new concepts for another decade, his great work was complete. He left the field to the suddenly numerous rising generation and, although living on the edge of the campus, seldom appeared in Fine Hall. In this younger days Alexander was well-known as an accomplished alpinist, but after an attack of polio he had to give up this activity. An exceptionally shy man, he loved music, photography and amateur radio. Alexander was elected to the American Philosophical Society (1928) and to the National Academy of Sciences of Washington (1930). He received honorary degrees from the universities of Bologna and Paris in 1947. He retired in 1951, and died 23 September 1971 at Princeton.

Hassler Whitney [32] has this to say about the relationship between Alexander and Lefschetz. 'They naturally had many discussions on topology. But Alexander became increasingly wary of this; for Lefschetz would come out with results, not realizing they had come from Alexander. Alexander was a strict and careful worker, while Lefschetz's mind was always full of ideas swimming together, generating new ideas, of origin unknown. I believe that Lefschetz never felt good about Veblen choosing Alexander, not himself, as one of the first professors at the new Institute for Advanced Study.'

(Harold Calvin) Marston MORSE was born 24 March 1892 in Waterville, Maine, where he had his early education and where he completed his undergraduate work in 1914 at Colby College. Three years later he received his Ph.D. from Harvard under G.D. Birkhoff, having meanwhile published his first research paper in 1916. His career was interrupted by the first world war. He served with distinction in the American Expeditionary Force and was awarded the Croix de Guerre with Silver Star for bravery under fire. Resuming academic life, Morse taught at Harvard 1919/20, Cornell 1920/25 and Brown 1925/26, before returning to Harvard as assistant professor (1926), associate professor (1928) and finally full professor (1930). Five years later he moved to the Institute for Advanced Study. Mathematics to Morse was a highly competitive enterprise; priority in publication was important to him. He needed an audience. As a consequence he sought collaborators and assistants, a substantial function of these individuals being to listen to his explanations of mathematical situations as he perfected his understanding of them.

In the second world war he served as consultant in the Office of the Chief of Ordnance. His invaluable work on military applications of mathematics was recognized by a Meritorious Service Award, conferred in 1944 by President Roosevelt.

After the war he was the prime mover in the establishment of the National Science Foundation. President Truman invited him to serve on its first board from 1950 to 1954. He represented the Vatican at the Atoms for Peace Conference of the United Nations (1952). He was president of the American Mathematical Society 1940/42; a vice-president of the International Mathematical Union starting in 1958; chairman of the Division of Mathematics of the National Research Council 1951/52, and so on. In his local community he served on the board of two private schools and an organisation concerned with making records for the use of the blind.

Among the many honours bestowed on Morse were honorary degrees from twenty institutions in the USA, Austria, France and Italy. These include the University of Paris (1946), Pisa (1948), Vienna (1952), Harvard (1965) and Modena (1975). In 1952 he became a Chevalier of the French Legion of Honour. He was elected in 1932 to the National Academy of Sciences of Washington and in 1956 as an associate member to the French Academy of Sciences. His affinity for France made the honours from that country particularly gratifying. He also cherished his election as a corresponding member of the Italian National Academy Lincei. A National Medal of Science was awarded to him in 1964 and presented by President Johnson at the White House.

Morse became emeritus in 1962 but for the remainder of his eighty-five years he continued his research activity. Essential to the remarkable prolongation of his long and brilliant career was the devoted care and understanding of his second wife Louise. As well as their five children there survives one of the two offspring of his first marriage. He died in Princeton 22 June 1977.

We now return to Europe. Until recently, topology in Eastern Europe followed a distinctive tradition, less dominated by the ideas of Poincaré than topology in Western Europe and America. In fact if topology is divided into point-set topology on the one hand and combinatorial topology on the other the emphasis in Eastern Europe was more on the former. One of the leading topologists of Eastern Europe was Čech [7].

Eduard ČECH was born 29 June 1893 in Stracov in northeastern Bohemia. He was the fourth child of Cenek Čech, a policeman, and Anna (née Kleplova). After studying at the Gymnasium in Hradec Kralove, he went to the Charles University in Prague to study mathematics in 1912. However, his university studies were interrupted by the first world war, and he did not graduate until 1920. By then he had become interested in projective differential geometry, the study of those features of the geometry of embedded curves, surfaces, and higher-dimensional spaces that are projectively invariant. Čech's work emphasized results about tangency, correspondences between manifolds, and (significantly for his later work in topology) the systematic theory of duality in projective spaces. He spent 1921/22 working with Fubini in Turin, and later they collaborated on two books on projective differential geometry.

In 1922 Čech returned to his native country and, after receiving his Habilitation he took up an appointment at the Masaryk University in Brno. He was promoted to full professor (1928) and, inspired by the papers in the Polish journal Fundamenta Mathematica, his research interests increasingly turned to topology. His first contributions to topology were characteristically broad and aimed at keeping the subjects of algebraic and point-set topology together. In his first two papers he developed a homology theory for arbitrary spaces, and established general duality theorems for manifolds, generalizing the classical duality of subspaces of projective spaces. Čech's approach to homology was deliberately intended to be very general, as the title of his 1932 paper in Fundamenta Mathematica makes clear ('General homology theory of an arbitrary space'). This is based on the idea of studying all the finite open coverings of a given space. Although many of the ideas in it can be traced back to earlier work of Vietoris and Aleksandrov, Čech's originality lay in using inverse limits to obtain homology groups independent of the choice of covering. This approach turns out to work well for compact spaces, and yields what is now called the Čech homology theory. The corresponding cohomology theory works less well for noncompact spaces, giving unexpected results even for the first cohomology group of the real line. Later Dowker replaced Čech's finite coverings with arbitrary coverings, and showed that with this modification, Čech cohomology satisfied all the Eilenberg–Steenrod axioms for a cohomology theory.

At the 1932 International Congress of Mathematicians in Zurich, Čech presented his ideas on the definition of the higher homotopy groups of a space. Unfortunately, he was discouraged from pursuing the subject. Independently Hurewicz, a few years later, recognized the importance of these invariants and proved the basic theorems about them. Aleksandrov was to single out Čech's contribution, when commemorating Čech's life and work in 1961, and to lament that it had been misunderstood.

After Čech reported on his researches at the 1935 international conference on topology held in Moscow, Lefschetz invited him to visit the newly-founded Institute for Advanced Study in Princeton. When he returned from there Čech founded a topology seminar in Brno which applied itself to the work of Aleksandrov and Uryson. In three years the seminar published 26 papers, including Čech's 'On bicompact spaces', which appeared in the Annals of Mathematics. In this paper he introduced the idea of the (Stone)–Čech compactification of a regular space. The seminar continued at the University until 1939, when the Germans invaded Czechoslovakia and closed the universities down. Thereafter it continued in the flat of Čech's student B. Pospisil until 1941, when the Gestapo arrested Pospisil. The seminar had a lasting influence on the development of mathematics in Czechoslovakia, because it introduced the practice of working on mathematical problems collectively.

After the war, Čech returned to the Charles University in Prague. By then in his fifties, he began an intensive career in administration. He was appointed Director of the Mathematical Research Institute of the Czech Academy of Sciences and Arts in 1947 and three years later became Director of the Central Mathematical Institute. In 1952 the Central Institute was incorporated into the Mathematical Institute of the Czechoslovak Academy of Sciences, with Čech as its first Director, and he also became head of the new Mathematical Institute at the Charles University. Nevertheless he also found time to redirect his mathematical interests; in the 1950s he wrote 17 papers on differential geometry. He also deepened his interests in the teaching of mathematics. He wrote seven textbooks for secondary schools and held seminars on elementary mathematics at both Brno and Prague. Čech continued to be active in mathematical life in Czechoslovakia until his death on 15 March 1960.

During the twenties and thirties a number of textbooks and monographs on topology were published. Several of these were written by Reidemeister [2].

Kurt (Werner Friedrich) REIDEMEISTER was born 13 October 1893 in Brunswick (Germany), the son of Hans Reidemeister and Sophie (née Langerfeldt). He went to school in Brunswick and then studied at the universities of Freiburg, Munich and Göttingen. The young man's studies were interrupted by four years of military service during the first world war. When that was over he returned to Göttingen and passed the Staatsexamen (Edmund Landau was his examiner) in mathematics and other subjects in 1920. After becoming assistant to Hecke at the University of Hamburg, he obtained his doctorate (1921) with a dissertation on algebraic number theory. He was also working on affine and differential geometry at this time.

In 1923 Reidemeister was appointed assistant professor at the University of Vienna, where he came into contact with Hahn and Wirtinger, amongst others. Two years later he became a full professor at Königsberg, where he worked with other young mathematicians, notably the algebraist Richard Brauer. Reidemeister's main research interests at this stage were in combinatorial topology and the foundations of geometry. The historical origin of mathematical and rational thought always fascinated him. Due to his opposition to the Nazis Reidemeister was expelled from his Königsberg professorship in April 1933. The previous year he had published his well-known monograph 'Knotentheorie', which remained the standard work on the theory of knots for several decades. From Königsberg he moved first to Marburg, a less important university, and then, in 1955, back to Göttingen, where he remained (apart from two separate years at the Institute for Advanced Study in Princeton) until he died on 8 July 1971. His wife Elizabeth (née Wagner), the daughter of a Protestant pastor at Riga, was a professional photographer, whose portraits of Dehn and Seifert illustrate their biographies elsewhere in this volume.

Topology arrived in the British Isles quite early, as we have seen, and there was particular interest in knots and graphs. However, the subject did not begin to flourish until the twenties when Newman [14] visited Vienna and came into contact with Wirtinger and others at a time when that university was exceptionally strong in topology.

Maxwell Herman Alexander NEWMAN [14] was born 7 February 1897 in Chelsea, London. His father, who had come from Germany (with the name of Neumann), was secretary of a small company; his mother was a farmer's daughter who had trained as a teacher and taught at elementary schools in London. Their only child, known to everybody as Max, was educated at the City of London School. He wrote very highly of the mathematical teaching he received there from a former fellow of St. John's College, Cambridge. In due course Newman won an Entrance Scholarship to that college.

Newman came up to Cambridge in October 1915, and resided until December 1916. He spent the next three years in war service. After his father was interned as an 'enemy alien' the young man changed his surname from Neumann to Newman by deed poll in 1916. At some point in this period he had a spell as a schoolteacher, and at another, in spite of poor eyesight, he served in the army as a paymaster.

When Newman returned to Cambridge in October 1919, to complete his studies, college teaching at St. John's was in the hands of Bromwich and Cunningham, both of whom Max regarded as exceptionally good teachers. Of the lecturers he thought Hardy stood out. In 1916 Newman had achieved first class honours in the first part of the Mathematical Tripos and now in 1921 he was equally successful in the second part. Two years later he was elected to a Fellowship at St. John's which he retained until 1945.

Newman spent the year 1922/23 at the University of Vienna, at that time one of the leading centres in Europe for topological research. He was strongly influenced by the ideas of Wirtinger, Hahn and Reidemeister. Afterwards he worked on a wide range of subjects, but he made his name by his early work on the foundations of combinatorial topology. Later in this period Newman worked on the Poincaré Conjecture and on the Hauptvermutung, but did not publish his findings. He spent 1928/29 in Princeton.

Newman had been appointed University Lecturer at Cambridge in 1927. He had a pioneering attitude to syllabus reform and was the first in Cambridge to use abstract vector spaces in presenting linear algebra. He was also the first to lecture on Gödel's theorem. About 1938 he started a joint seminar with Philip Hall on algebra and topology, which played a part in introducing the axiomatic point of view into Cambridge mathematics. He also wrote a textbook 'The Topology of Plane Sets' which some consider a minor masterpiece, and one of the best introductions to point-set topology considered as a part of mathematics as a whole.

In December 1934 he married Lyn (née Irvine), an authority on medieval poetry who also enjoyed some success as an author. They made their home at Cross Farm, Comberton, outside Cambridge. They had two sons, Edward (born October 1935) and William (born May 1939).

In May 1942 Newman was approached to work at a government institution on a matter whose importance to the war effort could be hinted at, although its nature could not. This turned out to be the celebrated code-breaking centre at Bletchley Park, midway between Oxford and Cambridge, where a remarkable group of British mathematicians not only succeeded in the immediate task but also talked about what they would do when the war was over. Newman entered into this important work wholeheartedly and found it deeply interesting.

In September 1945 Newman resigned his positions in Cambridge and went to Manchester to succeed Mordell as Fielden Professor of Mathematics. It was as leader and manager of a mathematics department that Newman showed talents rarely equalled among pure mathematicians. The list of those he recruited displayed exceptional judgement of ability. From 1945 to 1964 he raised his department to a very high standard among British mathematics departments. At the same time his own research was not neglected. Interest in combinatorial and geometric topology revived through the work of Moise, Mazur, Smale and others. From 1960 to 1966 Newman published work of a quality seldom achieved by mathematicians in their sixties.

Newman continued a Manchester tradition of support and hospitality to refugee mathematicians, such as Paul Erdős, Bernhard Neumann, Kurt Mahler, and Beniamino Segre. In contrast to some other mathematicians he was not a great traveller himself. He visited Princeton as a Rockefeller Research Fellow in 1928/29 and returned there in 1937/38. However, when the time came for him to lay down his responsibilities at Manchester Newman decide to spend a few years abroad and did so at the Australian National University, Canberra, and at Rice University, Houston.

The Newmans had kept their main home at Cross Farm, in fact their only home after he retired from Manchester, and their social life was based around Cambridge to a large extent. In 1973 Lyn died after a short illness. Later that year Newman married Margaret, the widow of the distinguished Cambridge scientist Lionel Penrose. Just over ten years later he died 22 February 1984.

Although the Swiss contribution to topology is important, after Euler it is not so easy to name topologists of Swiss origin who satisfy our criteria. This may be because they tended to be geometers first and topologists second. This was true of Schläfli, who was notable for his generalization of the Euler formula, and it was also true of de Rham. There does not appear to be a biographical memoir of the latter but some information can be found in a lecture 'Quelques souvenirs des années 1925–1950' which he gave in 1980 and in a memorial booklet 'Georges de Rham 1903–1990', privately published by his friends and colleagues.

Georges DE RHAM was born 10 September 1903 at Roche (Vaud), where his father Leon occupied the post of engineer in an important cement works. He went to school at Aigle, on the edge of the Valais, and by the time he was fifteen had already developed the passion for mountaineering which remained with him throughout his life.

In 1919 the de Rham family left Roche for an apartment in the Chateau de Beaulieu in Lausanne, and the future mathematician attended the Gymnasium of that city until 1921, when he proceeded to the University of Lausanne. There his studies, always in scientific subjects, increasingly tended towards mathematics. After graduating he continued at the university as an assistant, becoming particularly interested in the work of Poincaré, Lebesgue and Elie Cartan, especially the last. In the following years he was developing the theory to which his name is now invariably attached and which he presented as his doctoral thesis at Paris in 1931 under the title 'Sur l'Analyse Situs des Variétès à n Dimensions', in which he showed that the real cohomology of a differential manifold could be defined using differential forms. This celebrated result established his reputation internationally. After a few years as lecturer he was appointed associate professor at Lausanne in 1936 and then, almost immediately, to a similar post at the university of Geneva, an extraordinary combination. He was promoted to full professor at Lausanne in 1943.

During these years de Rham was becoming equally celebrated as an alpinist. He liked to tell how on one occasion on Mont Blanc he accidentally met the topologist Alexander and another time on the Weisshorn he met him again with a young companion who turned out to be Hassler Whitney. Among a long list of spectacular climbs in the Alps perhaps the one he was most proud of was his conquest of the south face of the Taschhorn with André Roch in 1943. He died at Lausanne 7 October 1990.

During the thirties, as the situation in Europe became increasingly threatening, a number of European mathematicians migrated to America. Even with the assistance of Veblen and other influential well-wishers it was not easy for them to find suitable positions in the aftermath of the economic depression. Among the topologists Dehn was one of the older migrants; there is an account of his life in the next section. Hurewicz [5, 20] was one of the younger migrants, who generally found positions more easily.

Witold HUREWICZ was born 29 June 1904 in Lodz (Poland), the son of a well-to-do industrialist. After graduating from the Lodz high school he went to study at the University of Vienna, at that time a leading centre for mathematics, and by 1926 he had obtained his Ph.D. under Hans Hahn. His early research was on dimension theory, particularly the extension to separable metric spaces of the results established by Brouwer, Menger and Uryson for subsets of Euclidean space. When Menger moved to Amsterdam Hurewicz followed him, and remained there until 1936. After the first year, during which he was supported by a Rockefeller fellowship, Hurewicz became lecturer and assistant to Brouwer, who was not only one of the creators of dimension theory but also of homotopy theory. Brouwer thought Hurewicz was a genius who might turn out to be a second Riemann or Poincaré. For some years Hurewicz added nothing to his earlier publications and it was not until 1935/36 that there appeared, in the Proceedings of the Royal Academy of Amsterdam, an amazing series of four short papers which set homotopy theory in motion. In these he defined the higher homotopy groups and discovered the fundamental theorem which links them to homology groups. The definition of the higher homotopy groups had occurred to others, in particular to Čech and Dehn, but they had not developed the idea. The last of these papers, on aspherical spaces may be regarded as the starting point of what became known as homological algebra.

These achievements led to a fellowship at the new Institute for Advanced Study in Princeton, where Hurewicz spent the years 1936/39. In the years 1939/45 he held a professorship at the University of North Carolina, Chapel Hill, but he was also working at Brown University and the Radiation Laboratory at Cambridge, Massachusetts, during the second world war. During this period he published an abstract, in which the notion of exact sequence appears for the first time, and an important joint paper with Steenrod containing a theory of fibre spaces; a similar theory had been developed independently by Eckmann in Europe.

At the end of the war Hurewicz was appointed to a professorship at the Massachusetts Institute of Technology, where he remained for the rest of his life. Unfortunately, Hurewicz had a tendency not to write up his ideas in detail. He intended to develop them into a book, which would cover the whole of homotopy theory as it then was. However, the theory was developing rapidly and so he was never able to complete it. The incomplete text seems to have been destroyed in a fire not long after his death. Fortunately, his earlier work, on dimension theory, led to a very successful book on the subject, co-authored with

Henry Wallman, published in 1941 and still the classic text over 50 years later. He wrote a few other papers, including an important one on ergodic theory, but his complete list of publications extends to no more than 14 items.

Mathematics was by no means his only interest. As well as music and literature he enjoyed learning languages, being fluent in English, French, Russian, German, Dutch and, of course, his native Polish, which he often used in his private life. Although he took out American citizenship he remained a European gentleman, most at home in a circle of close relatives and old friends of a similar cultural and social background. His untimely death, on 6 September 1956, occurred after he had been attending the important international symposium on algebraic topology which was held at the National University of Mexico that year. Hurewicz was taking the opportunity to visit the Mayan archaeological sites of Yucatan, and it was when he was exploring the ruins at Uxmal that he suffered a fall, sustaining injuries which proved fatal.

The British school of topology was founded by Newman, as we have seen. However, it was Whitehead [13, 25, 26, 34] who built it up into a school of major importance, especially after the second world war.

(John) Henry (Constantine) WHITEHEAD was born 11 November 1904 in Madras (India). His father was Bishop of Madras, and his mother the formidable Isabel Duncan, had been a mathematical scholar at Oxford; the mathematician and philosopher Alfred North Whitehead was his uncle. He was educated at Eton College and at Balliol College, Oxford, and was very proud of it. His great mathematical gifts did not include the knack of carrying out manipulations correctly, so that he was no schoolboy prodigy. At Oxford, however, he achieved high honours in both the mathematical examinations and his college awarded him an honorary scholarship.

Perhaps the mathematics taught to undergraduates at Oxford at that period had too strong a flavour of problem-solving to fire him. At any rate after he graduated Whitehead spent several years working at Buckmaster and Moore, a London firm of stockbrokers. Fortunately for mathematics he found this way of life unsatisfying and in 1928 began a new career by winning a Commonwealth Fellowship to study differential geometry at Princeton under Veblen. Differential geometry was a natural choice at that time, but he may have been influenced towards it by lectures on relativity he attended at Oxford. He worked mainly on differential geometry until about 1932, collaborating with Veblen in writing a monograph on the foundations of that subject.

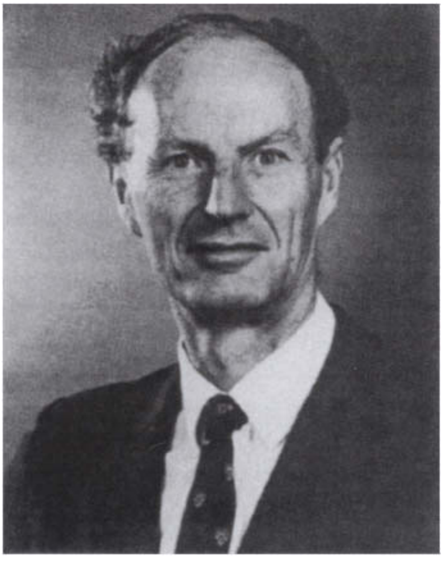
A number of important things happened to him in the next few years. In 1933 Balliol made him a fellow and tutor in mathematics. In 1934 he was married to Barbara (née Smyth), a concert pianist; together they made their successive homes places of welcome and entertainment for a host of friends. It was in the middle 1930's too that he began to publish papers in topology and algebra. He was drawn towards topology not only because it was then a great geometrical arena but also by his friendship with Newman; his early papers show clearly the influence of Newman's work and also that of Reidemeister. This early work, up to the time of the second world war, contains some of Whitehead's most original contributions to mathematics, the importance of which was not fully appreciated until much later.

During the war, after a spell at the Admiralty he joined Newman and other mathematicians at the celebrated code-breaking establishment in Bletchley. There was one useful thing, he said, he had learnt from his war-work, namely the policy of abandoning unpromising projects ritually and finally.

Henry Whitehead (1904–1960)

Witold Hurewicz (1904–1956)

Norman Steenrod (1910–1971)



Frank Adams (1930–1989)

In 1944 Whitehead had been elected a Fellow of the Royal Society; three years later he was appointed Wayneflete Professor of Pure Mathematics at Oxford and thereby became a Fellow of Magdalen College. Meanwhile, as he found at mathematical gatherings, he was beginning to be recognized as one of the world leaders in topology, attracting able research students and a stream of academic visitors so that Oxford became a Mecca for topology.

His work after the war was at first mainly devoted to a striking series of results on homotopy classification and to the development of his important pre-war papers. Although most of his earlier topological research had firm geometrical roots, he began to turn increasingly to algebraic methods. By the end of the fifties, however, exciting progress again began to be made with some of the classical problems of geometric topology. He had the satisfaction of seeing some of his earlier work playing an important part in these developments, and to participate himself. In 1959 he returned to Princeton for a sabbatical year. He was apparently in full vigour and spirits when in 11 May 1960 he collapsed and died, leaving a widow and two sons.

During the thirties a distinctive school [9] of topology was developing in Poland, especially Warsaw. Point-set topology was an important component of the work of this school but under the leadership of Borsuk [12] an alternative theory to combinatorial topology was created.

Karol BORSUK was born 8 May 1905 in Warsaw, the son of the well-known surgeon Marian Borsuk and Zofia (née Maciejewska). He received a master's degree (1927) and doctorate (1930) from Warsaw University, and became Privatdozent there in 1934. On 26 April 1936 he married Zofia (née Paczkowska); they had two daughters. Borsuk's principal nonacademic interests were reading and travelling, usually accompanied by his wife. At their country cottage 'Radachowka', some forty miles from Warsaw, they often entertained mathematical friends.

During the Nazi occupation of Poland Borsuk strove to keep intellectual life in Poland alive through an 'underground university'. Together with other 'illegal' activities this led to his imprisonment but he managed to escape and remained in hiding until the war ended. When Poland began to rebuild, Borsuk and Kuratowski began the work of restoring mathematical research in Warsaw. Borsuk was appointed professor at the university in 1946, director of the Mathematical Institute there from 1952 to 1964, and deputy director of the Institute of Mathematics of the Polish Academy of Sciences in 1956. He made several visits to the United States: the Institute for Advanced Study 1946/47, the University of California at Berkeley 1959/60 and the University of Wisconsin at Madison 1963/64. He was a corresponding member of the Polish and Bulgarian Academies of Science. He died in Warsaw 24 January 1982.

Although all the American topologists whose lives are discussed in this section were associated with Princeton it is only fair to say that some of their best work was done elsewhere. Whitney is a case in point.

Hassler WHITNEY was born 23 March 1907 in New York City. His grandfather was Simon Newcomb, a noted astronomer and the fourth President of the American Mathematical Society. His father Edward Baldwin Whitney was a Justice of the Supreme Court of New York, and his mother Josepha was an artist and active in politics. The young man graduated from Yale University with bachelor's degrees in physics (1928) and music (1929), and went on to Harvard University to obtain his doctorate in mathematics (1929). In 1931 he was awarded a National Science Foundation Fellowship to go to Princeton for two years. He returned to Harvard in 1933, where he advanced to the rank of professor. In 1952 he moved to the Institute for Advanced Study in Princeton as a Professor of Mathematics, a position he held until he became emeritus in 1977.

A pioneer in topology, Whitney combined very fruitful perspectives with great technical prowess. The ideas and methods he developed in the general theory of manifolds, the study of differentiable functions on closed sets, geometric integration theory, the geometry of tangents to a singular analytic space, as well as many others, have become a part of the very fabric of these subjects and have had a tremendous influence on subsequent work.

The celebrated Whitney Embedding Theorem was an important conceptual advance in the understanding of manifolds, for it ties together the extrinsic and intrinsic definitions. The theorem states that any n-dimensional manifold can be embedded in m-dimensional Euclidean space, for m sufficiently large; later he showed that $m = 2n$ was sufficient.

Whitney was also one of the founders of cohomology theory. Along with Čech he formulated the first clear and correct definition of the cup product in cohomology. He also pioneered the use of the vector and sphere bundles as a tool in the solution of topological problems. Stiefel–Whitney classes, which are important invariants of vector bundles, were discovered independently by Whitney and Stiefel around 1935.

Throughout most of his career Whitney was interested in the properties of smooth functions. His ideas were instrumental in the development of the field of differential topology, which in turn led to his work on the theory of analytic varieties. He helped to launch the theory of singularities and established that the generic singularities of maps from the plane to the plane are folds and cusps.

Whitney was intensely concerned with what he saw as the failure of the American educational system in mathematics and during the last twenty years of his life he pursued this concern with great energy, particularly on the elementary school level. The many honours and awards he received included election to the National Academy of Sciences of Washington (1945), the Wolf Foundation Prize (jointly), and the United States National Medal of Science. As well as music his nonmathematical pursuits included roller-skating, on which he was something of an authority, and mountain-climbing, particularly in the Swiss Alps. He died 10 May 1989 in Princeton at the age of 82, following a massive stroke; his ashes were placed at the summit of one of the Alpine peaks he loved.

Under R.L. Wilder the University of Michigan at Ann Arbor became an important centre for research in topology and for bringing on the young. Steenrod was one of his later students.

Norman (Earl) STEENROD was born 22 April 1910 in Dayton, Ohio, the youngest of three surviving children of Earl Lindsay Steenrod and his wife Sarah (née Rutledge). The Steenrods, reputedly of Norwegian origin, came to the United States by way of Holland before the Revolutionary War, and Norman Steenrod's ancestor Cornelius Steenrod, raised a company of soldiers who fought in that war. Both his parents were teachers – his mother for two years before her marriage, his father for some forty years as a high school instructor in manual training and mechanical drawing. Neither parent had any special interest in mathematics, although Earl Steenrod had a keen interest in astronomy, which he communicated to his son. From his mother Norman Steenrod acquired a lifelong interest in music, to which he devoted much of his spare time. Other interests included tennis, golf, chess, and bridge.

Steenrod attended the public schools in Dayton, finishing the twelve-year course in nine years. After graduation from high school he worked for two years as a tool designer, having learned the trade from his elder brother, and he did so again later. In this way he earned enough to help with his college expenses first at Miami University in Oxford, Ohio, (1927/29) and then at the University of Michigan (with interruptions) until 1932.

At Ann Arbor Wilder's course in topology was the only mathematics course that he enrolled in, all the others being in physics, philosophy and economics. The year 1932/33 was a hard one for Steenrod: unable to secure a fellowship, he went back home to Dayton, where he started on a problem given to him by Wilder. By the end of the year he had finished his first paper, on the strength of which Harvard, Princeton and Duke all offered him fellowships. He decided on Harvard and, to help meet his expenses there, worked for a time at the Flint Chevrolet plant as a die designer. This enabled him to spend a year in Cambridge. Next, in the spring of 1934 he was again offered fellowships at both Harvard and Duke, but turned these down when a similar offer arrived from Michigan. However, as it happened Wilder was going to spend that year at Princeton and decided not to leave Steenrod behind. With Lefschetz' support Wilder was able to persuade the Princeton fellowship committee to make an offer to Steenrod, although it took some persuasion to get him to accept.

By this time Steenrod's financial problems had eased. At Princeton he worked under Lefschetz, obtaining his Ph.D. in two years. He remained at Princeton as an instructor for three more years. In 1938 Steenrod married Carolyn (née Witter), and they moved to Chicago the following year. Their first child, Katherine Anne, was born there in 1942. However, Steenrod felt a strong attachment to Michigan and was reluctant to raise a family in a large city like Chicago, and so they returned to Ann Arbor the same year. Five years later their other child, Charles Lindsay, was born. Just afterwards the family moved back to Princeton, and Steenrod spent the rest of his career there.

Steenrod's work in algebraic topology is probably best known for the algebra of operators which bears his name. The problem of classifying by homotopy the maps of a complex into a sphere had long occupied the attention of topologists. The case where the complex has the same dimension as the sphere is classical, the work of Hopf in 1933, although subsequently improved upon. In 1942 Steenrod solved the next case, where the dimension of the complex may exceed that of the sphere by one. His solution was not only interesting per se, but by virtue of the new operations in terms of which the solution was expressed. These were the celebrated Steenrod squares. The power of the operations soon became apparent, for example, in relation to the famous problem about vector fields on spheres. The squares were defined in mod 2 cohomology; similar operations for mod p cohomology, p odd, were defined before long. Relations involving these families of operations were discovered by the Mexican mathematician José Adem, one of many able students who took their Ph.D. under Steenrod.

Algebraic topology underwent a spectacular development in the years following the second world war. From a position of minor importance, as compared with the traditional areas of analysis and algebra, its concepts came to exert a profound influence, and it is now commonplace that a mathematical problem is 'solved' by reducing it to a homotopy-theoretic one. To a great extent the success of this development can be attributed to Steenrod's influence.

As well as his early research on point-set topology there are two other aspects of Steenrod's work which have had a profound and lasting influence on the development of topology. One was his deep interest in the theory of fibre bundles, which culminated in his classic book, the first attempt to organise this important subject. The other was his interest in homology theories, which led to the book on the foundations of algebraic topology which he wrote with Eilenberg. To this might be added the monograph on cohomology operations, based on his Princeton lectures, and a useful compilation of reviews of all papers in algebraic topology and related areas.

Steenrod was elected to the National Academy of Sciences of Washington (1956) and gave the prestigious Colloquium Lectures for the American Mathematical Society (1957). In the spring of 1971, while on sabbatical leave at Cambridge University, he suffered an attack of phlebitis and, after his return to Princeton that fall, the first of a succession of strokes, to which he finally succumbed on 14 October 1971.

There are many examples of mathematicians of European origin who emigrated to North America. Dowker [30] is one of the few who moved in the other direction.

(Clifford) Hugh DOWKER was born 2 March 1912 at Parkhill, Western Ontario, and grew up in a rural community, where his family owned a small farm. His ancestors on his father's side were of Yorkshire origin, while his mother was of Scottish descent. Hugh was the first member of the family to attend high school. His elder brother Gordon left school at the age of thirteen and worked as a forester, while his younger brother Arthur followed the family tradition by going into farming.

When Dowker was seventeen he went to the University of Western Ontario on a scholarship, intending to become a schoolteacher. He studied a variety of subjects, including physics and economics, but his particular gift for mathematics was already evident. This was such that, after obtaining his BA (1933) he was encouraged to continue his studies at the University of Toronto, where he gained his MA the following year. He was then advised to go on to Princeton to study under Lefschetz. It was at Princeton that Dowker became fully aware of the power and beauty of mathematics. He specialized in topology and ran one of Lefschetz' seminars, obtaining his Ph.D. in 1938. Apart from Lefschetz, the mathematicians who were to have an important influence on Dowker's research included Aleksandrov, Fox, Hurewicz and Steenrod.

Dowker's first academic post was that of instructor (1938/39) at the University of Western Ontario, where he had earlier been a student. Next he moved back to Princeton to become assistant (1939/40) to von Neumann at the Institute for Advanced Study, after which he became an instructor (1940/43) at Johns Hopkins University in Baltimore. It was there that he met Yael Naim, who he married in 1944. Yael at the time was a young graduate student who had come to Johns Hopkins from Israel, and who was to become well-known for her work on ergodic theory.

In 1943 Dowker was seconded to the United States Air Force as a civilian adviser, and carried out work on gunnery and the trajectories of projectiles, which took him to Libya and Egypt. Then, from 1943 to 1945, he and Yael worked at the MIT Radiation Laboratory. After the war he became associate professor (1946/48) at Tufts, before going on to hold visiting positions at Princeton (1948/49) and Harvard (1949/50).

This was the period of McCarthyism, when the atmosphere in North American Universities was very difficult. Several of Dowker's friends in the mathematical community were severely harassed, and one had been arrested. In this situation Hugh and Yael decided to leave North America. They came to England in 1950, when Yael obtained a post at the University of Manchester and Hugh before long was appointed to a Readership in Applied Mathematics at Birkbeck College, London. Although Dowker is best known for his work in the purest and most abstract branches of mathematics, it is a mark of his versatility that he was capable of holding an applied post, in which he contributed to the theory of servomechanisms and projectiles. In 1962 he was appointed to a personal chair at Birkbeck, where he remained until his retirement in 1979.

Dowker's mathematical work lies mainly in the field of topology. Although the number of his published papers is not large they have been remarkably influential. They contain a wealth of ingenious examples, often answering difficult problems posed by other mathematicians. He was constantly concerned to find the 'right' basic definitions and axioms, and this led to his proving very general results under very few assumptions. While he is best known for his work in point-set topology he also made contributions to category theory, sheaf theory and the theory of knots. He had a long-standing interest in homology theory, for general spaces. Among many other important results he showed that the Čech and Vietoris homology groups coincide, for general spaces, as do the Čech and Alexander cohomology groups.

Dowker was widely travelled. In his early twenties he had twice crossed the United States and Canada, jumping on and off freight trains hobo-style. Later, as a mathematician, he held visiting positions in Russia, Israel, India and Canada. He also spent some time working on a kibbutz in Israel. He was able to speak Russian and knew some Georgian and some Hebrew. He loved the countryside and often went walking or mountain-climbing in the National Parks of Britain and Switzerland.

In manner Dowker was reserved and gentle, with an innate dignity and a penetrating wit. He possessed a high degree of integrity and moral strength which enabled him to endure seven years of illness uncomplainingly. He was unfailingly kind and generous, always ready to spend time helping others. With Yael he did a great deal of work for the National Association for Gifted Children. He had an affection for all young people and was known among his students for his helpfulness and patience. Three years after his retirement he died in London 14 October 1982 after a long struggle with ill-health.

After the second world war research in topology flourished vigorously in England, notably at Cambridge, Manchester and Oxford. Among topologists of the younger generation Adams [17] was outstanding.

(John) Frank ADAMS was born 5 November 1930 in Woolwich, near London, the elder son (there were no daughters) of William Frank Adams, civil engineer, and Jean Mary (née Baines), biologist, both of London. He was educated at Bedford School and then, after a year of military service, he went on to Trinity College, Cambridge. In due course he became a research student at Cambridge, first under A.S. Besicovitch and then, more significantly, under Shaun Wylie. His Ph.D. Thesis (1955) was on algebraic topology, which remained his main research interest for the rest of his life.

Adams spent the year 1954 at Oxford as a Junior Lecturer, where he came under the influence of Whitehead, then the leading topologist in the United Kingdom. Returning to Cambridge in 1956 as a research fellow at Trinity College, Adams developed the spectral sequence which bears his name, linking the cohomology of a topological space to its stable homotopy groups. The next step in his career was a two-year visit (1957/58) to the University of Chicago on a Commonwealth Fellowship. While he was there he used his new methods to prove the famous conjecture about the existence of H-structures on spheres. On his return from America Adams became fellow, lecturer and director of studies in mathematics at Trinity Hall, Cambridge. In 1961, on another visit to the United States, Adams enhanced his already high international reputation by solving another famous problem, concerning vector fields on spheres. For this purpose he invented certain operations in

K-theory, which later bore his name, and these have proved to be of fundamental importance.

In 1962 Adams left Cambridge for Manchester University, where in 1964 he became Fielden professor in succession to Newman, and was elected a fellow of the Royal Society at the early age of thirty-four. At Manchester he developed much further the powerful methods he had originated previously in a celebrated series of papers 'On the groups $J(X)$', which opened up a new era in homotopy theory. In the first of these he made a bold conjecture about the relation between the classification of vector bundles by stable isomorphism and their classification by stable homotopy equivalence of the associated sphere-bundles. Reformulated in various ways this Adams conjecture (later a theorem) is regarded as one of the key results of modern homotopy theory.

By 1970 Adams was the undisputed leader in his field. His reputation was such that he was seen as the obvious person to succeed Sir William Hodge as Lowndean Professor of Astronomy and Geometry at Cambridge. He was delighted to return to Trinity, his old college, although he never became very active in its affairs. Among Adams research interest in this later stage in his career three subjects predominated: H-spaces of finite type, classifying spaces of topological groups, and equivariant homotopy theory. Although he published important research papers on these and other subjects throughout this period he also began to publish more expository work, notably lecture notes on Stable Homotopy and Generalized Homology (1974) and a monograph on Infinite Loop Spaces (1978), based on the Hermann Weyl lectures he gave at Princeton. The latter, especially, gives a good idea of his magisterial expository style and particular brand of humour.

Adams was an awe-inspiring teacher who expected a great deal of his research students and whose criticism of work which did not impress him could be withering. For those who were stimulated rather than intimidated by this treatment, he was generous with his help. The competitive instinct in Adams was highly developed, for example, in his attitude to research. Priority of discovery meant a great deal to him and he was known to argue such questions not just as to the day but as to the time of day. Again, in a subject where 'show and tell' is customary he was extraordinarily secretive about work in progress.

Although Adams enjoyed excellent physical health he suffered a serious episode of depressive illness in 1965 and there were further episodes of depression later. To what extent his professional work was adversely affected by the nature of the treatments he received to control the condition is not clear but certainly his contributions to research in later years were not as innovative as those of his youth. Moreover, he never played the prominent role in the academic and scientific role to which his professional standing would have entitled him. Even so his influence was very great; those who turned to him for an opinion were seldom left in any doubt as to his views.

Adams' research achievements were recognized by the awards of the Junior Berwick (1963) and Senior Whitehead (1974) prizes of the London Mathematical Society and the Sylvester medal (1982) of the Royal Society. He became a Cambridge Sc.D. (1962), was elected a foreign associate of the National Academy of Sciences of Washington (1985) and an honorary member of the Royal Danish Academy of Sciences (1988), also he received an honorary Sc.D. (1986) from the University of Heidelberg. His collected works were published in 1992.

In 1953 Adams married Grace (née Carty), who soon afterwards became a minister in the Congregational Church. They had a son and three daughters (one adopted). Family life was extremely important to Adams, although he preferred to keep it separate from his professional life. The family used to do many things together, especially fell-walking in the Lake District. Adams acted as Treasurer of the local branch of the Labour Party and might be described as an intellectual Fabian in outlook. He died immediately following an accident on the Great North Road near Brampton, 7 January 1989.

Samuel EILENBERG was born in 30 September 1913 in Warsaw (Poland). At the University of Warsaw he was a student of Borsuk. His doctoral thesis, on the topology of the plane, led to a series of early publications on general topology but a research paper of 1938 on the action of the fundamental group on the higher homotopy groups of a space signalled a shift in the algebraic direction, where he was to make his reputation.

On his father's advice Eilenberg left Poland in 1939 and went first to Princeton. Before long a position for him was found at Ann Arbor, where at that time Wilder's research group of topologists included Steenrod. In this stimulating environment Eilenberg's interests rapidly broadened. At the end of the war he moved to Columbia University and made New York his main home. In later years he also spent a good deal of time in London where he was better able to pursue his non-mathematical interests. Of these the most important was the collection of small Asian sculptures, on which he became an expert; some of his trophies can be seen in the Metropolitan Museum in New York.

Eilenberg was one of the most influential algebraic topologists of the post-war period. His influence was spread not only through his many distinguished research contributions, often in collaboration with others, but also through his expository work, where he showed an extraordinary ability to clarify and systematize. Algebraic topology derived enormous benefit from this, but it went much further. In a long-term collaboration with Saunders Maclane he played a leading role in the development of homological algebra. One of the fruits of this joint work was the notion of category, which from being just a nice convenient way to look at certain kinds of mathematics has turned into a major speciality. Although algebra and topology remained Eilenberg's main mathematical interests he also contributed to the theory of automata.

Perhaps his most important book is the classic Homological Algebra (1956), which he wrote in collaboration with Henri Cartan. The earlier Foundations of Algebraic Topology (1952), of which Steenrod was coauthor, unfortunately did not progress beyond the first volume. He was an outstanding teacher, with a distinguished list of former research students, and the recipient of many academic honours. He died 30 January 1998 in New York, having been incapacitated by a stroke two years previously.

Bibliography

[1] W. Aspray, *The emergence of Princeton as a world center*, History and Philosophy of Modern Mathematics, W. Aspray and P. Kitcher, eds, Univ. of Minnesota Press, Minneapolis (1988), 346–366.

[2] F. Bachmann, H. Behnke and W. Franz, *In memoriam Kurt Reidemeister*, Math. Ann. **199** (1972), 1–11.

[3] R. Baltzer, *August Ferdinand Möbius*, Möbius Werke I, S. Hirzel, Leipzig (1885), v–xx.

[4] A. Borel, *The school of mathematics at the Institute for Advanced Study*, A Century of Mathematics in America, Vol. III, Amer. Math. Soc., Providence, RI (1988), 119–147.

[5] K. Borsuk et al., *Witold Hurewicz – life and work*, Collected Works of Witold Hurewicz, Kuperbeg, ed., Amer. Math. Soc., Providence, RI (1995).

[6] R. Bott, *Marston Morse and his mathematical works*, Marston Morse: Selected Papers, Springer, Berlin (1981), vii–xlvi.

[7] M. Cenkl, *Eduard Čech*, The Čech Centennial, M. Cenkl and H. Miller, eds, Contemporary Mathematics vol. 181, Amer. Math. Soc., Providence, RI (1995).

[8] S.-S. Chern, *Hassler Whitney*, Proc. Amer. Phil. Soc. **138** (1994), 465–467.
[9] K. Ciesielski and Z. Pogoda, *The beginning of Polish Topology*, Mathematical Intelligencer **18** (3) (1996), 32–39.
[10] L.W. Cohen, *James Waddell Alexander 1888–1971*, Bull. Amer. Math. Soc. **79** (1973), 900–901.
[11] G. Darboux, *Eloge historique d'Henri Poincaré*, Oeuvres d'Henri Poincaré, Vol. II, Gauthier-Villars, Paris (1916), vii–lxi.
[12] A. Granas and J. Jaworowski, *Reminiscences of Karol Borsuk*, Topological Methods in Nonlinear Analysis **1** (1993), 3–8.
[13] P.J. Hilton, *Memorial tribute to J.H.C. Whitehead*, Enseignement Mathematique **7** (1961), 107–125.
[14] P.J. Hilton, *Obituary: M.H.A. Newman*, Bull. London Math. Soc. (1986), 67–72.
[15] Sir W. Hodge, *Solomon Lefschetz 1884–1972*, Biogr. Mem. Roy. Soc. **19** (1973), 433–453.
[16] H. Hornich, *Wilhelm Wirtinger*, Monatshefte fur Mathematik **52** (1) (1948), 1–12.
[17] I.M. James, *John Frank Adams*, The Dictionary of National Biography 1986–1990, Oxford Univ. Press (1996).
[18] B. Laugwitz, *Bernhard Riemann 1826–1866*, Birkhäuser, Basel (1995).
[19] H. Lebesgue, *Notice sur la vie et les travaux de Camille Jordan*, Oeuvres de Camille Jordan, Gauthier-Villars, Paris (1964), x–xxiv.
[20] S. Lefschetz, *Witold Hurewicz, In Memoriam*, Bull. Amer. Math. Soc. **63** (1957), 77–82.
[21] S. Lefschetz, *A page of mathematical autobiography*, Bull. Amer. Math. Soc. **74** (1968), 854–879.
[22] S. Lefschetz, *Reminiscences of a mathematical migrant in the US*, Amer. Math. Monthly **77** (1970), 344–350.
[23] S. Lefschetz, *James Waddell Alexander (1888–1971)*, Biogr. Mem. Amer. Phil. Soc. (1973), Philadelphia 1974, 110–114.
[24] L. Markus, *Solomon Lefschetz: an appreciation in memoriam*, Bull. Amer. Math. Soc. **79** (1973), 663–680.
[25] M.H.A. Newman, *John Henry Constantine Whitehead*, Biogr. Mem. Roy. Soc. **7** (1961), 349–363.
[26] M.H.A. Newman and B. Whitehead, *A biographical note*, The Mathematical Works of J.H.C. Whitehead, Pergamon Press (1962), xv–xix.
[27] O. Perron, *Heinrich Tietze 31.8.1880–17.2.1964*, Jahresber. Deutsche Math. Vereinig. **83** (1981), 182–185.
[28] E. Pitcher, *Obituary: Marston Morse*, Biogr. Mem. Nat. Acad. Sci. (USA) **65** (1994), 223–240.
[29] G.-C. Rota, *Fine Hall in its Golden Age: Remembrances of Princeton in the early fifties*, A Century of Mathematics in America, Vol. II, Amer. Math. Soc., Providence, RI (1988), 223–236.
[30] D. Strauss, *Obituary: C.H. Dowker*, Bull. London Math. Soc. **16** (1984), 535–541.
[31] G.W. Whitehead, *Norman Earl Steenrod*, Biogr. Mem. Nat. Acad. Sci. (USA) **55** (1975), 452–470.
[32] H. Whitney, *Moscow 1935: Topology moving toward America*, A Century of Mathematics in America, Vol. I, Duren, ed., Amer. Math. Soc., Providence, RI (1988), 97–117.
[33] R.L. Wilder, *Reminiscences of mathematics at Michigan*, A Century of Mathematics in America, Vol. III, Amer. Math. Soc., Providence, RI (1988), 191–204.
[34] S. Wylie, *John Henry Constantine Whitehead*, J. London Math. Soc. **37** (1962), 257–273.

CHAPTER 33

Johann Benedikt Listing

E. Breitenberger

Department of Physics and Astronomy, Ohio University, Athens, OH 45701-2979, USA

1. Overview

Johann Benedikt (or Benedict) Listing, 1808–1882, was one of the founding fathers of our discipline even though he did not contribute any capital discoveries. He gave it the apt name "topology" and wrote the very first monograph on it, published originally as a paper in 1847, and re-issued as a slim booklet in 1848, under the modest title *Vorstudien zur Topologie*. Later he turned to aggregates of points, lines and polygons which he called "complexes", and wrote a novel survey under the title *Der Census räumlicher Complexe* ..., again published first as a paper and then as a book, both in 1862, and followed a few years later by a paper containing two addenda. He wrote nothing else on topology, but repeatedly lectured on the subject from 1848 onwards in various seminars and meetings.

Listing was a doctoral student of Gauß who became a close personal friend, but he should not be labelled a mathematician; from 1839 on he was nominally a professor of physics in Göttingen. Indeed he appears as one of the many minor universalists who lend so much colour to the history of 19-th century science. He was a founding father of modern ophthalmology, too, since he pioneered the study of the optical properties of the eye, described in a non-technical memoir called *Beitrag zur physiologischen Optik* in 1845; later he expanded this in a theoretical article which still reads strikingly accurate today. He made other contributions to optics; he studied the figure of the earth in minute detail; he made observations in meteorology, terrestrial magnetism, and spectroscopy; he wrote on the quantitative determination of sugar in the urine of diabetics; he promoted the nascent optical industry of Germany and better street lighting in Göttingen; he travelled to the world exhibitions in London 1851, Vienna 1873 and London 1876 as an observer for his government; he assisted in geodetic surveys; and more. Besides the word "topology", he invented a good many other terms, some of which have become current: "entoptic phenomena", "nodal points", 'homocentric light", "telescopic system", "geoid". Lastly, he coined the name "one micron" for a millionth of one metre; it entered general usage through the American Society of Microscopists who adopted the micron as the standard unit of length for microscopical work in 1879.

HISTORY OF TOPOLOGY
Edited by I.M. James

Johann Benedikt Listing (1808–1882)

The man himself was industrious and inquisitive, kind and helpful, gregarious and witty, good-natured to a degree, a true friend to many and an enemy to none. Yet he was not held in high esteem. Given his achievements, here lies a conundrum. Flaubert's dictum "*L'homme c'est rien – l'œuvre c'est tout*" (Never the person – solely the work) will not apply. For an elucidation the chronicler must yield to the biographer who champions the individual. Three reasons for Listing's small academic repute stand out.

First, the structure of the personality. Listing was a mild manic-depressive who most of his adult life oscillated between opposite frames of mind. Early on the extremes were approximately "carefree/dejected", later more "mellow/listless"; the waves lasted from one month to many. This is the judgement of a much later generation, of course. His contemporaries, including Gauß, a most astute observer, certainly did not see Listing in such a light, and he never saw through himself. Like many persons afflicted with this disposition he found it difficult to concentrate, both in the high and in the low phases. Thus his natural curiosity frequently drove him to linger over details; he then failed to see the wood because of all the trees, and did not reap the full fruit of his efforts. His writings often lack a grand sweep, and occasionally appear pedantic. Presumably the listeners in his lectures formed similar impressions.

The second reason was his wife. Pauline, 15 years younger, beguiling and profligate, must be characterised as an overweening hysteric. Her treatment of servants brought her before the magistrates any number of times, while her relations with landlords led to many moves for the family. In brief, she was no social asset.

Thirdly, the partners in this ménage had one fault in common: an astonishing inability to cope with money. Listing borrowed frequently and heavily, sometimes from usurers; Pauline habitually abused credit, and again ended up in court with some regularity. They tended to live beyond their means, and on one occasion barely avoided bankruptcy. No wonder that Listings were not invariably held in polite regard.

When Listing died, only one obituary appeared, by his correspondent P.G. Tait in Edinburgh. Thereafter this obliging and deserving man was nearly forgotten, although in 1905 Wilhelm Ostwald reprinted the *Beitrag* in his renowned series of classics of science. Even the word "topology" did not readily gain currency.

2. Youth in Frankfurt

Listing was born in Frankfurt-on-Main on 25 July, 1808. The father, who had the same names, came from a family of craftsmen; they were mainly cabinetmakers, but he made brushes. The mother, Caroline Friederike née Theißinger, was a peasant girl from the Hunsrück hills west of the city. He seems to have been an only child; there were several cousins, at least one of whom later emigrated to Central America. Hardly an affluent setting.

The boy showed promising gifts, including a special talent for drawing and fine writing. Already at age 13 he earned some pocket money giving calligraphy and sketching lessons. He soon received aid from the Bernus family of Frankfurt patricians and from the *Städelsches Institut*, a foundation which still exists and maintains one of Europe's major art museums.

In 1816 he entered the *Musterschule*, a very progressive institution with distinguished founders and a select staff. The mathematics and astronomy master and textbook author, Johann Heinrich Müller, 1787–1844 (not to be confused with the Freiburg physicist Johann

Heinrich Jakob Müller, 1809–1875, who was widely known as author and translator), soon gained influence over him. Interest in mathematical and scientific matters grew; eventually he envisaged academic study. Müller remained a mentor and friend until his untimely death.

From 1825 onwards Listing attended a *Gymnasium* and graduated in 1830. By then he was familiar with English, French, Italian and Latin; he knew some Greek; he had a working knowledge of analytic geometry and calculus; and he had a boundless curiosity about all the natural sciences. The Städel foundation honored his scholastic success with the award of a four-year scholarship for the study of "architecture and mathematics". Their charter emphasized the fine arts to such an extent that it would not have permitted support for the study of mathematics or science alone, hence the curious combination. The stipend was generous: 400 *Thaler* per year, an amount well above the needs of a frugal student.

3. Student in Göttingen

In the fall of 1830 Listing moved to Göttingen and registered as a student of mathematical and natural sciences. For a starter he even took a course in architecture, but the rest of his time he spent all over: mathematics, astronomy, anatomy, physiology, botany, mineralogy, geology, chemistry. Why worry? He was better off than the average Göttingen student, and the place, then as large as Oxford and Cambridge combined, buzzed with enticing activity.

From his third semester onwards he also went to lectures by Gauß (nominally professor of mathematics and director of the observatory). The *princeps mathematicorum* disliked formal teaching and had a public reputation for aloofness, but in his small inner circle he appeared rather different, for he possessed the precious gift of making friends with young people. Spotting Listing's industry, he thought him "quite promising" (*vielversprechend*). Soon afterwards he invited the lowly student to dinner with stars like Wilhelm Weber, 1804–1891, the brilliant young physicist he had brought to Göttingen in 1831, and his son-in-law Georg Ewald, 1803–1875, one of the leading semitists of the century. The relation intensified, enhanced no doubt by Listing's unselfish disposition, and remains close and even intimate to the last. Indeed, Listing was present when Gauß finally closed his eyes shortly after 1 a.m. on 23 February, 1855.

In this inner circle Listing also met Wolfgang Sartorius von Waltershausen, 1809–1876, a fellow student whose interests lay mostly in mineralogy and geology; as his family was well acquainted with Gauß he had grown into the status of another young friend and was welcome to drop in and attend a class any day. Being of the same age, the two soon formed a close friendship which was destined to endure a lifetime.

In those years after 1830 Gauß no more undertook the arduous field work for the geodetic surveys that had consumed so much of his time during the 1820's. Shifting interests had gradually led him to physics; first to Faraday's discovery of magnetic induction (1831). It may well have been a preoccupation with magnetically linked, closed circuits that brought Gauß back to some old ideas about the curvature of space curves and resulted in an integral formula for winding numbers jotted down in January of 1833. He never published it, but it was later incorporated in his *Gesammelte Werke*. For the present context, it is significant that he added the lament:

> Von der Geometria Situs, die Leibnitz ahnte und in die nur einem Paar Geometern (Euler und Vandermonde) einen schwachen Blick zu thun vergönnt war, wissen und haben wir nach anderthalbhundert Jahren noch nicht viel mehr wie nichts

(Of the *Geometria Situs* presaged by Leibniz, and of which only two geometers (Euler and Vandermonde) had the privilege to gain a faint glimpse, we still know and have next to nothing although one and a half centuries have gone by), which aroused some attention after Maxwell alluded to it in his classic *Treatise on Electricity and Magnetism* of 1873. In fact, Gauß had topological matters on his mind, off and on though never intensively. He even talked about them. Listing reports "sketchy remarks" (*hingeworfene Äußerungen*) and records various conversations on "analysis situs" in his diaries.

Another topological fact may have lingered in Gauß' thoughts at about that time: the impossibility of defining "right" and "left" without reference to a material object. He seems to have encountered it rather early, when he had to come to grips with aprioristic philosophy. Kant regarded this impossibility as proof that space was a category of pure reason, whereas Gauß recognized it as merely a matter of symmetries. He first spoke out on the subject in a non-technical manner in 1831, in the preliminary announcement of his second memoir on quadratic residues (which is also famous for his geometrical representation of complex numbers). This was just after Listing had arrived in Göttingen.

Terrestrial magnetism was another concern from physics. It is not always appreciated what a tireless empiricist Gauß was. He had constructed sensitive magnetometers, and introduced the new, metric and "absolute", system of "Gaussian" units of measurement. He had also built a magnetic observatory without iron components or nails; it was the world's first geophysical research institute and can still be visited in Göttingen. He was interested in both the local and the temporal variations of the earth's field. Thus a synoptic approach was indicated. Together with Weber, he founded a Magnetic Union, the very first instance of international scientific cooperation. The accumulating mass of data later permitted him to calculate the positions of the magnetic north and south poles by means of least squares, his pet method. Needless to say, the extensive observations absorbed much time and effort, but willing helpers always turned up. In Göttingen, Listing as well as Sartorius assisted regularly.

In a short while, Listing had become a young friend, a recipient of sketchy thoughts on topology, and an experimentalist collaborator. It may have followed of itself that he also did his dissertation under Gauß. He made good progress and obtained the Dr. phil. on 30 June, 1834. The dissertation, *De superficiebus secundi ordinis*, establishes a connection between surfaces of the second degree and the ternary forms studied by Gauß in the *Disquisitiones arithmeticae*. Gauß rated it "ingenious" (*scharfsinnig*). Listing acknowledged an old indebtedness by dedicating the dissertation to his teacher Müller: *"Viro doctissimo summe venerabili"*.

4. Italy and Hanover

What next? An ambitious young graduate would have reached for the best academic position available. Listing was not so career-minded. Instead, he embarked on a three-year adventure.

Sartorius the geologist was interested in volcanoes and intended to study the barely understood Mt. Etna in Sicily. He was also willing to combine an expedition to the Mediterranean with further geomagnetic work for Gauß. Both undertakings called for a qualified assistant. Who was better suited than his good friend Listing? Funding was no problem, as he was wealthy enough to finance everything himself.

And so, Listing having graduated, the two went to Frankfurt, procured the needed gear, and on 27 July, 1834 set out southwards on their own coach, Listing in the driver's seat, helpful as ever. At intervals they stopped for a week or two to make magnetic and barometric measurements; they also arranged to meet various friends and colleagues. Karlsruhe, Stuttgart, Bavaria, Salzburg, Innsbruck, Verona, Milan were among the early stops. Then they continued on a circle of latitude to Venice, went on by the shortest route to the island of Elba, and again by the shortest way via Rome to Naples, all in accordance with the plans of the Magnetic Union.

After more than a year, in the fall of 1835, the pair finally reached Sicily. While the magnetic observations continued, the huge Etna massif now became the focus of labour. Geology and topography had to be elucidated; extensive geodetic surveys were undertaken. Sartorius later made full use of these early results in a *magnum opus* on Etna (1876).

Beyond his duties in the field Listing still found enough time for various studies. In particular, he roamed over bits and pieces of topology. His notebooks and diaries from this period contain many elegant sketches of crystals and polyhedra knight's moves on the chessboard, ornamental calligraphic flourishes with discussions of their symmetry properties, instances of the Königsberg problem, and the like. Shortly before April 1836, the word "topology" makes its first appearance. It may have helped him to unify his scattered thoughts. At any rate, he felt a need for a summary essay, and under date of April 1st, 1836, wrote up his preliminary ideas in a lengthy letter to his teacher Müller. Its main contents were later incorporated in the *Vorstudien* and need no comment. The introductory passage is relevant, however; in shortened paraphrase, he explained to Müller that he is dissatisfied with the term "*geometria situs*" introduced by Leibniz because the word "geometry" should not denote a discipline in which distance and quantity are irrelevant, and anyway, because the word should remain reserved for the then current term "*géometrie de position*" (due to Carnot, 1803). The entire doctrine being rather new, he felt justified to give it a new name and therefore called it "Topology", which he thought more appropriate.

Some time afterwards, Sartorius fell severely ill, was given up by the physicians and barely survived, desperately nursed by Listing. Then Listing fell ill for a month. Still, if life was not rosy, a future beckoned. The *Höhere Gewerbeschule Hannover*, instigated by Gauß, wrote to offer a position as teacher of applied mathematics, machine design and engineering drawing. This institution was an ambitious polytechnic which later became a *Technische Hochschule*, and today is a fully-fledged university. Listing answers politely, yet does not commit himself.

A return to Göttingen had been intended for the university's centennial in the fall of 1837. Now the cholera intervened. First the Italian mainland became inaccessible, then the epidemic reached over to Sicily. At an opportune moment, Sartorius and Listing catch a Danish brig going to Rio de Janeiro with a cargo of Sicilian wine. Thus they reach Gibraltar, then continue on another vessel to Lisbon, and from there on one of the newfangled steamers to Liverpool. They took the unplanned opportunity and stayed in Britain for a couple of weeks: Sartorius to look at the extremely involved geology, Listing, perhaps soon a teacher in a polytechnic, to look at the industry which then led the entire world. And so to Hamburg.

Later in the fall Listing presents himself in the city of Hannover at the *Gewerbeschule* to interview and is given the position at once, to start in November 1837. The new sphere of activity perhaps did not suit his inclinations exactly, but he seems to have liked it. In any event, he could have carved himself a comfortable niche after a few apprentice years.

It was not to be. On 2 March, 1839, he is summoned before the cabinet secretary in charge of the affairs of the University of Göttingen who offers him point blank a position as professor extra-ordinary of physics. Stunned but quickly overjoyed, he accepts, to start at Michaelmas 1839.

5. Back to Göttingen

The abrupt change of Listing's fortunes arose from historical events which call for a digression. In 1837 occurred the death of William IV, King of Great Britain, Ireland and Hanover. In London he was succeeded by his niece, eighteen-year-old Victoria. In Hanover the law of succession differed, though: a female could ascend the throne only if the entire male line was extinguished. Thus Ernest Augustus, Duke of Cumberland, a younger brother of William, became King of Hanover.

Ernst August, as he henceforth called himself, was a Tory, reactionary, headstrong, and a sworn foe of all ideas emerging from the French revolution. His Whig opponents in the House of Lords called him "the most unpopular prince of modern times". Hanover had a fairly liberal, written constitution, promulgated in 1833, loosely connected with the British Parliamentary Reform of 1832, and initiated partly by student unrest in Göttingen in 1831. Ernst August rescinded it with a stroke of the pen. His embittered subjects were placated a little by lowered taxes, but not all Göttingen academics gave in. Seven of them, including Weber and Ewald, signed an eloquent letter of protest. They made the mistake of addressing it not to the King but to a cabinet minister. Besides, the contents leaked to the public. Ernst August reacted with the instant dismissal of all seven, and, moreover, commanded three of them, including the famous brothers Grimm, to leave Hanover within twenty-four hours on pain of incarceration. This happened shortly after Listing had begun at the *Gewerbeschule*.

Another and unintended victim was Gauß: he lost not only his son-in-law, but also his friend and indispensable collaborator Weber. Eventually the ministry asked him to name a successor to Weber. In the faint hope of finding someone who might "to some extent" (*einigermaassen*) replace Weber for him, he forwarded three names with Listing's in third place. Negotiations with the first two foundered, whence Listing, just 31 years old and still without publications, emerged as a junior professor in renowned Göttingen as related above.

However, a close collaboration with Gauß did not materialize, for two unforeseen reasons. First, from about 1835 onwards the naval powers more and more appreciated the importance of terrestrial magnetism to navigation. In the interest of both strategy and commerce, they increasingly muscled into research; the pre-eminence of Gauß and Weber faded, and the Magnetic Union slowly crumbled. In 1838 the Royal Society awarded Gauß the Copley Medal for his geomagnetic work, but he gradually lost interest and abandoned the subject.

Secondly, the brutal dismissal of the "Göttingen Seven" caused an outcry throughout Europe followed by the spontaneous formation of a committee-in-aid which collected massive subscriptions and continued to pay the salaries of the sacked Seven while they were without positions. In particular, Weber was enabled to stay on in Göttingen for over five years to cope with the affairs of the Magnetic Union until in March 1843 he moved to Leipzig.

Thus Listing unexpectedly remained free to do what he pleased, although his personal relation with Gauß remained close as always, with mutual visits and dinners and many discussions. Then his friend Theodor Ruete, 1810–1867, a former fellow student and now an aspiring young ophtalmologist in Göttingen, aroused his interest in the optics of the human eye. Coincidentally, Gauß furnished indispensable theoretical ideas inasmuch as just at that time, prompted by weaknesses in a paper of Bessel, he wrote up his theory of thick lenses which he had worked out much earlier but never published (maybe Listing's reports of the collaboration with Ruete also nudged him a little). In any case, Listing found himself in a unique position with access to specialist guidance as well as to a new chapter of optical theory. He made commendable use of it. After protracted, difficult and minute observations he had enough material for the *Beitrag zur physiologischen Optik*; the little work without mathematical formulae appeared at the end of 1845 and became a classic not least through the medical illustrations which he had drawn and lithographed by his own hand.

What next, may again be asked. On the whole, he had spent six years learning how to teach general physics, how to cope with faculty politics, and how to understand the workings of the human eye, and he had enjoyed a bachelor's merry social life, but he had hardly done much to keep up with the rapid advances of physics and applied mathematics. Now he went his own way and returned to topology where pioneering work beckoned.

6. The *Vorstudien*

Topology had never been far from his mind. Gauß also continued to supply stimuli; for instance, Listing's diaries report another discussion of "Geometria situs" on 2 January 1845, when the human eye was still his prime concern. Soon after the *Beitrag* is out, early in 1846, Listing begins to write and has a long essay done one-and-a-half years later. He calls it *Vorstudien zur Topologie*. The term means "preliminary studies" and should be taken literally, for he was conscious all along that he had no comprehensive vistas to offer. After a cursory but correct historical survey, he defines his own standpoint:

> Unter der Topologie soll also die Lehre von den modalen Verhältnissen räumlicher Gebilde verstanden werden, oder von den Gesetzen des Zusammenhangs, der gegenseitigen Lage und der Aufeinanderfolge von Punkten, Linien, Flächen, Körpern und ihren Theilen oder ihren Aggregaten im Raume, abgesehen von den Maß- und Größenverhältnissen.

(By topology we mean the doctrine of the modal features of spatial objects, or of the laws of connection, of relative position and of succession of points, lines, surfaces, bodies and their parts or their aggregates in space, always without regard to matters of measure or quantity.) The stress on "connection" recurs in other passages and hints at a deeper understanding of continuity which seems to be Listing's own, for it cannot be felt in the posthumous papers of Gauß. Still, a lot of water had to flow under the bridge until the spirit of Klein's Erlangen program (1872) prevailed, and Listing's snake of words was replaced by the brief call for invariance under continuous 1–1 transformations.

In an essentially combinatorial section entitled "On Position" (*Von der Position*) he discusses the relative orientations of two Cartesian axis systems with parallel axes. There are

48, allowing for all right-left inversions and permutations; they may serve to symbolize the relative positions of two objects such as dice. His discussion is long-winded because he does not possess the basic notions of group theory. He even writes a sum for the product of two transformations. Nor does he perceive subgroups or inner automorphisms. On the other hand, he remains not content with abstractions and applies his insight to the object-image relations in various optical instruments; they had come to his close attention during many observational pursuits. What he has to say sounds lengthy but far more lucid than most optics textbooks; they often leave their readers wondering and fail to stress the basic symmetries inherent in reflections and refractions which could be readily illustrated by pairs of symbols such as b and d, L and Γ, or 6 and 9.

The rest of the essay is a medley under the title "On Helices" (*Von der Helikoïde oder Wendellinie*). He defines helices in quite a general manner and shows first that topologically they differ only in their handedness. Like many early authors he uses right- and left-handed in a sense opposite to the one that is now standard; the former usage survives to the present only in the designation of the handedness of circularly polarized light. Typically he adds many examples from botany and zoology, right back to Linnaeus. Then he passes on to double and multiple helices as in pine cones, threads, strings and ropes. He also describes simple and multiple screws and screw surfaces.

When a simple or multiple helix is closed, the complication arises that it may have been knotted before the endpoints were joined. In order to study knots he flattens the extended object into a plane. Consideration of the simplest possible cases shows how two knots may be equivalent in the sense that they can be (continuously!) deformed into one another; if not equivalent, they are topologically different. The simplest possible forms of a given knot he calls "reduced" but he does not ask how to find them in some systematic manner; it took another eighty years until Reidemeister showed how to do that. He does, however, attempt to classify knots by the nature of the meshes in a reduced form (including the part of the plane outside the knot). He does not carry the classification very far; it was extended thirty years later by Tait. Still, some of his examples are astonishingly involved.

When a closed helix is not just flattened into the plane but projected, the result is a closed curve with double points. Such curves may again be classified by the nature of the meshes. The white-on-black figure on the cover of the *Vorstudien* reproduces his Figure 20 as an illustration of the method. Closed curves with multiple points lead Listing on to the Königsberg problem; he gives the general solution and shows that it also holds in the 3-dimensional case.

Much of all this would nowadays not be considered part of topology. At that pioneering stage, however, the subject had no conventional boundaries yet. One may even give Listing credit for the polymath breadth of his outlook. For instance, his emphasis on the importance of symmetries to science was altogether prophetic. In this context he was quite right to quote the classic *Treatise on Crystallography* by W.H. Miller (1839) in a long footnote. In another place, he was also right to mention a paper by Charles Babbage, *On a method of expressing by signs the action of machinery*, which in 1826 had attempted to classify machines by the nature of the connections between elements; for Listing, this was indeed topology.

Much of it was also incomplete. No matter, it was original. The section on knots was even the very first anywhere. No one had written a single word on knots before (Vandermonde in 1771 had written about plaits, knits and knight's tours, but said nothing on knots).

7. Middle years

Coincident with (and perhaps detrimental to) the writing of the *Vorstudien* Listing travelled the path to matrimony. Pauline, née Elvers, was the daughter of an eminent jurist in Kassel, the nearby capital of Electoral Hesse. In January 1846 she accompanied her mother on a private visit to Göttingen, met Listing and captivated him instantly. In April they became engaged, in September they married. After a honeymoon in Bavaria she commenced as a housewife, exceeded her monthly allowance in three weeks and needed more money. Domestic misery has begun.

Listing had just acquired a spouse, moved into a new apartment and published his second book, the *Vorstudien*, when his career was once more affected by extraneous forces. The revolution of 1848, launched with the abdication of King Louis Philippe in Paris in February, quickly laps over into Germany. The autocratic bastion of King Ernst August begins to quake in March; by September it has collapsed. Ernst August is forced to grant a constitution which goes farther in its liberality than any other in Germany. Moreover, in Göttingen everyone, town as well as gown, remembers the dismissed Seven and begins to agitate for their return. As early as April, Weber is approached officially; in May he comes over from Leipzig to negotiate, and at the end of August he is appointed to his old post. Ewald, too, eventually returns.

Listing had to be newly accommodated in this altered setting. He hardly lifted a finger to safeguard his own interests; it was Weber, always a gentleman, who insisted from the outset on cooperation and avoidance of conflicts. In the end, Listing is promoted to professor ordinary and charged with pursuing "mathematical" physics, whereas Weber takes the "experimental" part. This representation of physics by two separate chairs was the first in the world. It owed its emergence to the historical incidents of 1837 and 1848, but it was also promoted by an express desire on Weber's part to see his discipline enhanced in its standing as a basic science. The division into an experimental and a theoretical part should not be taken too seriously, for the differentiation which is now common only hardened much later. Weber and Listing continued to do just what interested them most.

If Listing had gained something, he took losses elsewhere. He had to transfer two thirds of his laboratory space to Weber, and his income did not rise substantially. While the cost of his household grew, waves of price increases also swept over the country repeatedly. Thus his and Pauline's housekeeping inabilities developed into an endless spiral of financial woe.

A daughter arrived in the summer of 1848, another in 1849. Both had musical gifts and received a thorough and probably costly training in voice and instrumental music. The younger later became a music teacher and organist. The older in 1881 married Wilfrid Airy, a son of the Astronomer Royal Sir George Airy; she died a year later after the birth of a daughter (Listing's only grandchild).

After the death of Gauß in 1855 Listing obtained some financial relief when he was awarded the rent-free apartment in the observatory where Gauß had lived. A second such apartment existed there. In May of 1858 it was given to Riemann, already a professor extraordinary but still at very low pay. Bernhard Riemann, 1826–1866, came from a tubercular family and had lost parents, brothers and sisters until only the two youngest sisters were left to him and his care. He had explored the general theory of complex functions as a doctoral student of the "experimental physicist" Weber whose *Assistent* he afterwards remained for some years. Listing knew him well, through academic governance and from common walks; he seems to have recognized his genius early, and respected him highly.

Now they found themselves neighbours with a common terrace in front of their doors. Yet no friendly contacts between the households developed. Perhaps the tuberculosis was responsible. In the northern part of Europe it was believed inherited because it ran often enough in families, but in the south it was held to be contagious. Listing may well have learned that during his mediterranean travels, and avoided close contact with Riemann and his sisters in order to protect his two young daughters.

When the *Vorstudien* were done Listing returned to various matters from optics. Then he travelled to the historical Great Exhibition of 1851 and used the occasion to make more acquaintances in Britain through visits in and around London. He entered optics again, to write the substantial theoretical article about physiological optics which also includes a flawless exposition of the theory of thick lenses with many impressive figures, published 1853 in a large handbook of physiology. Then he turned to a number of subjects, interrupted by the task of integrating Göttingen into the network of meteorological stations promoted by Heinrich Dove, the founder of synoptic meteorology (Copley Medal 1853). And at long last, back to topology.

Listing's diaries note several conversations about topology with Dedekind who was first a student and then a *Privatdozent* in Göttingen until 1858, and with Dirichlet who had succeeded Gauß in 1855. According to his notes, he also knew Riemann's short but important paper of 1857 which clearly and convincingly defines the connectivity of a surface by means of separating and nonseparating cuts. This was Riemann's only publication in topology but Listing also knew of some unpublished material; thus they must have had talks on topological matters although the diaries do not mention any. However that may be, Listing later used the idea of connectivity merely in an altered manner that did not entirely do justice to the clarity of Riemann's insight. It remains a puzzle why these two kindhearted men never discovered how much they had to tell each other. The failure was mutual inasmuch as Riemann did not adopt the convenient word "topology" instead of his "analysis situs". Of course, their personalities were quite different, Listing expansive, Riemann very shy. Their thinking habits also differed greatly, Riemann always aiming straight at the heart of the matter, Listing often dawdling with detail. A creative artist and an assiduous stamp collector facetiously come to mind. How should a fluid exchange of ideas have grown up?

Listing's new topological concerns originated from a long-standing preoccupation with crystals, acquired perhaps under the influence of Sartorius the mineralogist. It guided him to polyhedra and from there to related items such as tents, scaffolds, adjacent cells as in a foam, maps in the plane, and so on. He was the first to recognize that all these objects had topological properties; in particular, he perceived relations between the numbers of their elements, similar to Euler's formula for simple polyhedra. The novel insights led him to write another milestone of topology.

8. The *Census*

This product of much labour was published in 1862 under the baroque title *Der Census räumlicher Complexe oder Verallgemeinerung des Euler'schen Satzes von den Polyedern* (Census of spatial aggregates, or generalisation of Euler's theorem on polyhedra). It consists of 84 pages of text, followed by an alphabetic list of 31 terms none of which has survived into the present, and by copperplates containing 64 figures. Through long discussions larded with new terminology it proceeds towards an apex consisting of a single proposition, called the *Census-Theorem*. From remaining notes and drafts it is known that Listing had considered half a dozen titles mentioning topology, but in the end he opted for the word "census", meaning a taxonomy according to his peculiar criteria. A full, technical description should not be the aim of a biography; hence, with some license only a few principal cues will be outlined.

Start with Euler's formula $C + F = E + 2$ for the number of corners, faces and edges of a simple polyhedron. Listing asks himself, why 2? He realizes that for an object like a picture frame, the 2 becomes a zero. Today we say neatly that, when the polyhedron is not simple but equivalent to a sphere with g handles, the 2 becomes $2 - 2g$. But the 2 can also become an odd number. E.g., for a bounded map in a plane, a 2-dimensional analog of polyhedra, it becomes 1. Listing guesses even more. Just as in the *Vorstudien*, he pays attention to the space outside a geometrical object (which he calls the *Amplexum*) and interprets the 2 as the number of pieces into which a closed, simple polyhedron separates the 3-dimensional space, whereas he sees the 1 play the analogous role for a plane map. Furthermore, when he writes Euler's formula as $C - E + F - 2 = 0$, he notes that the odd-dimensional components, the edges and the pieces of space, occur with a $-$ sign, and the even-dimensional ones, the corners and faces, with $+$ sign. Thus dimensionality as well as its parity should be watched.

In pursuit of the highest generality, he thereafter considers aggregates (or "complexes") of constituent points, lines and polygons which must be connected but remain otherwise arbitrary, and he reckons the resultant space pieces as parts of the aggregate. To illustrate: a tetrahedron may be a scaffold made up of six rods, or a closed surface made up of four triangles with an extra space piece now belonging to it; or the bottom triangle may be omitted to leave a tent, one of the triangles may be detached like a tent flap, a tent pole may be sticking out at the top, or a flag may be put at the top; and so forth. The edges and faces may also be bent. Connectivity is now introduced by means of a process he calls a *Cyclose*. On any given constituent, whether polygon or polygonal face or space piece, he seeks to construct two closed, linked simple curves, one lying entirely within the constituent, the other entirely outside. He then spans the latter by a diaphragm which necessarily cuts the constituent. The process is repeated until the constituent would separate. The maximal number of cycloses possible, which he calls the *cyclomatisch* number, is evidently Riemann's connectivity minus 1, only it is introduced in a more rigid fashion. A simply connected component or a single point or the whole space he calls *acyclomatisch*. Through a delicate discussion he needs to clarify how the cycloses are to be established for more complicated aggregates having loops that are knotted, or wound around each other; it anticipates Felix Klein's later distinction between intrinsic and relative (or embedding) properties. For constituent surfaces it also becomes necessary to state whether or not they are closed; he calls closedness the *Periphraxis*.

Through a sequence of propositions he ultimately reaches the *Census-Theorem* for an aggregate of a corners, b edges and c faces with d resultant pieces of space. He writes it as

$$(a + \alpha) - (b + \beta) + (c + \gamma) - (d + \delta) = 0,$$

where each number α, β, γ, δ is called an *Attribut* and is made up by properly counting cycloses, periphraxes and the number of constituents extending out to infinity, if any. Furthermore, he shows that for compound aggregates made up of p nonconnected aggregates the zero on the right becomes $p - 1$. Many examples follow.

Despite its startling formal elegance the *Census-Theorem* has not entered the canon of present-day topology. Not even the existence of such a general relation is well remembered. The reason is mainly that the "attributes" remain obstreperous. They are often not easy to ascertain; they commingle dimensionality, connectivity and extension; and they do not foreshadow the combinatorial invariants which were later found to furnish trenchant criteria of wide applicability. In the end, only partial cases of the *Census-Theorem* were absorbed into various branches of topology, such as graph theory, where other concerns demanded different emphases and left Listing's contribution hard to recognize.

Yet the *Census* remains a milestone. It established topological aspects where no one had seen them before. One is also impressed by the strictly topological nature of Listing's proofs. Not once does he mention angle or distance, area or volume. And where it sounds verbose, the utter generality of his approach still rests upon an exactness of definition which for his time was remarkable. In brief, it is a pioneering work which like many other bold explorations was quickly superseded by more practical undertakings.

Throughout his meticulous survey of variegated examples, Listing fails to notice the existence of one-sided surfaces. However, the Möbius strip appears in the *Census* as Figure 3. He refers to it only once, in a footnote, together with another multiply-connected surface which occurs in the adjoining Figure 4 but is two-sided. Listing says that these two examples, both bounded by a single, unknotted curve, have "properties quite different" (*ganz andere Eigenschaften*) from those just described in the main text for a simple diaphragm, namely, how it can be continuously contracted to a point, and how in order to pass from one side to the other the boundary curve must be crossed. The statement is faintly unclear, but it definitely does not indicate that he perceived one-sidedness.

The strip is named after Ferdinand August Möbius, 1790–1868, who was a student with Gauß in 1813/1814, became nominally a professor of astronomy in Leipzig, and wrote his two seminal papers on topology when he was older than 65. His biographer dates the earliest consideration of the strip "with fair certainty" (*mit ziemlicher Bestimmtheit*) to the last quarter of 1858. Listing's notes mention the strip quite often, for the first time in July of 1858. Hence, independent, prior discovery is sometimes ascribed to Listing. It has been widely overlooked that among the papers left behind by Möbius was a note on bordered, one-sided surfaces which said of the strip that it possesses

> ... noch die merkwürdige Eigenschaft, dass man von irgend vier in ihrem Perimeter auf einander folgenden Puncten P, Q, R, S den ersten mit dem dritten und den zweiten mit dem vierten durch zwei Linien [PR] und [QS] verbinden kann, welche in der Fläche selbst liegen und dennoch einander nicht schneiden ...

(... also the remarkable property that on its perimeter one may mark any four successive points P, Q, R, S and connect the first with the third and the second with the fourth by two lines [PR] and [QS] which lie entirely within the surface and yet do not intersect each other ...). In brief, the perimeter admits non-intersecting diagonals. Möbius adds:

> Nach einer mündlichen Mitteilung von Gauß. Wodurch G. zur Betrachtung dieser Fläche geführt worden ist, ist mir unbekannt.

(After an oral communication from Gauß. What led G. to consider this surface is not known to me.) Oddly enough, these non-intersecting diagonals are also mentioned by Listing in his quoted note of July 1858 as the characteristic property of the strip. Is this a convergence of ideas? Or did Gauß tell Listing as well as Möbius? Whatever the truth, priority in time belongs to Gauß, yet the all-important one-sided nature was recognized by Möbius.

When a ladder is twisted into a Möbius strip, its rungs are readily seen to become such non-intersecting diagonals. A widely reproduced woodcut by the Dutch artist M.C. Escher shows a Möbius strip made from a ladder-like material, with nine huge ants crawling around it in unending single file. Metaphorically, then, Gauß and Listing saw only the rungs, but Möbius also noticed the ants. The strip is properly named after him.

9. Later years

While still writing the *Census*, in 1861, Listing was elected one of the very few members of the mathematical section in the Göttingen Academy (then called *Königliche Gesellschaft der Wissenschaften*). After the publication he took up a variety of minor matters, mostly in spectroscopy, and in instrumental, atmospheric and physiological optics. In the midst of all this the Listing family very nearly slithered into a catastrophe due to overwhelming debts, and had to be rescued with a direct intervention by the ministry in Hannover (set in motion by Sartorius, in stalwart return for Listing's loyalty). Of course, the embarrassing event could hardly remain a secret and resulted in increasing social isolation.

Next came a short foray into topology. The extreme generality of the *Census* had exempted Listing from the need to define what is meant by "polyhedron". He touched upon it in a paper dated 1867 which contains mainly a retrospective critique of L'Huilier who in 1812 had noted exceptions to Euler's theorem as it was then understood, but substituted an alternative definition that was again full of lacunae. The paper also dealt with closed, plane curves having only double points, as in the white-on-black figure on the cover of the *Vorstudien*. Listing shows that the rule "The number of pieces exceeds the number of crossings by 2" is contained in the *Census-Theorem* as a special case. After that, he left topology altogether. Thus the reception of his ideas by others took place without his active participation and need not be described here; see especially the articles on knots (M. Epple) and on graphs (R.J. Wilson).

An isolated occurrence may be mentioned, though. Maxwell in his *Treatise* had quoted both the *Vorstudien* and the *Census* in the context of integrability conditions, line and surface integrals, and solid angle calculations. He also had other concerns. In particular, various studies of statics had led him to networks of force vectors, and to frameworks of rigid members. In four papers on reciprocal figures, force diagrams, hills and dales, dated 1864–1870, he repeatedly quoted Listing and the *Census*. However, his results did not find their way into engineering practice and remained a dead end as far as Listing's standing was concerned.

In the 1870's Listing again worked on optics, including a book on reflection prisms, and then undertook an intensive study of the figure of the earth. Undoubtedly he had been introduced to this ramified topic by Gauß who in the 1820's had provided many accurate data through his geodetic labours, then enriched cartography and differential geometry hand in hand, and also contributed the main elements for the later definition of the geoid. Listing attempted a synthesis in two large, painstaking memoirs in 1872 and 1877. Both were superseded in 1878 by Heinrich Bruns (he of the algebraic integrals in the three-body problem). With a booklet of only 47 pages Bruns put the entire subject on a firm, lasting foundation through a sound application of potential theory. Listing had missed the salient point over all the detail. Still, just then some recognition arrived from abroad. In 1877 the University of Tübingen awarded him the Dr. med. h.c. for his contributions to ophtalmology, and in 1879 he was elected a member of the Royal Society of Edinburgh. And he remained active to the end, busy in many areas, and directing dissertations on subjects ranging from centrifugal pumps through meteorology into electromagnetic theory.

Listing was felled by a stroke on 24 December, 1882.

10. Sources

Listing has been neglected by historians and biographers alike. My own last-minute (1985) contribution to *Neue Deutsche Biographie* **14**, 700–701, is still inaccurate. For adequate documentation, see my article "*Gauß und Listing: Topologie und Freundschaft*", Mitteilungen der Gauss-Gesellschaft, Nr. 30 (1993), 2–56, with much collateral material, a census of Listing's publications, and the entire text of the letter to Müller.

When Leibniz called for a *Geometria situs* he envisaged not topology but a vector calculus of sorts. Gauß misunderstood, like many others (although not Listing in the *Vorstudien*). For the full story, see Michael J. Crowe, *A history of vector analysis*, University of Notre Dame Press, Notre Dame (1967), pp. 3–5, with accurate references and translated sources. Further documents, hitherto unknown, have been published with commentary in G.W. Leibniz, *La caractéristique géométrique*, J. Echeverria, ed., J. Vrin, Paris (1995).

The posthumous papers of Gauß on topology are in his *Gesammelte Werke*, Vol. V, 605, and Vol. VIII, 271–286 (with terse commentary by Paul Stäckel). The early discussion of "right" and "left" appeared in his (anonymous) announcement of the second paper on quadratic residues in *Göttingische gelehrte Anzeigen* **1** (1831), 625–638; also *Werke*, Vol. II, 169–178.

On the broader influence of Gauß, see Jean-Claude Pont, *La topologie algébrique des origines à Poincaré*, Presses Universitaires de France, Paris (1974). He also comments extensively on Möbius, with the conclusion

> Si Euler, Listing, Riemann et d'autres ont donné des béquilles à la topologie, Möbius lui a donné des ailes.

(If Euler, Listing, Riemann and others provided crutches for topology, Möbius gave it wings.)

For detailed, technical appreciations of the *Vorstudien* and the *Census*, see Pont, *op. cit.* (with some translation and printing errors); also Angelo Tripodi, "*L'introduzione alla topologia di Johann Benedict Listing*", Atti e Memorie della Accademia Nazionale di Scienze, Lettere e Arti di Modena **13** (1971), 3–14, and "*Sviluppi della topologia secondo Johann Benedict Listing*", *ibid.*, 15–24. The *Vorstudien* have been translated into Russian: *Predvaritel'nye issledovaniya po topologii / s Iogann Benedikt Listing*, Gosudarstvennoye tekhniko-theoretitcheskoye izdatel'stvo, Moscow (1932), with introduction and technical comment by the editor E. Kol'mana.

Listing's note of July 1858 has been published by Paul Stäckel, "*Die Entdeckung der einseitigen Flächen*", Mathematische Annalen **52** (1899), 598–600. On the papers of Möbius, see Curt Reinhardt, "*Mitteilungen aus Möbius' Nachlass*", in August Ferdinand Möbius, Gesammelte Werke, Vol. 2, Felix Klein, ed., Hirzel, Leipzig (1886), pp. 513–708, especially pp. 517 and footnote 197 on pp. 540–542.

The definition of "polyhedron" has remained vexing to the present day. See Branko Grünbaum and G.C. Shephard, "*A new look at Euler's theorem for polyhedra*", Amer. Math. Monthly **101** (1994), 109–128; and Walter Nef, "*A new look at Euler's theorem for polyhedra: A comment*", *ibid.* **104** (1997), 150–151; with much literature but without reference to the *Census-Theorem.*

CHAPTER 34

Poul Heegaard

Ellen S. Munkholm and Hans J. Munkholm
Department of Mathematics and Computer Science, Odense University, Odense M, Denmark DK-5230
E-mail: hjm@imada.ou.dk

1. Poul Heegaard's autobiographical notes

When we started our investigation of Heegaard's life and career, it was easy enough to locate his mathematical publications, but we found only very few accounts of his life, in general. In particular, we could locate only one obituary. We then searched the Internet for persons carrying the name Heegaard. This led us to contact a number of e-mail addresses in Norway, Denmark, USA, Sweden and Switzerland. A few of the persons we reached this way knew that they were related to "our" Heegaard. Among those was Poul E. Heegaard, a Ph.D. student of computer science at Trondheim University, Norway, and a great grandson of Poul Heegaard. He gave us the very welcome news that Poul Heegaard had actually left roughly 130 pages of handwritten autobiographical notes, [17], and he generously supplied us with a copy.

The notes were written in 1945 (in Norwegian) when Heegaard was 73 years old and they were meant as a family history told to his children and grandchildren, but they do contain a lot of information that is relevant to our study. Unfortunately, a few pages are missing precisely at two critical points in Heegaard's life. Nevertheless, the notes supply much more information about Heegaard's life than any other single source we have found, and we have chosen to use them as a skeleton for the following account. Our rather extensive quotes from this source are both indented and between quotation marks, as in the following example which refers to the semester Heegaard spent in Paris in 1893.

> "Later, I have always regretted that I accepted the advice not to attend lectures by Poincaré, who was claimed to be untelligible. His very intuitive exposition has later on been of great importance to me when I met it in printed books."

Heegaard wrote long, rather complicated sentences, probably influenced by his regular use of German. In our translation we have attempted a compromise between the original style and current usage in English.

HISTORY OF TOPOLOGY
Edited by I.M. James

Poul Heegaard (1871–1948)

A transcript of the autobiographical notes (in Danish) has been produced at Odense University and is available on the Internet (`http://www.imada.ou.dk/~hjm/heegaard.html`).

In our account we also quote from other sources, again in our own (occasionally somewhat free) translation. Such quotes are not indented. Here is an example from a letter that Heegaard wrote to Jakob Nielsen in 1935:

> "I have been terribly busy with University work. Therefore, I have not written to you, nor to Dehn. Where is Dehn at the moment? I made all the mathematicians sign an application to the Science Academy for 2000 Kr. so that he could come up here in April of 1936, attend the Congress and give lectures and exercises. Now the board is trying to figure out that a grant would be against the statutes."[1]

2. The early years (1871–1883)

Poul Heegaard was born 1871 in Copenhagen where his father, Sophus Heegaard,[2] was a professor of philosophy. As a young student, Sophus Heegaard had wavered between theology and astronomy when he registered at the university. Influenced by his father, he settled for theology, but for his dissertation he switched to philosophy. He retained an interest in science, and it is reported that he pursued mathematical studies with great eagerness throughout his life.

As we shall see, Sophus Heegaard managed to transfer at least three of the above four areas of interests to his son even if he died before Poul turned 13.

At first, there was little indication of a future mathematical career for the schoolboy Poul. Thus, in his autobiographical notes, [17], Poul Heegaard reports that he was not very adept at arithmetic. In particular, when he was examined in the addition table, he would always secretly try to reach the result by counting on his fingers.

> "I really only learned the addition table in [my high school freshman year] when the use of logarithms forced me to take it up on my own. I still remember that the occasion for this was my discovery that I had consistently believed seven plus eight to be seventeen. ... I never learned any mental arithmetic, a fact that has later been a great nuisance for me. By the way, this matter deteriorated even further towards the end of elementary school [i.e. around the age of 10 or 11] when our teacher of arithmetic, a young student ... discovered that I had a flair for algebra. Apparently it amused him to replace the dry teaching of basic arithmetic by such abstract teaching of mathematics. This proved fateful to me in two ways. For one thing, it further weakened my basic arithmetic skills. For another, it developed my mathematical abilities at an early stage – and thus led my surroundings to drive me towards an occupation with mathematics for which I did have a talent, but for which I do not have the burning interest that I have met in others. I feel so particularly, when I compare with my interest in astronomy. ... [The interest in] astronomy grew during the dark, starry nights. I would often sit astride the ridge of the roof and compare a star atlas with the firmament itself. Later, during starry nights, I have often felt the well known constellations to be faithful friends."

[1] This may refer to the dispute concerning the participation of Germans in the 1936 International Congress in Oslo.

[2] Information about Sophus Heegaard is from [23].

3. High school and University of Copenhagen (1883–1893)

Nevertheless, in Poul Heegaard's description of his high school years (1883–1889, Metropolitanskolen, Copenhagen) he notes that

> "The teacher who became most important for me was the mathematics teacher, Eigil Schmidt ... an excellent teacher for those students who took an interest in the subject. But he did not have sufficient patience for the backward pupils."

Later in high school, the same teacher

> "Eigil Schmidt gave an excellent course in mathematical physics, sufficient in scope for my minor in physics at the university."

After high school graduation in 1889, Poul Heegaard started studies at University of Copenhagen. His father's death had left the family in a bad financial situation, so Poul had to finance his studies by tutoring other students and grading papers at the nearby Technical University (Polyteknisk Læreanstalt). Nevertheless, he finished his Master's Degree in mathematics with a minor in astronomy, chemistry and physics, in less than four years. In mathematics, his best known teachers were H.G. Zeuthen at whose lectures on enumerative geometry he was the only student, and Julius Petersen whose contributions to complex function theory would later become a model for much of Heegaard's dissertation. Lectures in astronomy by Thiele included an introductory and an advanced series as well as a special series on quaternions. Heegaard's strong high school background in physics mentioned above was still present, so

> "the physics curriculum I picked up on my own. I found chemistry very interesting but the teaching was very abstract and unpedagogical. Jul. Thomsen did arrange for complete fireworks of experiments, but you remembered very little of it. And S.M. Jørgensen was very dry. We only had to memorize his big textbook on inorganic chemistry, and we got no real impression of the substances from watching him shake some glass jars containing a white powder."

When Heegaard registered for the final examination in December of 1892,

> "Zeuthen thought it premature, but I had become secretly engaged to Magdalene and longed for my degree."

The first part of this comprehensive examination was a "thesis" which had to be worked out between January 23 and March 4 1893.

> "Its purpose was to study Chasles' description of algebraic curves in a surface of second order by characterizing them in terms of the number of intersection points with the generators in the two generating systems. I recall as a wonderful time these six weeks when I could concentrate completely on the thought processes. Zeuthen was very satisfied with my work."

In May of 1893, there followed two written six hour exams in mathematics and one in physics, as well as an oral examination where

> "all four professors, Zeuthen (mathematics), Thiele (astronomy), Christiansen (physics) and Jørgensen (chemistry) sat around a semi circular table at the center of which I had been placed. The examination started at 10 o'clock and the torture lasted until 2 o'clock."

The final outcome, presented by professor Zeuthen immediately after the oral on May 6, 1893, was the desired "Admissus".

4. In Paris and Göttingen (1893–1894)

After his examination, Heegaard wanted to go on a study trip abroad. He managed to secure some funds from a private foundation and a much needed addition came from the university.

> "Personally, I would like to go to Göttingen to study with Felix Klein with whose intuitive exposition I felt so confident. But Zeuthen was strongly pro-French and wanted me to study in Paris. This I then went along with."

Thus, in August of 1893, Heegaard left for Paris, bringing along letters of introduction to a number of French professors. He reports to have been well received first by Mannheim and later also by Picard,

> "... but nothing more came of this courtesy call.
> Professor Darboux' behaviour was more remarkable. In his private residence the maid told me that he would receive only during his office hours at the Sorbonne. ...
> I went there and delivered Zeuthen's letter of introduction to a Cerberus in the anteroom. After I had waited there for three quarters of an hour, the Cerberus said: 'Monsieur might as well leave immediately, for the professor tossed the letter in the wastepaper basket after reading it'."

Heegaard had arrived early in order to improve his command of the language, but in this he did not succeed. Nor did things improve when the lectures started.

> "The French were extremely withdrawn towards foreigners. Therefore, I only associated with the Danish Norwegian colony, much to the detriment of my progress in speaking French.
> ...
> I attended lectures by Picard. He covered his book 'Leçons d'Analyse' word for word. When he came into the room, met by applause, a caretaker ... would precede him at a light trot. He would pour water into a glass and place small bits of sugar beside it. Some of the students kept an account of the number of sugar bits consumed by Picard during each lecture, and intended to expand the results in spherical harmonics after the semester.
> ...
> I also heard lectures by Jordan at the College de France. Nor from him did I get exciting expositions. He went through the proof sheets of his 'Cours d'Analyse'. Occasionally, he would pause and pencil in a correction."

The autobiographical notes give no indication that Heegaard worked (hard, or even otherwise) on mathematical research, but he reports that his mathematics was of some assistance to a Danish professor Tscherning who was investigating Listing's law on eye movements. Also, he made the drawings for a medical dissertation, submitted to University of Copenhagen by a Danish MD, named Rée.

> "The scant return on my mathematical studies matured in me the thought that I would spend the next semester in Göttingen. After an exchange of letters with Zeuthen, I finally got a very friendly letter from Klein. He wrote that it would be a special satisfac-

tion for him if I would end up feeling that I had studied better at Göttingen than Paris. As a preparation, I had to read a paper by Riemann on the $\mathcal{P}$ function. I rushed to the Bibliothèque Nationale and got hold of the relevant issue of Mathematische Annalen in the reading room. It was heavy stuff, but I did get far enough to be orientated when I arrived at Göttingen.

The remaining time I spent in part by fulfilling an old, rather peculiar, desire, namely to learn Chinese. . . . I wallowed in the studies, in part at Bibliothèque Nationale, in part at Bibliothèque Mazarin and Bibliothèque Sainte Geneviève."

In Göttingen, the reception was much more pleasant. Within hours of Heegaard's arrival, Klein himself had taken him to his new quarters. And very soon Heegaard presented his first seminar lecture – about harmonic functions in Tait's mathematical [theory of] electricity. Later, he attended two lecture series by Klein (an elementary one including angle trisection and the like, and a more advanced one on differential equations) and one by Heinrich Weber (Higher Algebra). In addition,

"Klein had me give two lectures in the 'Mathematische Gesellschaft' with a summary of Zeuthen's work on enumerative geometry. He also discussed with me the idea that would later form the basis for my dissertation. Altogether, there was a scientific atmosphere which stimulated me very much – stronger than anything I have ever met again.

. . .

When the semester ended, early October [of the year 1894], I returned to Copenhagen, very satisfied with the result of my study visit. In particular, I had the idea for my dissertation. Now, the object was to get a secure occupation so that I could marry."

5. Work on the dissertation (1894–1898)

The desired occupation Heegaard found as a teacher of mathematics in two high schools. Later on, more high school jobs were added along with some tutoring at the Technical University, and in 1896 the financial situation finally allowed Heegaard to marry Magdalene. Their first child, Lorenz, was born a year later. Under these circumstances, Heegaard recalls that

"of course, the dissertation progressed only slowly. Moreover, I felt strongly out of it among the mathematicians. While I had been in Göttingen, two closed coteries had formed, and I was made to feel odd man out in many ways. When I had announced a lecture in the Danish Mathematical Society, and it ended up being sabotaged away, I resigned from the Society at the end of the year. Presumably, this was a tactical error, but I have always been timid where it looks like war."

It is easy to speculate that this represents an early start of those infights that would make Heegaard resign his chair at University of Copenhagen more than 20 years later. However, one must bear in mind that the text quoted here was written fifty years after the events took place.

One day, apparently in the fall of 1897,

"one of the older mathematicians who knew of the problems I worked on [said to me] 'It is unpleasant for you that Poincaré has solved your problem.' He had seen it in an article Analysis Situs in the Journal de l'École Polytechnique, [obviously [30]]. At the time, this journal was in circulation among members of the Society of Sciences and Letters, and it would be available to the public in the University Library only after a long time.

> I now ordered the issue in question through my bookseller. I noticed immediately that Poincaré's treatment was based on a gross mistake. There was a reference to a book by Picard and Simart, [presumably [29]], on functions of two independent variables. When I got hold of this, I saw that Picard had the same mistake. During the Christmas break I then plunged into a critique of Poincaré's and Picard's accounts, in which I demonstrated the mistake. Already in January [of 1898], I could hand in a dissertation, [1], entitled 'Preliminary studies of a theory of connectivity for algebraic surfaces'."

6. The dissertation (1898)

Heegaard's counterexample to Poincaré's original formulation of the duality theorem and the role this example played in the development of algebraic topology (by "forcing" Poincaré back to the matter) is well documented, e.g., in [24, 31, 34], and we shall not go into that aspect here. Nor shall we go directly into the other main contributions from the dissertation, *Heegaard decompositions* and *Heegaard diagrams*. Here we refer the reader to [34] and the article by C. Gordon in the present volume.

The Poincaré duality conterexample and Heegaard's stay in Paris had led us (and many others, we have reason to believe) to assume that Poincaré had been an unofficial advisor for Heegaard. As we have seen in the above description of the time in Paris, this was certainly not the case. Actually, we have found no indication that Heegaard ever met Poincaré, then or later.

For a dissertation there was no official advisor, but we have seen that Heegaard's discussions with Klein certainly played a role. So did also Julius Petersen's lectures on Function Theory, [28], which Heegaard had attended (in an earlier version) at University of Copenhagen and in which the role of Riemann surfaces as a tool in the study of (complex) algebraic functions of one variable, had been emphasized. Furthermore, the Riemann surfaces themselves had been studied by puncturing them and deforming the result into a normal form. It is this approach that Heegaard sets out to generalize to the case of algebraic functions of two variables. In his own words from the introduction to the dissertation:

> "To carry out an analogous investigation of the connectivity of an algebraic surface $z = f(x, y)$, one must first form a fourfold infinite collection of elements to which one can associate all the value pairs (x, y) that can be obtained by letting x and y assume all possible complex values, independently of one another."

Today, it is easy for us to say that Heegaard is looking simply for $S^2 \times S^2$, but such abstraction was not available, so Heegaard begins by describing how one may interpret $(x_1 + ix_2, y_1 + iy_2)$ as a point (x_1, x_2, y_1) in ordinary 3-space equipped with a contour number y_2; how lines, planes, and (flat, 3-dimensional) spaces look in this set up; also circles, rotations, angles, distances, etc.

Actually this description of points in 4-dimensional space had already been given by Lie [12] some thirty years before, but we see no indication that Heegaard knew of Lie's work when he wrote his thesis. We shall return to this question later when we describe Heegaard's contribution to the publication of Lie's collected works.

Once Heegaard has the 2-variable analogue of the Gauss sphere at his disposal, he turns to the generalized Riemann surfaces as we see in the following continuation of the quote above:

> "By covering this several times, introducing 'branching surfaces' in a suitable way, and connecting these by means of 3-dimensional creations through which the different 'layers' can be connected to one another, one can create a 4-dimensional manifold in which the connectivity of the algebraic surfaces can be studied."

Here the word "connectivity" could be interpreted to mean the Betti numbers, but Heegaard really wants more:

> "Already before I knew ... [Poincaré's and Picard's work, [30] and [29]] ... I had decided to try a different road than that of Riemann and Betti, viz. to attempt a generalization of Jul. Petersen's Puncturing Method ....
> ... it appeared unfortunate to me that the n connectivity numbers were not sufficient to characterize a manifold, topologically, when $n > 2$. When I became acquainted with ... Poincaré's [work] I began to falter in my choice as I compared the elegant methods I met here with the somewhat hard and clumsy theory that I worked on. But since I seemed to discover that the road chosen by me would throw light over circumstances which were not clearly exposed in the other way, and since furthermore I got tools in hand to find *sufficient* conditions for the equivalence of n-dimensional manifolds I decided to continue in spite of the difficulties I met."

The difficulties were so great that Heegaard never really got to the 4-dimensional case. But he illustrated the idea in three dimensions and thereby immortalized his name.

The defense act went well but did not lead to mathematical acceptance in Denmark as we can see from the following excerpt.

> "I had sent my dissertation to Picard and Poincaré. The latter asked me about different things that he had not understood in the Danish text. Thus I wrote a summary of the dissertation in French for him. This led him to write a paper supplementing his original treatise, and thus my dissertation became known abroad even if it was written in Danish. In Denmark public opinion held it worthless and completely ridiculous. One of my foreign friends noted this in a conversation with one of the older mathematicians who had to admit at the same time that he had not read it."

A French translation, [5], appeared in 1916. A preprint from Odense University, [32], contains a translation into English of the latter half of the dissertation.

7. The Dehn–Heegaard Enzyklopädie article (1907)

After his dissertation Heegaard taught for twelve years in a variety of naval and military academies in Copenhagen. Hours were long – eight hours a day, six days a week, typically. But Heegaard remembers this period fondly

> "I was now on firm ground and could live happily and unaffected by whatever the coteries of mathematicians might be up to for a score of years. But, of course, there was extremely little time for advanced mathematical production."

In spite of this, when he was asked to report on Analysis Situs in the *Enzyklopädie der Mathematischen Wissenschaften*, Heegaard accepted, and

> "started the work with great pleasure, and – lack of time notwithstanding – finished an outline and a bibliography. However, it was difficult to get the time and quiet needed to work out the theoretical introduction. Moreover, quite senselessly, I let myself be influenced by a variety of malicious comments on my work in topology. Therefore,

> I asked Franz Meyer [the Enzyklopädie editor] for an assistant. It was then arranged that I should write the article with the young German mathematician Max Dehn, Dr. from Göttingen.
>
> ...
>
> In the meantime, Dehn had become Privatdozent in Kiel, and in the summer of 1905 I went down there to work with him. I now initiated him into my viewpoints and he began to work on the general introduction, which he finished beautifully during the next winter."

Here, Heegaard seems to think of Dehn as a junior author, but the official version, appearing in a footnote in the article itself, [2], is different:

> "Of the two authors, Heegaard did the preliminary literature studies, and also took an essential part in the work. Responsibility for the final form of the article is Dehn's."

In his Heegaard obituary, [25], Heegaard's student, Ingebrigt Johannsson, gives a version, which he probably heard from Heegaard at some stage. This has Heegaard and Dehn meeting at a conference in Kassel in 1903. It further reports that they discussed foundational problems in topology on the train back between Göttingen and Hamburg and continues:

> "Dehn believed that that one should postulate just enough axioms to let the topological essence stand out clearly, something that had never been done before. Here, in the railroad compartment, combinatorial topology was created. Heegaard was enthusiastic, and proposed that they would write the article jointly."

The axioms referred to by Johannsson are purely combinatorial. They treat abstract *polyhedral complexes*; *subdivisions*; *n-manifolds*, defined to be complexes in which each vertex has an $(n-1)$-sphere as its link, just the way we expect it today; *orientations*, under the name indicatrices; *homeomorphisms*; *singular subcomplexes in manifolds*, defined somewhat clumsily by today's standards, but workable; *homotopies* that are more restrictive than we might anticipate today; *isotopies*; etc.

Such an approach was not universally admired. Thus, Klein, [27], calls it

> "... quite abstractly written ... begins with the most general formulations of basic notions and facts from Analysis situs, construed by Dehn himself. From there, everything else is then deduced by pure logic. This is in complete contrast to the inductive presentation that I always recommend. To be fully understood, it really presupposes a reader who has already worked through the field thoroughly in an inductive manner."

The authors are well aware that the intuitive content is important. In fact they characterize Analysis situs as *a part of combinatorics, distinguished by its "anschauliche Bedeutung"*, i.e. by its intuitive/visual interpretation/impact/importance. In a section called "Das Anschauungssubstrat" immediately after the combinatorial axioms, they record axiomatically those properties of the ordinary three dimensional space which make it possible to interpret the abstract developments concretely. And they explicitly state that it is only through such interpretation that the whole theory acquires its value.

What then is the value that it acquires, i.e. what do Dehn and Heegaard do with the formal apparatus they have built? In the following summary we follow Dehn and Heegaard's break down of the material according to the methods used.

Complexus is their name for that part of the theory in which neither subdivision, nor homotopy is used. In this part they give a thorough overview of the existing literature on graphs ("Liniensysteme"). And in higher dimensions, a treatment of homology, Betti numbers, torsion coefficients, and Euler's formula. Even though this necessitates some use of subdivision, expositions of Poincaré duality and Poincaré's counterexample to the (original) Poincaré conjecture are also given under this heading.

Nexus is the heading used when subdivisions, but no homotopies, are allowed. The main problem here is to find necessary and sufficient conditions under which two given manifolds are homeomorphic. Their axiomatic approach allows them to establish a normal form for surfaces (as a disk with a number of twisted bands or double bands attached, and finally capped off by another disk) and thereby solve the main problem in dimension two. Among the applications given is the so called *proper Euler formula* expressing the Euler characteristic of a surface in terms of the number of boundary components, and the maximal number of disjoint, closed, simple curves that do not separate the surface. They argue that this form is much deeper than the one with Betti numbers. And especially point to the fact that the latter is easy to establish in all dimensions whereas the former has no known generalization to higher dimensions, "because, so far, no presentation in normal form is known for M_n with $n > 2$".

Connexus, finally, denotes the theory obtained when homotopy and isotopy become essential. One section reports on Jordan's, [26], classification up to homotopy of all closed curves through a given point on any orientable surface. In modern terms this is, of course, the determination of the surface fundamental group, and Poincaré's recent introduction of that notion is briefly recorded. In another section, there is an extensive report on the existing literature on knots and links.

Dehn and Heegaard did not find it possible to fit the study of singularities into their complexus/nexus/connexus classification scheme. A separate section, entitled *Manifolds with singularities*, consists mainly of a survey of Riemann surface theory. Also mentioned are Gauss' study of singularities for curves and Boy's proof that any surface can be realized in $\mathbb{R}^3$ with singularities no worse than double curves with one threefold point. The latter of course refers to Boy's construction of the immersed surface carrying his name, [21].

8. Not quite turning astronomer

In 1901, Heegaard's childhood stargazing experience matured in the form of a series of popular articles that appeared in a Danish weekly magazine. In book form, under the title "Popular Astronomy" (1902), it became very popular and it was later translated into Swedish and German. Moreover, it became the starting point for a long series of popular lectures on astronomy, often accompanied by small publications.

During his visit with Dehn in Kiel, Heegaard participated in professor Schur's seminar on astronomy, and also otherwise

> "I had the opportunity to cultivate my old love, astronomy. In Copenhagen I had often wanted to to get this opportunity, at times I had even had plans to study astronomy. After my dissertation I went so far as to lecture on astronomy at Copenhagen University. But this crossed Professor [of astronomy] Thiele's plans, and I never got any further. Especially, I never had the opportunity to learn observation skills. But here, I was met favourably by Professor Harzer, the Director of the Kiel observatory. On the one hand, I performed observations with the observer, Professor Kobold, using the big meridian instrument, along with a sister instrument in South Africa the largest on Earth. On

the other hand, I worked with a Swedish astronomer, Dr. Strømgren ... I recall an experience, which really throughout the years contributed to the cooling of my desire to be an observing astronomer. One starry night at around 2–3 o'clock I was sitting in the meridian house, assisting Professor Kobold. He would give the times for the star passages, and I would record them. Then he got tired and wanted to take a break. We were both looking through the slit in the roof where innumerable stars sparkled in the dark night. Then he suddenly said: 'Eigentlich ein lächerliches Geschäft'. Here sat a man, who had reached all I longed for, the insight, the position, and one of the world's best instruments – and then basically he found the work 'a ridiculous business'. I began to fear that it might turn out the same way for me."

A year after Heegaard's visit to Kiel, in 1906, Professor Thiele retired from the astronomy chair at University of Copenhagen, and several friends urged Heegaard to apply.

> "Zeuthen spoke very diplomatically when he answered one of my most insistent friends: 'Dr. Heegaard is probably the best judge of his own qualifications'. In that he was right. For although I did not have a bad standing concerning theoretical knowledge, I lacked sufficient experience in observation and I completely lacked scientific publications. It therefore never occurred to me to apply."

However, this did not mean that Heegaard gave up his interest in astronomy. He became a cofounder and the first chairman of the Astronomy Society in Denmark in 1916, and his popularizing lectures continued, also after he went to Norway. So did the small accompanying publications, including one called 'The Childhood of Astronomy. Lectures for prisoners of war' published 1917 by University of Copenhagen and The Danish Red Cross and translated into French, Italian, Russian and German, [6].

9. Professor at University of Copenhagen (1910–1917)

Heegaard's mentor, Professor Zeuthen, retired at the age of almost 71 on February 1, 1910. Already a few months earlier, applications to succeed him had been solicited, and Heegaard reports that he was assailed with calls to apply,

> "but definitely did not feel qualified. My best years of preparation had elapsed under a great work load and without any support from the Carlsberg Foundation, so I had only a scant scientific production to show, really only my dissertation and the Enzyklopädie article with Dehn. Moreover, I had a well paid job, around 15.000 Kr. a year, which I filled to the satisfaction of everyone. Also, in these circles I lived protected against the hostile plots as long as I kept quiet. When all the calls to apply were lost on me for a period of six months, my friends ... turned to a different angle. They began to work on my mother. ... At last, the pressure was so great that I poured out my troubles to Zeuthen. He completely concurred that I felt unqualified, but said that I ought to submit an application anyway. He would then see to it that I would not be appointed. Thus I submitted a very short application.
>
> This, however, ... gave the coterie of university mathematicians the means to prevent the Technical University mathematicians from forcing their entry into the Mathematical Faculty of the University. They had secured very laudatory references from Poincaré, Picard, Klein and Franz Meier, probably also from Hilbert.
>
> There followed a succession of dramatic events which I shall not report on here since I had nothing to do with them and stood by quite powerlessly. As a result I was appointed in February of 1910."

The above is one of the many instances where Heegaard refuses to state explicitly who are his enemies. And what are they really fighting about. Heegaard had only two competitors for the chair. One of them, J. Hjelmslev, already held a chair at the Technical University, and thus must have had a formal advantage over Heegaard. Was it him that the University mathematicians wanted to keep out? If so, they only succeeded temporarily: When Heegaard resigned seven years later, Hjelmslev became his successor.

Heegaard reports that his time at University of Copenhagen was difficult, also financially. As a professor his annual salary was only 3.000 Kr. He had a wife and six children, he had no private means worth mentioning, and he had been used to a yearly income of 15.000 Kr. To make ends meet, he kept a job at the naval cadet school, where he made 5.000 Kr. a year. But the work load made research conditions difficult.

> "After all, my best youthful years had passed. My teaching experience did make it possible for me to attend to my lectures to the satisfaction of the audience. But in many ways I was hampered by the other group of mathematicians who also had their allies in the Faculty of Science."

The above paints a bleak picture of Heegaard as a professor at Copenhagen University, but there are also successes to be reported.

Already before his appointment, in 1908, Heegaard had been elected Danish representative on the international committee for the teaching of mathematics, IMUK, an organisation that had been founded earlier that year, at the International Congress of Mathematicians in Rome, with Klein as chairman. Heegaard took part in the first meeting in Bruxelles, and in 1912, IMUK published his comprehensive report on the teaching of mathematics at all levels in Denmark, [3].

A mathematical laboratory had been created by the University in 1907. According to [33] it was Heegaard who first organized a library there and who undertook the creation of a collection of mechanical and kinematic models for teaching use, probably inspired by what he had seen in Göttingen. Furthermore, he did a lot of the practical work connected with the 2nd Scandinavian Congress of Mathematicians which took place in Copenhagen in August of 1911.

Two years later, Heegaard also participated in the 3-rd Scandinavian Congress in Kristiania (later to be called Oslo) where he presented a short paper on graph theory. This, in 1915, became the third mathematical research paper, published by the, then, 43 year old professor.

As mentioned earlier, in 1916 the French Mathematical Society published a translation of Heegaard's dissertation, [5]. Quite understandably, Heegaard was pleased to see this happen and he chalks it up as a "retraction by the vulgar real world" of the slighting press that the coteries of mathematicians had given the dissertation. But apparently it did not inspire him to continue the line of research from the dissertation.

In his letter of resignation (see below) Heegaard complains that his many duties do not leave him time for such a continuation. As we have seen, he did have the time for an extensive popularizing activity in astronomy. He also wrote on the geometry of Trondheim cathedral and several high school related publications. Was it really time he lacked – or was he closer to the truth when he wrote (as quoted earlier) "... mathematics for which I did have a talent, but for which I do not have the burning interest that I have met in others"?

10. Resigning the chair in Copenhagen (1917)

The University of Copenhagen 1917 Yearbook, [19], records a letter of resignation from Heegaard, dated January 5. The reasons given are as follows:

> "1. My work at the University comprises Geometry, Rational Mechanics, Elementary Mathematics, History of Mathematics and General Mathematics for Actuaries; in addition I am librarian at the Mathematical Laboratory Library. My work has been of such magnitude that neither have I had any otium for research, nor – in spite of all the energy spent – have I been able to discharge my work to my own satisfaction.
>
> 2. In addition, often my views on various matters in the Faculty have been so different from those of my colleagues as to further contribute to making my University work too onerous for me."

The Faculty did not immediately recommend the resignation to the University senate. Instead, concerning point 1 it asked Heegaard to propose changes in his duties that would alleviate the problem. As for point 2, it asked for explicit examples where disagreements with colleagues had been serious enough to make Heegaard's work onerous.

Heegaard answered that any discussion of the work load should be taken up with his successor. And that he had no desire, now or ever, to further discuss the matters in his point 2.

Kurt Ramskov, in his thesis [33], has touched upon the matter of Heegaard's resignation because of its relation to the main character of the thesis, Harald Bohr. Based on letters from Nørlund[3] to Mittag-Leffler at Stockholm University, and on contemporary accounts in the Danish tabloid Ekstrabladet, Ramskov reports that the fight could be construed to start with the Danish sale of the (then Danish) Virgin Islands to the United States. This sale was finalized in 1916, at a price of 25 million $ of which more than 1 million $ was (at some stage) intended as extra funding for the University. With the exchange rate and the salary level of the time, the interest on such a sum would probably finance 40–50 new positions at the University. This naturally raised hopes and expectations among young researchers. Among the hopefuls were Harald Bohr and Hjelmslev. They both held positions at the (less prestigious) Technical University, but they now hoped for (additional) University appointments. However, Heegaard and his colleague, Professor Niels Nielsen, were opposed, and at first they won the battle (even if it may be said that Heegaard lost the war). Thus, on January 6, 1917, Ekstrabladet brings the following account:

> "Harald Bohr's and Hjelmslev's strong desires, and no less vigorous work and propaganda, for admission into the Faculty as Docents [$\sim$ Adjunct Professors or Instructors] has created a lively mood, and it has long been felt that one walked among loaded bombs.
>
> Yesterday, the first bomb exploded: Professor Heegaard submitted his letter of resignation. In academic circles it is no secret that he has not been an unequivocal admirer of Professor Harald Bohr, Esq. He is certainly not alone with this opinion but he is probably among those least likely to pull their punches. Now an explosion occurred. And Poul Heegaard preferred to leave, and for that one cannot blame him.
>
> ...
>
> We have told that recently Harald Bohr has threatened the Faculty of Science in a letter. If he did not get an affiliation with the University, he would leave [the country]. Groningen lay

[3] At the time Professor at Lund University, Sweden.

open to him.[4] And even if he was damned opposed to going to Groningen – it is always worth using as a threat, isn't it.
But it did not work out. In the end it was decided that Bohr would not be appointed Docent, nor Hjelmslev. When Bohr realized that his threat did not avail and that nobody feared the loss of him, he got furious. He went home, unpacked his suitcase, and said No to Groningen: now he would take revenge on the gentlemen of this country by staying and making it hot for them. And then there is the chair that Heegaard vacated. But on that, the least said the better."

When Heegaard wrote his autobiographical notes, he may or may not have violated his promise never to discuss his reasons for resigning. We cannot tell because, unfortunately, two pages of the notes are missing at the critical spot. This could be a coincidence. Or Heegaard may have written an account after all, but then decided to destroy it later on. Whatever the reason, it seems certain that the missing pages, 103 and 104 in the handwritten version, have dealt with his situation around the time that he left University of Copenhagen, for p. 105 picks up as follows:

"she: 'What would father have said to this?' At that moment the telephone rang. A voice said: 'This is Reverend Dalhoff. I have just returned from a visit with my son-in-law, Professor Guldberg, in Norway. He has requested that I ask you whether you would accept it, if you were to be nominated for a chair at the University of Kristiania.' Reverend Dalhoff was the clergyman with whom my father spoke at his conversion to active Christianity. This, therefore, seemed like an answer to my mother's question, and a greeting from my father."

11. Professor at Kristiania (= Oslo) University (1918–1941)

Towards the end of the year, Heegaard's appointment at Oslo University became official, and on December 12, 1917, the Danish newspaper Nationaltidende reports that

"We met Dr. phil. Heegaard this morning at the Naval Academy where his Royal Highness the Crown Prince and other cadets were passing by during a break, and we again used the Professor title as we addressed him.

'Well, well', he answered while stroking his beautiful, slightly greying whiskers, 'but the Royal signature is still lacking.'

'Stortinget [The Norwegian Parliament] promises you 10.000 Kr. as annual revenue.'

'Yes, and that is more than I had here at the University; it is twice as much.'

'Norwegian newspapers report that you resigned down here for personal reasons.'

'They do? Let us not get into that matter. I have long since decided to bury everything that caused my leaving University of Copenhagen, and it will stay that way. But I can tell you that it had no connection to Kristiania, for the Norwegian nomination did not exist at that time.'

'Have you not taught at all since then?'

'Yes, I have had much to do, both here at this Academy and at other schools, and I am also occupied with actuarial work. But academic teaching appeals strongly to me and I have accepted the offer from Norway with pleasure.'

'And you will leave Denmark?'

'Yes, in the spring, I believe, once the signature of King Haakon is available.' "

[4] In December of 1916, Harald Bohr had been offered a chair at Groningen University in the Netherlands.

Actually, Heegaard arrived in Kristiania already in January of 1918. He tells of becoming a member of the country's oldest freemason lodge 'St. Olaf of the white leopard', where he soon became speaker and acquired many friends.

> "Thus I quickly took root and felt very happy. A question which stirred the mind strongly at the time was the restoration of Trondheim cathedral. As already mentioned, while in Copenhagen I had written an article about the geometric system of Macody Lund.[5] It was the first favourable contribution and thus had attracted attention in Norway. Now the architect Sinding Larsen and Macody Lund welcomed me with enthusiasm. In my inaugural lecture[6] I touched upon the question."

During his twenty-odd years at Kristiania University, Heegaard followed the same pattern as in Copenhagen. He was interested in many things and they seemed to keep him away from mathematical research to a large degree. And when he did enter into a major mathematical publication project (see later about the collected works of Lie), he was too busy to finish it.

He was active in the creation of *Norsk Matematisk Forening* and became a founding editor of its journal *Norsk Matematisk Tidskrift*. He chaired an influential welfare committee for students (Den akademiske dyrtidskomité) 1920–1925, and he arranged a geometric study circle for advanced students.

His interest in popularization manifested itself strongly when he became chairman for the Oslo University Folkeakademi (~ Extramural Department). In this latter context, he continued to give popular lectures on astronomy, and in a country with the enormous distances (and the climate) of Norway, this could be an arduous job. Thus he tells about a lecture tour that took him from Kristiania to Kautokeino in the extreme North of the country. First overland, 400 miles to Trondheim, then another 700 miles by boat along the coast to Hammerfest, and finally inland again for at least 150 miles, a part of which took place in a reindeer drawn sledge. All along the tour he would stop to give lectures, and up in the north these

> "were translated, sentence by sentence, into Lappish by the storekeeper who acted as an interpreter. When I asked the local minister how the translation had been, he said 'Oh, rather free'. For example, on one occasion after 1/2 hour of lecturing I found all this translation bothersome so I said: 'Now the Laps may leave. I shall continue for the Norwegians'. The Laps rose, but immediately sat down again. My words had been translated as: 'What kind of a disturbance is that you are making. Sit down again!' "

12. The collected works of Sophus Lie

On this topic Heegaard writes:

> "I had received a call to take part in the publication of Sophus Lie's collected works. Already before the world war there had been plans for such an edition, in six volumes plus a seventh volume of 'Nachlass'.

[5] In a book 'Ad quadratum' (1919) and earlier in other writings, the Norwegian historian Macody Lund (1863–1943), claimed to have found the geometrical system underlying medieval church architecture, and proposed to use this system for the restoration project.

[6] Entitled 'Incidents from the History of Geometry'.

> Now that the first world war was over, the matter was considered again. Professor Engel already had most of the manuscript ready. He had worked out most of the annotations. There were only a couple of the geometric papers, especially the first one, not easily accessible, that he had not considered. The Norwegian Research Foundation promised to give the necessary, considerable, financial support, but then a Norwegian had to join as a coeditor. In a way I was not particularly keen on the idea. The work would take me a long time since I had not earlier occupied myself with the work of Sophus Lie, and I would thereby be hindered in my plans to resume the investigations in my dissertation. But in this way the appearance of the edition would depend upon my consent, and as a professor of geometry, the very subject in which Sophus Lie had worked, I felt morally obligated to support the publication. I visited several times with Professor Engel in Giessen ..."

Heegaard does not indicate when the above took place, but according to the obituary, [25], the editing extended over the period 1921–1937. The work did turn out to be considerable. In the two volumes to which Heegaard contributed one finds 47 of Lie's publications totalling more than 1300 pages, and with detailed annotations running to more than 400 pages (not counting indexes and the like). As mentioned earlier, Heegaard had to give up before the finishing line. In the introduction to volume I, the editors note that

> "Since Engel can in no way pass himself off as a geometer – nor would he be accepted as such by Lie himself – the other one of us, Heegaard, constituted the necessary completion in order to edit the geometric articles. Unfortunately, his teaching and his constantly appearing new duties have placed such demands on him that Engel has had to do the annotations alone from article number XI."

Heegaard did sign the preface to volume II which has no special mentioning of the parts played by the respective authors, so it seems natural to assume that here the original plan had been restored so that Heegaard had the main responsibility for the annotation in this part.

Naturally this is not the place to go into a description of Lie's work or Heegaard's annotation thereof. However, one feature must be included. As mentioned earlier, in his dissertation Heegaard describes the points in 4-dimensional space as points in ordinary 3-space equipped with contour numbers. Presumably, it had not been known to him at that time that some of Lie's very first articles take precisely the same view point. At any rate, here some 25 years later, he takes a great interest in this early part of Lie's work: The annotations to these papers run to almost twice the length of the papers themselves.

Moreover, three papers published by Heegaard in the period 1928–1930 seem to be direct spinoffs. The first one, [8] was his contribution to the International Congress in Bologna in 1928, where the explicitly stated purpose is to draw attention to Lie's work. One year later, at the seventh Scandinavian Congress in Oslo, he presents a generalization of the viewpoint to three complex dimensions, [9]. This generalization is also the topic of [10].

Shortly before his retirement, Heegaard finally finds the time to return to his dissertation. This leads to three papers in the period 1938–1941. As shown by the following review of [14] in Zentralblatt für Mathematik, Lie's view of four space again figures prominently.

> "It is the purpose here to utilize Lie's presentation of points in the complex plane as points in space equipped with contour numbers in order to approach a visual investigation of the Riemann surface of an algebraic function. The method is explained in some detail, albeit not completely, through the example of the sphere $x^2 + y^2 + z^2 = r^2$. H. Kneser"

The remaining two papers, [15, 16], use similar methods to study neighbourhoods of the origin in two specific complex surfaces.

13. The four colour problem

Heegaard's interest in graph theory and the four colour problem first shows up at the 3-rd Scandinavian Congress in Kristiania in 1913, [4], and he takes up the thread again in 1933, [11], and at the 1935 International Topology Conference in Moscow, [13]. His main contribution is a reduction of Heawood's congruences to a single congruence. But this section is not here to report on what Heegaard *did* in graph theory, but rather to let himself describe what he did not (manage to) do. The description appears in three letters[7] that he wrote to Jakob Nielsen in the Spring of 1946. On May 9, he writes:

> "Hold on to your hat! I believe I have solved the four colour problem!! At least I cannot find any mistake. But considering the many false proofs that previously have been given I would be exceedingly grateful if you would look through the small paper enclosed. If you can get other Danish mathematicians, e.g., Hjelmslev and Steffensen to look it over, I would also be glad. Since I would like to publish the paper in America, I shall send it to Veblen. I am not well up in the writing of English so I shall ask Veblen to have a student revise the language, at my costs. ..."

Jakob Nielsen seems to have answered quickly, because May 28 Heegaard continues:

> "I am sorry that I wrote you at a time when you have so very much to do. Thus, I'm even more grateful for your inspection. It shows me that at least there is no obvious mistake. I was ill at the time that I sent my original proof[8] to Künneth; otherwise I would probably have discovered the flaw myself. Honestly, on that occasion I was very annoyed with myself for having overlooked the [special] case. And since Künneth apparently had worked hard, but in vain, to repair the matter, I gave up working on it. But then some weeks ago, three days before I had to be admitted to the hospital for an operation, the solution literally fell down from Heaven, and the matter was in order in a couple of minutes. Myself, I now do not believe that there could be any real mistake. I do realize that formally a lot of objections can be raised. But I had to go into surgery and I had no guarantee that everything would go so well. ..."

The third letter is not dated, probably because of the agitation that may be construed from the heading:

> "Professor Poul Heegaard.
> Nordli, Aurdal, Valdres, Norway, Earth, pro tem The Universe.
>
> Dear Jakob Nielsen.
> This is only to tell you that you may throw the manuscript in the wastepaper basket. Indeed, there is a mistake in Section 17 (Possibility 2). It is incredible how easily a mistake may sneak into this complicated matter. In 1942, I believed I had constructed a vortex graph transformation which would solve the problem but a German expert pointed out a case where it did not always exist. Now after a life threatening case of pneumonia I had to go in for a hernia operation, and it occurred to me that both Richard Birkeland and Palmstrøm had passed away

[7] We thank Lektor Bjarne Toft, IMADA, Odense University, for drawing our attention to these letters in the Jakob Nielsen collection at Matematisk Institut, University of Copenhagen.

[8] Probably the one also referred to in the third letter, see below.

after an operation. So I got the idea that I would try over the following three days to mend the flaw in the proof. I had a list of all possible initial vortex graphs. And I noticed that with one of them the criminal case disappeared. As I now hastily changed the whole exposition in accordance with this, I did not notice that the conditions for a different case had been changed. Thereby the four colour devil found a new loophole. . . ."

14. Heegaard's spiritual life

When we first set out to investigate Heegaard's life we were struck by the scarcity of obituaries. Heegaard had been an editor of three journals published in Scandinavia, *Acta Mathematica*, *Nyt Tidsskrift for Matematik* and *Norsk Matematisk Tidsskrift*, but none of them published any obituary. Actually, apart from two small pieces in Danish newspapers, the only obituary known to us is [25], which has the following on Heegaard's spiritual life:

". . . he was a high ranking free mason, and for many years he was active in the Oxford Movement.[9] Most of all it was all kinds of philosophical, political and religious systems that attracted his attention and kept him captive. Clearly he was a searching soul in pursuit of a meaning in life. He had to try everything. But he abhorred doubts, and his desire to believe occasionally made him almost blind to the less sympathetic aspects of the systems and their followers. Towards the end, his inquiring mind apparently found peace in Catholicism."

These words were written three years after Norway's liberation from five years of Nazi rule. And the extensive Heegaard bibliography in [25] suggests that indeed Nazism was one of the systems, to whose imperfections Heegaard had been blind. In fact, in 1945 Heegaard published a two page note 'Meine Ahnentafel. Auswärtige Vorfahren' ('My pedigree. Foreign ancestors') in a periodical 'Norsk-tysk Tidsskrift' ('Norwegian-German Journal') that appeared in Oslo precisely during the years of occupation. In it, Heegaard advocates genealogy as an interesting means to stimulate international contacts and he presents a few examples of his own studies in the area. Thus, the actual content of the note is completely innocuous, but in Norway, at that time, the very existence of such a note might be enough to ostrasize its author.

Lest the reader believe that this is a case of overinterpretation on our part, we quote from the centennial volume of the Norwegian Science Academy, [20]

"Poul Heegaard whose sympathies definitely lay with the authoritarian states stopped attending the meetings after 1941."

The Norwegian society of the 1940's may have seen another indication of such sympathies in a series of causeries, "Naturvitenskapens hövdinger" ("Chiefs of Natural Science"), presented by Heegaard in the Norwegian radio, NRK, in 1944 and 1945. One of Poul Heegaard's grandchildren, Rese Hjelle, gave us access to the manuscripts of these causeries, and once again, we note a complete absence of politics. However, to appreciate the political impact that may nevertheless have resulted from such an activity, one only has to know that during most of the occupation, Norwegian families were allowed to own a radio set only if a majority in the household were members in good standing of the Norwegian Nazi party. All others had their sets confiscated in 1941, and this lowered the number of registered listeners from more than half a million to less than fifteen thousand [22].

[9] In [17], Heegaard himself speaks of the Oxford Group, aka Moral Re-Armament, rather than the Oxford Movement.

After the war, Poul Heegaard rewrote the causeries slightly (in Danish), to make them form the major part of a book manuscript, [18], that he apparently intended to have published in Copenhagen. We borrowed this manuscript from the above mentioned Rese Hjelle. A transcript is available on the Internet at the address `http://www.imada.ou.dk/~hjm/heegaard.html`.

In spite of the above, the Academy did accept the presentation of the obituary [25] at its meeting November 1, 1948, so it must have been felt that Heegaard's connection to the Nazi ideology had been relatively innocent. One could speculate that the unpleasant sympathies simply sprang from Heegaard's long standing love and respect for German mathematics. Certainly this background was noticeable after the *First* World War when Heegaard [7] tried to counteract an international attempt to exclude German mathematicians from future international cooperation, but then, so did many other mathematicians who would later take a clear stand against the Nazi movement.

Naturally, we would have liked to see Heegaard's own account of this period in his life, but the autobiographical notes finish abruptly

> "... I attended many meetings [in the Oxford Group] but gradually became more and more skeptical as I noticed how superficial many people took the matter. And after April 9, 1940[10] I retired from the movement because"

Once again we wonder whether some pages have been deliberately destroyed.

An interesting piece of information was presented to us by Poul Heegaard's great grandson, Bror Magnus Heegaard. He reports that, according to his father, one of the leaders of the Norwegian resistance movement has praised Poul Heegaard for inspiring the formation of the resistance leadership, Milorg, and has dedicated a copy of his book about Milorg to Poul Heegaard. We have not (yet?) seen a copy of such a dedication, but if, indeed, it exists, it would certainly reinforce our image of Poul Heegaard as an innocent, albeit politically naive, admirer of German science and Germany rather than an adherent of a fascist movement.

On the religious side, we have to return to Poul Heegaard's father, Sophus Heegaard. Throughout most of his life, Sophus was an outspoken agnostic, but three years before his death, following his own serious illness and the death of Poul's older sister, Henny, he became

> "... a practicing Christian. In a new edition of his book on upbringing and education, he changed the preface in a Christian direction declaring that previously he had relied on science, but when the storms of life came these anchor cables snapped like threads and he found peace only in a simple Christian faith. The new preface caused a great stir and rejoinders. It was claimed that his conversion was caused solely by illness and feebleness of the brain."

This happened when Poul was around 10 years old and made a lasting impression which came to a climax in 1916 when Poul's 11 year old son, Aage, was admitted into a Copenhagen hospital with meningitis.

> "... from the corridor I heard him say the Lord's Prayer. When I came into the room he was unconscious and he never regained consciousness. This affected me violently. I had to find my school catechism to learn the Lord's Prayer. A few days later there was a call to Magdalene and me telling that the end was approaching. When we entered the

[10] The day of the German invasion of Norway.

> sick room a nurse was standing with a telephone at the ear connected to a stethoscope. Something like a grey shadow raced over his face and the nurse said: 'Now his heart stopped beating'. Silently I folded my hands and said the Lord's Prayer. This was the end to a long fight. I had had to give in to God."

Heegaard's faith manifested itself also in his academic activities. Over the years a number of small articles on science appeared in various religious periodicals. And at the University of Kristiania he took a special interest in Christian students, e.g., by letting them use his office for morning prayers.

15. Conclusion

We have described some aspects of Heegaard's life and career largely by following his own account. This was written at a late stage of his life, and in an earlier version of this article we speculated that the turbulence about his position during the Second World War might have made the author a bitter and depressed man. However, except for short remarks in a couple of letters, we had no real evidence. Also, an interview with the granddaughter Rese Hjelle, who spent most of her childhood and youth in Poul Heegaard's house, contradicted our speculations. In fact, Rese Hjelle recalls her grandfather "Bobo" as a very happy and entertaining old man who would spend a lot of time at his desk, reorganizing papers.

Which conclusions can we draw from our study? First of all, it is undeniable that Heegaard's mathematical research production was small. Probably no Dean of today would even grant tenure on the basis of such a publication list. On the other hand, Heegaard's criticism of Poincaré did play a vital role for the foundation of algebraic topology, and the study of 3-manifolds is still very much based on notions that he created. Thus a search for his name in current review journals might lead our hypothetical present day Dean to reconsider. Indeed, over the last ten years Mathematical Reviews lists more than 200 reviews where the word Heegaard occurs either in the title or in the review itself.

It is also undeniable that Heegaard's publications in astronomy are more numerous than the mathematical ones. However, we do not believe that he would necessarily have been a happier man, or more productive in the strict academic sense, if he had chosen a career in astronomy. Maybe, deep down he was really an expositor and an educator more than a researcher. After all, he *does* characterize the teaching years between his dissertation and his appointment at University of Copenhagen as a happy time; it *does* appear that he would always let other activities interfere with his research; and it *is* remarkable that his autobiographical notes record *only one* instance where mathematical research is described as an urgent endeavour, viz. in his description of the Christmas vacation where he worked out his counterexample to Poincaré's formulation of the duality theorem.

Acknowledgements

From the Heegaard family we thank *Poul E. Heegaard*, whose importance for this report has been described in the opening section; *Lars and Rese Hjelle* with whom we had a fruitful discussion and from whom we borrowed a variety of material including pictures and the book manuscript [18]; *Jørgen Anker Heegaard*, who – in addition to pictures – gave us much information about the Heegaard family in general, and Poul Heegaard's father, Sophus, in particular; and *Bror Magnus Heegaard*, who told us about his great grandfather's (possible) role as an inspiration for the Norwegian resistance forces.

Also, we want to acknowledge the contributions of a number of friends and colleagues: Lektor *Kirsti Andersen* of Aarhus University, Docent *Jesper Lützen* of University of Copenhagen and Adjunkt *Kurt Ramskov* of University of Copenhagen for useful information in general; Lektor *Bjarne Toft* of IMADA, Odense University, for directing us to Heegaard's letters to Jakob Nielsen; Librarian at Odense Matematiske Bibliotek, *Margit Christiansen*, for assistance in locating many publications; Secretary at IMADA, Odense University, *Lisbeth Larsen*, for carefully transcribing [17, 18]; and last, but not least, Professor *Ioan James* for giving us a very interesting assignment.

Bibliography

The following list of references is organized with all Heegaard works first. We have chosen not to include any of his purely didactical publications. *In [25] there is a much more extensive Heegaard bibliography.*

References to Heegaard's work

[1] P. Heegaard, *Forstudier til en topologisk Teori for de algebraiske Fladers Sammenhæng*, Dissertation, Copenhagen (1898); see also [5].

[2] M. Dehn and P. Heegaard, *Analysis situs*, Enzyklopädie Math. Wiss., Vol. III.1.1, 153–220.

[3] P. Heegaard, *Der Mathematikunterricht in Dänemark*, Gyldendalske Bogh., Copenhagen (1912).

[4] P. Heegaard, *Bidrag til Grafernes teori*, Beretning om den 3. Skandinaviske Matematikerkongress, Kristiania (1913), A.W. Brøggers Boktrykkeri A/S, Kristiania (1915).

[5] P. Heegaard, *Sur l'Analysis Situs*, Bull. Soc. Math. France **44** (1916) 161–242. Translation of [1].

[6] P. Heegaard, *The Childhood of Astronomy*, Lectures for Prisoners of War, University of Copenhagen and the Danish Red Cross, 16 pp. + inserted plates (Translated into French, German, Italian and Russian).

[7] P. Heegaard, *Faut'il reprendre les relations intellectuelles avec savants allemands*, La Paix par le Droit **31** (1921), 226–227.

[8] P. Heegaard, *Repreésentations des points imaginaire de Sophus Lie et sa valeur didactique*, Atti del Congresso Intern. dei Mat. Bologna (3.–10. Settembr. 1928), Bologna (1930).

[9] P. Heegaard, *Om en generalisasjon til rummet af Sophus Lies fremstilling av imaginære elementer i planet*, Den syvende Skandinaviske Matematikerkongress i Oslo (19–22 Aug. 1929), A.W. Brøggers Boktrykkeri A/S, Oslo (1930).

[10] P. Heegaard, *Eine raümliche Erweiterung der Lie'schen Darstellung von den imaginären Elementen der Ebene*, Opuscula Math. A. Wiman Dedicata, 11/2 (1930), Lund, 157–165.

[11] P. Heegaard, *Über die Heawoodschen Kongruenzen*, Norsk Mat. Forenings Skrifter, Ser. II (1933), 47–54.

[12] *Sophus Lie: Gesammelte Abhandlungen*, Herausgegeben von Friedrich Engel und Poul Heegaard, Vol. 1 (1934), Vol. 2 (1935), Teubner, Leipzig.

[13] P. Heegaard, *Bemerkungen zum Vierfarbenproblem*, Mat. Sb. **43** (1936), 685–694.

[14] P. Heegaard, *Die Topologie und die Theorie der algebraischen Funktionen mit zwei komplexen Variablen*, Neuvième Congrès des Mathematiciens Scandinaves, Helsingfors (1938), Mercators Tryckeri, Helsingfors (1939), 15–22.

[15] P. Heegaard, *Die Umgebung von Origo im Imaginären auf der Fläche* $y^2 + z^2 = 2z$, Avhandlinger utg. av Det Norske Videnskabsakademi i Oslo I, Mat.-Nat. Klasse **1** (1939).

[16] P. Heegaard, *Beiträge zur Topologie der algebraischen Flächen. Die Umgebung von Origo im Imaginären auf der Fläche* $y = x^3 + z^3 - 3xz$, Avhandlinger utg. av Det Norske Videnskabsakademi i Oslo I, Mat.-Nat. Klasse **5** (1941).

[17] P. Heegaard, *Små livserindringer fortalt for mine barn og barnebarn*, Handwritten Notes in Norwegian, dated Fåvang, February 2, 1945, with several lacunae and inserts, pp. 1–136. A transcript (in Danish) is available on the Internet: `http://www.imada.ou.dk/~hjm/heegaard.html`.

[18] P. Heegaard, *Naturforskningens Stormænd indtil det attende Aarhundrede, Almenfatteligt fremstillet af Professor Poul Heegaard*, København (1947), Book manuscript in Danish. A transcript is available from the above Internet address.

Other references

[19] *Aarbog for Københavns Universitet*, Kommunitetet og Den Polytekniske Læreanstalt, Københavns Universitet (1917).
[20] L. Amundsen, *Det Norske Videnskabs-Akademi i Oslo, 1857–1957*, Aschehoug og Co., Oslo (1960).
[21] W. Boy, *Über die Curvatura Integra und die Topologie geschlossener Flächen*, Dissertation, Göttingen (1901); and Math. Ann. **57** (1903), 151–184.
[22] H.F. Dahl, *Dette er London. NRK i Krig 1940–1945*, J.W. Cappelens Forlag, A-S (1978).
[23] *Dansk Biografisk Leksikon*, 3die udgave, Sjette bind, Gyldendals forlag, København, Danmark (1980).
[24] J. Dieudonné, *A History of Algebraic and Differential Topology 1900–1960*, Birkhäuser, Boston/Basel (1989).
[25] I. Johansson, *Minnetale over Poul Heegaard*, Det norske Videnskabsakademi i Oslo, Årbok (1948), 38–47. (This contains an extensive bibliography for Poul Heegaard.)
[26] C. Jordan, *Des contours tracés sur les surfaces*, J. Math. Pure Appl. (2) **11** (1866), 110–130.
[27] F. Klein, *Elementarmathematik vom höheren Standpunkte aus, II*, Verlag von Julius Springer, Berlin (1925).
[28] J. Petersen, *Forelæsninger over Funktionsteori*, Carl Schønberg, Copenhagen (1895).
[29] F. Picard and G. Simart, *Théorie des Fonctions Algébriques de Deux Variables Indépendentes*, Tome I, Gauthier-Villars, Paris (1897).
[30] H. Poincaré, *Analysis Situs*, J. École Polytech. (2) **1** (1895).
[31] J-C. Pont, *La Topologie Algébrique des Origines à Poincaré*, Presses Universitaires de France (1974).
[32] J. Przytycki, *Knot Theory from Vandermonde to Jones (with the translation of the topological part of Poul Heegaard Dissertation by A.H. Przybyszewska)*, Preprint no. 43 (1993), IMADA, Odense Universitet, Denmark.
[33] K. Ramskov, *Matematikeren Harald Bohr*, Institut for de Eksakte Videnskabers Historie, Aarhus Universitet, Aarhus (1995).
[34] J. Stillwell, *Classical Topology and Combinatorial Group Theory*, Springer, Berlin (1980).

CHAPTER 35

Luitzen Egbertus Jan Brouwer

27.2.1881 Overschie – 2.12.1966 Blaricum*

Dirk van Dalen

Department of Philosophy, Utrecht University, Heidelberglaan 8, PO Box 80.126, 3508 TC, Utrecht, The Netherlands

E-mail: Dirk.vanDalen@phil.uu.nl

When Brouwer entered the mathematical scene, the Netherlands had already produced some outstanding mathematicians, notably Christiaan Huygens, Simon Stevin and Thomas Jan Stieltjes. Only the latter one belonged to the nineteenth century, but he did not practise his mathematics in the Netherlands. In fact, the level of mathematics had in the eighteenth and nineteenth century remained far behind that of the surrounding countries; the same could be said for the exact sciences, but those disciplines had at the end of the nineteenth century already joined the international research community. Van der Waals, Kamerlingh Onnes, Lorentz, van 't Hoff, Hugo de Vries and others belonged to the top of their profession.

Brouwer was born on 27 February 1881 as the son of a schoolmaster in Overschie (now a part of Rotterdam); after half a year his father accepted a position in Medemblik where the family spent the next 11 years before moving to Haarlem, where his father became the headmaster of a secondary school. In Medemblik two more sons were born, Lex and Aldert.

The high school years of Brouwer were without problems. He entered high school (the Hogere Burger School, HBS), in 1890 at the age of nine – a record at the time.

In Haarlem he again visited the HBS, in 1894 he entered the gymnasium (the successor of the old Latin school) while at the same time preparing for the final examinations of the HBS. In 1895 he obtained the HBS diploma and two years later (compressing three school years into two) he graduated from the gymnasium α and β. All the time he was the number one student in class, his only weak spot was the art class.

The gymnasium diploma gave him the right to study at a university. The university of his choice was the Municipal University of Amsterdam, where physics was dominated by Van der Waals. In mathematics there were two professors: Diederik Johannes Korteweg

*The paper relies on a number of historical presentations: Freudenthal's account in [52], the biography of Brouwer, [60], and [2–4, 64, 66, 67, 75].

HISTORY OF TOPOLOGY
Edited by I.M. James

Professor L.E.J. Brouwer – 1912 (Courtesy of Brouwer Archive) and A.J. Van Pesch. The latter is not known for striking contributions, but Korteweg was an extremely competent applied mathematician. He was the first doctor of the young Amsterdam University, where he obtained the doctor's degree from Van der Waals. Three years after his doctorate he became a professor at the Amsterdam University. Nowadays his fame is mostly based on the Korteweg–de Vries equation, but in fact he contributed to the most diverse parts of mathematics, ranging from thermodynamics to philosophy of mathematics. The mathematical underpinning of Van der Waals' physical theories was for a large part the work of Korteweg (e.g., the investigation of the Van der Waals surface, folding of surfaces). Furthermore the edition of Huygens' collected works was largely the work of Korteweg.

Brouwer learned his mathematics mainly from Korteweg; the second big influence in his early years was Gerrit Mannoury, a more or less self-made mathematician.

Brouwer was not a terribly quick student, it took him three and a half years to pass the *candidaatsexamen*, the half-way examination, which could be taken after two years. On 16 June 1904 he passed his final examination, the *doctoraal examen*. In all it took him seven years to obtain the doctoral degree (comparable to the M.Sc.). Certainly nothing to be proud of for such a brilliant student – for a brilliant student he was indeed: he passed both examinations *Cum Laude*. It should be pointed out that he had managed to publish three research papers before his final examination: [5–7]. The papers dealt with the decomposition of rotations in four-dimensional Euclidean space. In modern terminology Brouwer showed that $SO_4 \cong SU_2 \times SU_2/\pm(1,1)$. His proof was purely geometric; in the third paper he added an algebraic derivation. The paper brought him some recognition, but also his first priority conflict. The Berlin professor E. Jahnke claimed to have proved the main result first. Brouwer, who had not been aware of Jahnke's paper, analysed Jahnke's result and showed that apart from a certain similarity in formulation, there was no ground for Jahnke's claim.

The main reason for the long drawn out study was Brouwer's poor health, in combination with the national service in the army. His student years, and the next few years, were marked by nervous collapses and general ill health.

The story of Brouwer's student years can be read in the correspondence of the young student and the poet Carel Adama van Scheltema, cf. [57].

The second half of his study was also interrupted by the military service. Although Brouwer was no stranger to physical exertions – he was fond of walking, football, swimming, and in 1899 he made a trip to Italy on foot, he could not, however, adept himself to the military milieu. Pestered by his fellow conscripts and mocked by his superiors, he had his worst breakdowns in years.

Brouwer's interest was divided between mathematics and philosophy, the latter was rather a hobby and not a formal study. The true love of mathematics was strongly encouraged by Mannoury, a Jack of many trades. Gerrit Mannoury, the son of a merchant navy captain, had come to mathematics via the teacher's career. He had finished high school in 1885 and obtained his teacher's diploma three months later. He taught at many schools until he became a full professor in 1918. Although he never got a university degree, he published a few significant mathematical papers in the years before Brouwer's mathematical activity. In 1898 he published a paper "*Les lois cyclomatiques*" [73], the first Dutch paper in the new area of topology. The paper treats a generalised form of the Euler–Poincaré formula. Mannoury proved in this paper a theorem which Van Dantzig [61] has called 'Mannoury's duality theorem'. In Hopf's words

> The theorem expressed by the [indicated] formulas, which you correctly call 'Mannoury's duality theorem', belongs completely to the area of modern duality theorems, and that Mannoury knew it in 1897 shows how far he was ahead of his time. It is a pity indeed, that he did not continue this work, he was very close to the duality theorems of Alexander [61, p. 7]. .

Mannoury was also the first person to introduce Peano's symbolic logic in the Netherlands. In spite of the efforts of Korteweg, who gave Mannoury private tutorials on Sundays, and who allowed Mannoury the use of his private library (there was as yet no mathematical library at the University of Amsterdam), Mannoury could not find the time to study for the formal degree. But his mathematical talent was well recognized, so that in 1903 the University of Amsterdam admitted him as a *privaat docent,*[1] Mannoury's original and playful approach to mathematics greatly influenced Brouwer. The lectures of Mannoury eventually appeared in print [74].

In the Dutch academic tradition one studied at a university, obtained a doctoral diploma, with the title of *doctorandus*, and then either chose a profession or continued to work for a Ph.D. degree. In mathematics the profession invariably meant 'teaching', but even a doctorate was no guarantee for a scientific position. The universities had few mathematics professors, and hardly any lower positions. If one was lucky, some professor would die or retire and the faculty would offer the vacancy to the mathematical doctor. In the meantime one taught at a gymnasium or a HBS, published dutifully and hoped for luck.

Brouwer was temperamentally ill-suited for the teaching profession, thus his choice for a continued study seemed logical. Even after finishing his university education he was not certain whether to become a mathematician or a philosopher. Mathematics came out on top, but not before a brief excursion into the domain of philosophy.

In 1905 Brouwer gave a series of lectures on what he called 'moral philosophy', published under the title "*Life, Art and Mysticism*", cf. [80]. In these lectures Brouwer expounded a mystical view of the world. A number of topics are of interest for his later scientific and foundational work, e.g. his negative view on the role of language and the conviction that the domination of nature or fellow creatures is sinful.

Already before his dissertation was written, Brouwer published a few papers on vector analysis: *The force field of the non-Euclidean spaces with negative curvature* [9], *The force field of the non-Euclidean spaces with positive curvature* [10], *Polydimensional vectordistributions* [8]. In the first paper Brouwer used the tools of contemporary differential geometry, but added the novelty of 'parallel displacement', thus being the first to use the notion, albeit in a special case.[2] The third paper contained a proof of the higher-dimensional version of Stokes' theorem. It had apparently escaped him that Poincaré had already formulated the theorem in 1887 and 1899. In a sequel [39] he set the record right by acknowledging that Poincaré had enunciated the theorem, but, he added, "without a proof, however".

The dissertation, which was defended on 19 February 1907, was a rather mixed bag of topics. It appears from the correspondence with the Ph.D. advisor Korteweg that Brouwer originally wanted to incorporate a substantial philosophical part. Korteweg refused his permission for this exposition, he urged Brouwer to stick to the mathematical topics, see [79].

[1] The equivalent of the German *Privatdozent*. A rather poorly paid position at the University, with an opportunity to teach and stay around until an opening would appear somewhere.

[2] Schouten, in his book *Ricci-Kalkül* acknowledged Brouwer's priority, but later authors seem to have forgotten about it.

In a letter to Korteweg Brouwer claimed that the mathematical part of the dissertation was sufficiently substantial; he had, he said, solved three of Hilbert's problems – no. 1, the continuum problem, no. 2, the consistency of arithmetic, and no. 5, the elimination of the differentiability conditions from the theory of Lie groups.[3]

The claim may surprise the modern reader, but one should realize that Brouwer considered the first two problems in a constructive framework, not in an axiomatic or metamathematical framework. His solutions would thus lose their meaning in another setting. The solution to the fifth problem, however, was completely mathematical, albeit only a partial one. In the first chapter Brouwer eliminated the differentiability conditions from the special case of the one-parameter case by means of a meticulous construction.

The dissertation thus contains the first topological investigation of Brouwer. He presented an updated version at the International Mathematics Congress in Rome [14] and subsequently published his results in the Mathematische Annalen [15]. He continued his research in [15, 21], the second paper dealt with the two dimensional case. In a letter to Urysohn (9. 4. 1924) Brouwer mentioned that he had material for another paper, which unfortunately never appeared, nor has a manuscript been found. The material may have been lost in a fire that destroyed Brouwer's house. The Rome Conference may have been the inspiration for one of Brouwer's topological researches in the following years. Poincaré's lecture, the future of mathematics, (which was not delivered by Poincaré himself), contained a section on 'differential equations' in which he had advocated a "qualitative discussion of curves defined by a differential equation". Brouwer treated the subject in a series of papers "On continuous vector distributions on surfaces" [12, 17, 18]. He started with Peano's existence theorem and only made use of continuity properties. Poincaré, on the contrary, had in his earlier publications on the topic exploited algebraic and analytic features. In spite of Poincaré's lecture as a source of inspiration it seems that Brouwer did not know Poincaré's actual publications on the subject (cf. [52, p. 423]). Strange as this may seem, it may be explained as a certain limitation in his mathematical education. Compare his quoting Poincaré on Stokes' theorem only in 1919, and not in 1906.

Among the results of the first paper there is the well-known theorem on the existence of singular points: "A vector varying continuously on a simply connected, two-sided, closed surface must be indeterminate in at least one point".

In the second paper structure theorems for singular points and for the behaviour of the field in the neighbourhood of singular points are proved. The paper contains a purely topological definition of 'winding number', and the notion of homotopic change of vectorfields. The methods of the 'vector distribution'-papers are completely elementary topological.

While preparing his Lie group papers for the Mathematische Annalen, Brouwer discovered that some of the topological results borrowed from the monograph of Schoenflies were not satisfactorily proved, or even false (letter to Hilbert 14. 5. 1909). Schoenflies' book was the second volume of a comprehensive survey of set theory, including point set theory. It was commissioned by the German Mathematical Society, and the choice of Schoenflies was probably based on the fact that since 1899 he had devoted his efforts to the new discipline of set theoretical topology. He was known for his converse of the Jordan curve theorem. The combination is now known as the *Jordan–Schoenflies theorem.* Brouwer had, to his regret, found out that Schoenflies' arguments were far from perfect. He scrutinised Schoenflies' notions and proofs, and sent a report of his investigations to

[3] Brouwer to Korteweg 5.11.1906, see [56].

Hilbert for publication in the *Mathematische Annalen.* After some correspondence, which was not devoid of its painful moments, Brouwer and Schoenflies published in the same issue of the *Mathematische Annalen* a critique of Schoenflies' *Bericht* and a reply.

Brouwer's paper '*Zur Analysis Situs*' contained a number of comments on Schoenflies' topology. The theory of curves had to bear the brunt of the attack. Brouwer gave a number of counter examples, among which the curve that splits a square in three domains has become famous. The curve gave the first example of an indecomposable continuum. Wada gave a nice suggestive version of the construction: Consider an island in a salt water sea, with a sweet water lake. By alternatively digging suitably chosen canals from the sea and from the lake the whole island is 'eaten' away by canals and there is one curve separating the sweet and the salt water. A similar construction with two lakes (say with red and black water) will yield a curve separating three domains.

The "*Zur Analysis Situs*" paper elevated set theoretic topology to a new level of exactness; it became obligatory reading for the new generation of topologists.

At roughly the same time Brouwer started the publication of another series of papers, "*Continuous one–one transformation of surfaces in themselves*" [11, 13, 25, 26, 33]. The starting point was the question "whether a one–one continuous mapping of a sphere into itself is possible without at least one point remaining in its place"?

This, evidently, is the beginning of Brouwer's occupation with fixed points. In the first papers, the methods are still elementary. The whole series contain 8 papers, the last one appearing in 1920.

In [11] the fixed point theorems on the sphere are proved: "A continuous one–one transformation in itself with invariant indicatrix of a singly[4] connected, two sided closed surface possesses at least one invariant point" and the corresponding theorem for simply connected one-sided closed surfaces. The second paper contains a first, and defective, formulation and proof of the *plane translation theorem.* The precise formulation of the theorem reads:

> Let f be an orientation preserving homeomorphism of $\mathbb{R}^2$ without fixed points, then each point p belongs to its domain of translation for f. Here a domain of translation for f is an open connected subset of $\mathbb{R}^2$ with a boundary $L \cup f(L)$, where L is the image of a proper embedding of $\mathbb{R}$ in $\mathbb{R}^2$, such that L separates $f(L)$ and $f^{-1}(L)$.

Brouwer proved it rigorously in his "*Beweis des ebenen Translationssatzes*" [30]. In the fifth paper in the series Brouwer acknowledged that in the first version he had relied on defective arguments from Schoenflies' "*Bericht*". See also [63].

A resumé of the first part of the series appeared in the *Mathematische Annalen* [20].

Another line of research took Brouwer to, what Freudenthal has called "Cantor–Schoenflies (style) topology" – basically point-set topology. He published two papers "*On the structure of perfect sets of points*", in which he proved his extension of the Cantor–Bendixson theorem [19, p. 790], and introduced the first topological group which was not a Lie group.

The second paper contains another generalisation, in addition Brouwer formulated and proved his reducibility theorem ([27, p. 138], cf. [1, p. 123]).

Brouwer gave also a new proof in the spirit of elementary topology of the Jordan curve theorem; the proof stands out as one of the most elegant elementary proofs. It was highly praised by Hilbert.

[4] Brouwer's terminology, present term: 'simply'.

There are a few more Cantor–Schoenflies style topology papers in Brouwer's œuvre: *Some remarks on the coherence type* η [35], which contains higher dimensional generalisations of Cantor's categoricity theorem (all countable dense linearly ordered sets without end points are isomorphic) and finally two papers on G_δ sets.

Brouwer led, so the speak a double life; he was at heart a stern constructivist, but his mathematical leanings were very much geometric, in particular topological. The period 1909–1913 is completely taken up by his topological work. In fact, his appointment as a "*privaat docent*" was primarily intended as a reinforcement of geometry at the Amsterdam University. His inaugural lecture (12.10.1909) had the title "*The nature of geometry*" (Dutch, see [51, pp. 112–120]), it gave a survey of contemporary geometry, including geometric reflections on the theory of relativity. Brouwer concluded that no a priori arguments can serve to single out parts of geometry as privileged. His definition of geometry was strikingly liberal: "Geometry occupies itself with the properties of spaces of one or more dimensions. In particular it classifies the point sets, transformations and groups of transformations, which are possible in those spaces" [16, p. 15].

In the same address he mentioned some open problems, e.g.:

> "In how far are spaces of distinct dimension-number different for our group [of homeomorphisms]". He added "It is very likely that this is always the case, but it seems most difficult to provide a proof, and it will probably remain an unsolved problem for a long time."
>
> ... one has no certainty that the 3-dimensional Cartesian space is split into two domains by a closed Jordan surface, i.e. the one–one continuous image of a sphere.

Brouwer ended his lecture with a plea for basing mathematical theories on analysis situs, the prime example being the topological treatment of geometry (as found in Hilbert's "*Über die Grundlagen der Geometrie*" [65]. See also *Grundlagen der Geometrie, Anhang IV*). In the case of geometry, coordinates can be introduced afterwards by using Van Staudt's techniques.

> "And so", Brouwer concluded, "coordinates will not have to be banned from other theories, if one succeeds in founding them on analysis situs, but the formula-free 'geometric' treatment will be the point of departure, the analytic one will become a dispensable tool.
> It is this possibility and desirability of this priority of the geometric treatment, also in parts of mathematics where it does not yet exist, that I have mainly wished to point out in the above lines".

This is a clear statement of Brouwer's geometric credo, to which he adhered in his own mathematics.

Indeed, Brouwer was occupied with the fundamental problems of topology. At the end of 1909 he made an important breakthrough. He was spending the Christmas vacation with his brother in Paris when he apparently hit on the idea of "mapping degree". In a letter to Hilbert (1.1.1910) he indicated the notion of 'degree' (cf. [52, p. 421]).[5] In this letter the approach to the degree is still algebraic, but it is clear that Brouwer saw the importance and implications. He also formulated generalisation of his fixed point theorems on spheres to higher dimensions and to not necessarily bijective continuous maps.

[5] Freudenthal describes the history of the discovery of the mapping degree and its uses in the topological volume of the Collected Works of Brouwer.

In March 1910 Brouwer wrote to Hilbert that he had a partial solution to the dimension problem: odd and even dimensional spaces are not homeomorphic (letter to Hilbert 18.3.1910).

Somewhere in the spring or early summer Brouwer must have overcome the difficulties and found a satisfactory proof of the invariance of dimension. He submitted a paper of 5 pages to the *Mathematische Annalen* with the title "*Beweis der Invarianz der Dimensionenzahl*" [22].

The paper contains essentially the techniques of the mapping degree and simplicial approximation, the invariance of the mapping degree under homotopic change, simplicial mapping, ... One can see this with hindsight, but at the time it was considered as a clever but inaccessible, complicated proof. When the managing editor Blumenthal (who ran the *Mathematische Annalen* for Hilbert) told Lebesgue about Brouwer's paper during a trip to Paris in the summer vacation, Lebesgue remarked that he had already several proofs of the invariance in his possession. One of them was published as an extract from a letter to the editor in the same issue of the *Mathematische Annalen*, immediately following Brouwer's proof. Brouwer was shocked; he almost immediately saw that Lebesgue had indicated a beautiful principle from which the invariance followed straightforwardly, but he also saw that Lebesgue's proof was totally wrong. The principle that Lebesgue had formulated, without actually proving it, was the famous and elegant *paving principle*.

A long and complicated correspondence between Brouwer, Blumenthal, Lebesgue, Baire and possibly others ensued. When challenged, Lebesgue promised a correct proof, but in spite of Brouwer and Blumenthal's pressure no proof was forthcoming until 1921 [71], and even then, according to Brouwer, it was essentially Brouwer's 1913 proof.

Lebesgue in the meantime had submitted alternative proofs of the invariance of dimension to the *Comptes Rendus* [70].

Brouwer commented in [23]:

> The first one [69] is not sufficient. The second one [70] is with respect the content identical to mine: the differences make the line of thought only more complicated.

In a letter to Baire (5.11.1911) Brouwer wrote that

> I have already proved the Lebesgue *Annalen* theorem a few days after its appearance, but I do not publish the proof because I wish to give Lebesgue the opportunity to do his duty [52, p. 441].

The Lebesgue–Brouwer conflict had not only negative consequences, it spurred both parties to outwit each other. Lebesgue introduced new notions such as the paving principle and linking varieties, and Brouwer pushed his methods to their logical limits. He devised different proofs of basic facts, and laid the basis for the topology of the following years (or decades). Freudenthal, in his comments on Brouwer's papers in the Collected Works, has analysed and described the course of the conflict. Further information can be found in [67] and in the biography of Brouwer, [60].

On October 1910 Brouwer presented his invariance of dimension-proof at a meeting of the Dutch Mathematics Society. The proof is the one of the *Annalen* paper, but for the occasion the presentation was a didactical polished gem. In 1911 a series of papers in the New Topology appeared in rapid succession in the *Annalen*: *Über Abbildungen von Mannigfaltigkeiten* [28], *Beweis der Invarianz des n-dimensionalen Gebiets* [23], *Be-*

weis des Jordanschen Satz für den n-dimensionalen Raum [24], *Über Jordansche Mannigfaltigkeiten* [29].

In particular, the first one of the series is of tremendous importance for the development of topology, it contains all the tools that are implicit in the invariance of dimension paper: simplex star, simplicial manifold (his first definition of manifold occurs in [14]), simplex, indicatrix, simplicial decomposition, simplicial approximation, mapping degree, homotopy for maps, invariance of mapping degree under homotopic change, singularity index.

The paper further contains a generalisation of his singular point of a mapping on a sphere to spheres of even dimension, and it ends with a section on fixed points of continuous mappings of balls, culminating in the famous fixed point theorem.

The second paper contains the proof of the invariance of domain theorem: the homeomorphic image of a domain is a domain. In [31] another proof of the invariance of domain is given, this time based on the mapping degree.

Brouwer considered the invariance of domain paper as much more important than his invariance of dimension paper. Indeed the theorem plays an important role in analysis, and it was the missing piece in the proof of the fundamental theorem of automorphic functions (or uniformisation).

Once having solved the invariance of domain problem, Brouwer looked for a convincing application. He soon found one that had baffled the masters of analysis. After some thought, and probably some consultation, he settled on the theory of automorphic functions and uniformisation.

The uniformisation problem had occupied some of the finest minds of the nineteenth century, including Klein and Poincaré. Klein had indicated a method for solving the problem, the so-called "*continuity method*". This method called, however, for a deep homeomorphism result.

Brouwer had the good forture to realise that the invariance of domain theorem provided exactly the missing link. There is some correspondence with Blumenthal and Poincaré, but there is no doubt that Brouwer found the application himself. He was immediately invited to present his result at the special symposium on automorphic functions which was part of the annual meeting of the German Mathematical Society in Karlsruhe, 27–29 September, 1911.

The successful vindication of the continuity method somehow annoyed the leading automorphic function specialist Paul Koebe who had solved the uniformisation, simultaneously with Poincaré in 1906, by other means. A most unpleasant period followed, in which Koebe feverishly worked to beat Brouwer on his own territory. Many letters were exchanged, Koebe asked for Brouwer's manuscript in exchange for his own one, and then he did not comply. Some wild stories resulted from the episode, including tampering with Brouwer's printer's proofs. At the end Brouwer was sorry to have engaged in uniformisation.[6]

The year 1912 brought more results in the New Topology: the invariance of the closed curve (claimed but not proved by Schoenflies [32]); the introduction of homotopy class (under the name 'class') in "*Continuous one–one transformations of surfaces in themselves, V*" [33], including the theorem that maps of the same degree belonged to the same class. The latter theorem was the topic of Brouwer's talk at the International Congress of Mathematicians in Cambridge 1912 [34].

6 Cf. [52] and [60].

Brouwer was the first to introduce and investigate a number of notions in topology. Most of them have already been mentioned above. He coined a few new terms, but in general he was quite content to use descriptions instead of short, suggestive names. E.g., he referred to homotopy classes just as 'classes', someone else had to introduce the name 'homotopy'. He also used the term '*Zyklosis*', which probably derived from Listing. Brouwer used it for the predecessor of the fundamental group, cf. [32, 42]. Vietoris used the term in the framework of homology theory. The use of "*topological mapping*" in the modern sense was introduced by Brouwer in 1919 [40].

The last paper in the vein of the new methods was his "*Über den natürlichen Dimensionsbegriff*" [36]. In this paper Brouwer gave an intrinsic, topological definition of the notion of dimension. Poincaré had already in his "*Pourquoi l'espace a trois dimensions*" [77] given a first version of such a definition, but it suffered from a number of inadequacies.

Brouwer adopted Poincaré's idea and gave an exact definition. The definition used the notion of separation, and it ran:

> The expression π has the general dimension degree n, in which n denotes an arbitrary natural number, will mean that for each choice of ρ and ρ' [disjoint closed subsets of π] a separating set π_1 exists which has the general dimension degree $n-1$, but not, however, that for each choice of ρ and ρ' a separating set π_1 exists which has a lower dimension degree than $n-1$.

On the basis of this definition Brouwer showed that $\mathbb{R}^n$ has dimension degree n, thereby once more proving the invariance of dimension. The proof used the paving principle, for which Brouwer gave a short proof using the mapping degree.

This particular paper became in the twenties a nail in his coffin. It is the subject of the Brouwer–Menger conflict. Before we pass on to the next episode in Brouwer's life, a comment on his topological methods is in order. Many of Brouwer's contemporaries have remarked that he was hard to read, and up to a point they were right. Brouwer stubbornly stuck to his geometric approach, either unaware of the potential of homology as initiated by Poincaré, or just preferring the direct geometric attack, cf. [62]. Since nobody asked him the question, we can only guess.

It is well established, however, that Brouwer was not a victim of any compulsion to produce results. He usually practiced his mathematics as an artist, free from economic pressure: he loved mathematics for the sake of beauty and the satisfaction it brought, but following up a gold vein he had discovered was not to his taste. After proving the basic facts, he was happy to leave the area to the more ambitious professionals. In that sense we may be grateful to Lebesgue, he may have got more out of Brouwer than any kind well-meaning counsellor!

The year 1913 saw the end of Brouwer's first and incredibly productive, topological period. It is as if Brouwer's topological appetite was stilled.

At the end of this first topological period Brouwer was well recognised both nationally and internationally. The recognition at home had come only slightly later than the international one. When the leading mathematicians were already convinced that this young man had achieved beautiful and difficult results and offered new ideas to mathematics, his Ph.D. adviser was still struggling to get Brouwer a secure place in the academic world. Brouwer became a '*privaat docent*' in 1909, as such he could teach a course here and there, but privaat docents were only minor satellites to the professors and the faculties. Moreover, the salary was more symbolic than real. Korteweg, fully realising that Brouwer could easily drop out of mathematics altogether, started in 1910 an action to get a lecturer's position for him. The board of the university (of Amsterdam) refused, however, to go along. In order to build a better case, Korteweg then started a campaign to get Brouwer into the Royal Academy at Amsterdam as a jumping board for a university job. He wrote to the mathematical authorities of the day: Hilbert, Borel, Poincaré and possibly Klein for recommendations. Only Hilbert's reaction has been preserved; it was unconditionally favourable. The first time, in 1911, the action failed, when Brouwer did not get enough votes, but in 1912 he was elected a member of the Academy. Even then drastic measures where required to get Brouwer a real position. In 1912 Brouwer got an offer from Groningen. Korteweg, who would have been sorry to lose his star student, convinced the board of the University of the necessity to appoint Brouwer. This resulted in an appointment as extrarordinary professor, a post with few duties, but also with a marginal salary. Then Korteweg made a surprisingly generous proposal to the board of the university: he offered to step down himself and pass his chair on to Brouwer. This generous action was crowned with success. Brouwer became a full professor in 1912 and Korteweg stayed on as an 'extraordinary' professor till his retirement in 1918.

The First World War isolated Brouwer more or less from his second scientific home – Göttingen. And so the war years saw a return to his first love: the foundations of mathematical philosophy.

The advances in the foundations were closely related to his teaching. In 1912/13 and again 1915/16, 1916/17 Brouwer taught (among other courses) a course in set theory. From his lecture notes one can more or less reconstruct his progress. The first courses were basically on point set theory, we would nowadays say "theory of real functions". They were conducted in the style of his 1907 constructivism, and nonconstructive parts of the theory were labelled as such.

In 1916/17 he repeated the course of 1915/16, but this time there was an innovation: the introduction of choice sequences. In the margin of his 1915/16 notes he added his new insights on choice sequences. The simplest case of a choice sequence is an infinite sequence of natural numbers determined in a more or less arbitrary way. Given the fact that choice sequences were highly unpredictable objects, Brouwer saw that their weakness was at the same time their strength: if one knew that to every choice sequence a natural number was assigned, then their very undeterminedness forced a continuity property. In mathematical terms, he accepted on the grounds of a conceptual analysis of his choice sequences the following *Continuity Principle*: All functions from the set of choice sequences to the natural numbers are continuous [37].

Furthermore he joined during the war a philosophical society, which later became known as the *Signific Circle*.

In 1915 one of the highest honours that a mathematician could receive from his colleagues befell Brouwer: he was appointed member of the editorial board of the *Mathematische Annalen*.

The importance and impact of Brouwer's topological work was such that the Universities of Göttingen and of Berlin offered him chairs in 1919. Brouwer seriously considered the Berlin offer, but after some generous concessions of the Amsterdam University he decided to stay where he was. The concession took the form of an extension of the mathematics group in the faculty. The later economic crises practically prevented the fulfillment of the promises.

After the war Brouwer resumed his topological activities, but on a more moderate scale, also without new revolutionary insights. His heart was clearly drawn towards the foundations of mathematics. In a series of papers he started to rebuild mathematics along the line of his new intuitionism, i.e. the intuitionism with choice principles [37, 38, 41, 44, 45]. At the same time he published some fifteen papers on the topology of surfaces. A number of enumeration results, and also a paper which extended results of Nielsen on fixed points on the torus [76, 43].

The topological papers of 1921 would probably have been his last activity in the area, had not the developments in dimension theory called him back.

Brouwer had become a full-time intuitionist, and the new program asked all his attention. In 1919 he got the enthusiastic support of Hermann Weyl, whom he told about the new intuitionism during a holiday in the Engadin in Switzerland. Weyl vigorously promoted intuitionism in his provocative paper "*Über die neue Grundlagenkrise in der Mathematik*" [81]. A year later Brouwer gave his first public lecture on the topic at the *Naturforscherversammlung* in Bad Nauheim (September 1920). The challenging title of his talk was "*Does every real number have a decimal expansion?*"

Weyl's paper introduced famous (or notorious) expressions in the mathematical vernacular, such as "Brouwer is the revolution", "Pure existence is paper money". The paper fired the imagination of the readers, and it is not an exaggeration to say that it was the opening volley in the *Grundlagenstreit* (cf. [59]).

In 1923 Brouwer quite unintentionally returned to topology. The event that caused this was the lecture of the young Russian topologist Pavel Urysohn, who was at the annual meeting of the German Mathematics Society in Marburg. Brouwer was at the same meeting, delivering a talk on intuitionism. Urysohn had successfully attacked a number of problems in topology, among other things the definitions and theories of curve and dimension.

Urysohn had come with Alexandrov to visit colleagues, when visiting Göttingen he was told about Brouwer's 1913 paper. In his Marburg lecture he mentioned a mistake in the dimension definition of Brouwer. The "closed" in the definition of separation gave the wrong dimension. Urysohn provided a simple counterexample, cf. [46, p. 637]. Indeed, Brouwer had apparently slipped in the condition 'closed' unintentionally. Freudenthal's precise and convincing analysis shows that Brouwer had immediately after publication seen the, what he called, clerical mistake. He had inserted a remark to the effect in the proofs of Schoenflies' new edition of the *Bericht* [78], but Schoenflies seemed to have ignored the note, probably thinking that it was irrelevant. The question (which became central in the discussion with Menger) is, did Brouwer know the right notion of connectedness (i.e. the modern one)? Lennes gave the modern definition in 1911, and Brouwer himself gave in the same year (probably independently) the same definition [72, 23]. Moreover, Brouwer was the referee for Lennes second paper, so he could not have missed the fact that Lennes and he had given the same (modern) definition.

The evidence for the answer 'yes' seems therefore to be overwhelming (cf. [52, pp. 548 ff.]). The matter is of some importance, as the confusion about the credit for the right dimension notion would have been avoided if Brouwer's correct version had been known to Urysohn – and subsequently to Menger. The Brouwer–Menger conflict would probably have been avoided.

In 1924 Urysohn and Alexandrov again visited western Europe. This time they visited Brouwer, who was most favourably impressed by the two Russians. He was particularly taken with Urysohn, for whom he developed something like the attachment to a lost son. He saw in Urysohn the rightful inheritor of his own topology. After visiting Holland Urysohn and Alexandrov travelled on to France and there they rented a cottage in Brittany. There Urysohn tragically drowned during one of their regular swimming sessions.[7]

Brouwer was broken hearted. He decided to look after the scientific estate of Urysohn, as a tribute to the genius of the deceased. Together with Alexandrov he acquitted himself of this task.

In the following years various topologists visited Brouwer: Alexandrov, Vietoris, Menger and later Freudenthal, Hurewicz, Newman and Wilson. The latter wrote a Ph.D. thesis with Brouwer.

Menger had studied with Hans Hahn in Vienna, where he started his research in topology, as part of Hahn's seminar. He had investigated, starting in 1921, independently of Urysohn (and Brouwer) the notions of curve and dimension.

After some correspondence Menger joined Brouwer as an assistant in March 1925, in May of the same year Alexandrov arrived. In the fall Vietoris came over and Newman dropped in for a short visit. Brouwer's Ph.D. student Wilson also took part in the topology discussions. The members lived in Laren and Blaricum (towns not far from Amsterdam) and the workshops were conducted in Brouwer's house in Blaricum.[8]

During the Christmas Holiday Emmy Noether, with whom Brouwer had been in contact since the Karlsruhe meeting in 1912, stayed with Brouwer. She gave some informal talks, also attended by B.L. van der Waerden , about the definitions of the Betti groups of complexes and related subjects.

Eventually Brouwer and Menger fell out, mostly about the distribution of credits in dimension theory. This developed into a long drawn conflict with numerous letters and recriminations. Even the mediation of Hans Hahn (who was the teacher of Menger and a good friend of Brouwer) was of no avail. In 1927 Menger accepted a chair in Vienna, where he became a key figure in the mathematics community. For detailed information the reader is referred to [67, 75], Freudenthal's commentary in [52], and the forthcoming Brouwer biography, volume 2.

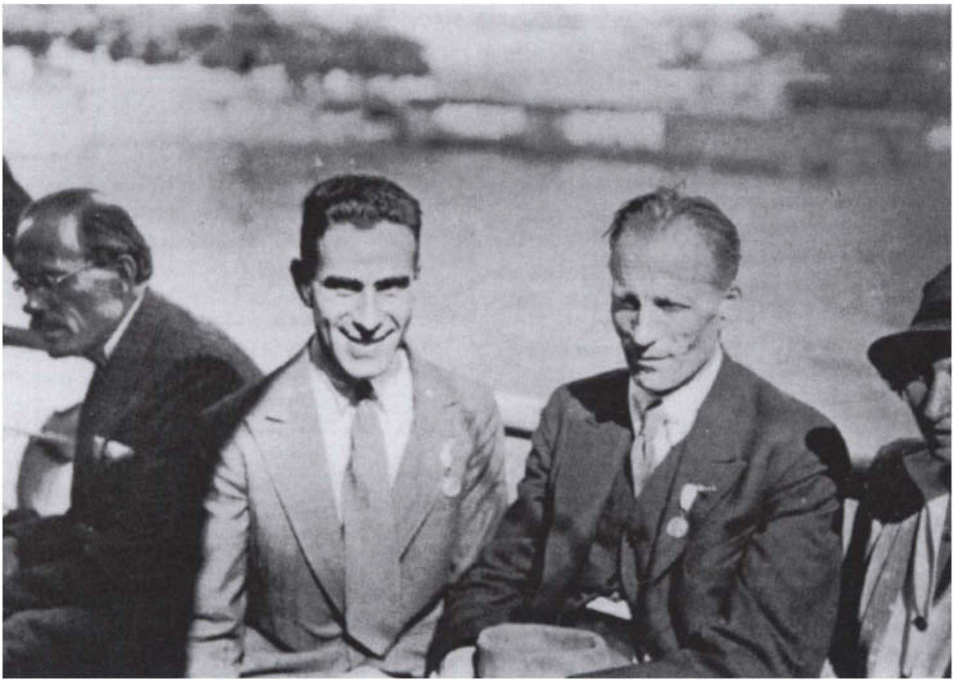
In the meantime Brouwer had carried on his intuitionistic program with considerable success. He had found means to exploit the properties of choice sequences. At the same time Hilbert was developing his proof theory as an answer to the intuitionistic challenge. The foundational discussion assumed definitely nasty proportions. Where Brouwer went out of his way to avoid provocation – his publications bordered on the impersonal, Hilbert attacked his opponents with all available means. After many an exchange the factual *Grundlagenstreit* ended when Hilbert fired Brouwer from the editorial board of the *Mathematische Annalen* (cf. [58]). Brouwer withdrew from the debate and for more than 10 years hardly published anything at all. The intuitionistic activity of the twenties yielded some topological papers, which mainly were intended to show that the adoption of the intuitionistic viewpoint did not lead to such disastrous amputations as some claimed. Among the results are (intuitionistic versions of) the Heine–Borel theorem, the dimension definition and the accompanying 'soundness theorem' (the invariance of dimension), the Jordan theorem.

In reaction to his exit from the Mathematische Annalen, Brouwer founded his own journal, Compositio Mathematicae. Its first issue appeared in 1934. Contrary to the expectation

---

[7] Cf. [3, 4].

[8] Brouwer got most of his visitors paid jobs. Here is a short list of Brouwer's assistants: 1925/26 – Belinfante, Menger, Alexandroff, Vietoris; 1926/27 – Menger, Hurewicz; 1927/28 – Menger, Hurewicz, Gawehn; 1928/29 – Hurewicz, Gawehn, 1929/30 – Hurewicz, Gawehn; 1930/31 – Hurewicz, Gawehn, Freudenthal; 1931/32 – Hurewicz, Freudenthal.



Hurewicz and Brouwer (Courtesy of Brouwer Archive)

of some veterans of the foundational war, the journal turned out to be a respectable, normal mathematics journal, not a pulpit for preaching intuitionism. The journal was closed down in the first year of the occupation, 1940.

The topological activity in Amsterdam was carried on by Freudenthal and Hurewicz, who became Brouwer's assistants. During the period of Brouwer's active mathematical career, further topological research in the Netherlands was carried out by D. van Dantzig, J. de Groot, A. van Heemert, B.L. van der Waerden, and E.R. van Kampen. Van Kampen's work was not influenced by Brouwer (other than through the literature), he came to topology through contacts with Alexandrov and Van der Waerden. The development of homotopy theory was the result of Hurewicz' and Freudenthal's research. Hopf in Zürich took up Brouwer's methods and pushed them far beyond their known limits.

In the late twenties Brouwer created furore with his lectures in Berlin (1927, [53]) and Vienna (1928, [47, 48]). The subsequent *Mathematische Annalen* conflict, unintendedly, put a halt to Brouwer's foundational crusade.

Brouwer was now in his fifties, and the mathematical revolutionary had become a respectable scientist. His recognition in the Netherlands may be inferred from the fact that Queen Wilhelmina made him Knight in the Order of the Netherlands' Lion (1932). International recognition had already come his way, when the *Königliche Gesellschaft der Wissenschaften zu Göttingen* (1917) and the *Leopoldinische–Carolingische Deutsche Akademie* at Halle (1924) adopted him as a member. Furthermore, he received an honorary doctorate from the Oslo University in 1929, together with 16 other mathematicians (including some of friends and associates, such as Hadamard, Landau and Weyl), at the occasion hundredth commemoration of the death of Abel. In the same year the Prussian Academy elected Brouwer as a member.

In the thirties Brouwer mainly worked in private, and hardly published. There is an exception, in [49] Brouwer published the triangulation theorem for differentiable manifolds. Freudenthal subsequently produced another proof. Both were not aware that Cairns had already established the theorem [54, 55], see also [68].

During the Second World War Brouwer published a few intuitionistic papers. After the war he resumed his activity and published a series of papers showing that intuitionistic analysis diverged in a strong specific way from classical analysis.

The postwar novelty, which was already implicit in his Berlin Lectures, eventually became known as the *method of the creating subject.* It provided strong (negative) results instead of the weak results of the form "we cannot affirm at the moment that ..." (the so-called Brouwerian counterexamples). In a paper in the proceedings of the Royal Society [50] Brouwer presented an intuitionistic form of the fixed point theorem (for any $\varepsilon > 0$ there is an x such that $|f(x) - x| < \varepsilon$, for a continuous f).

The aftermath of the war brought Brouwer sad disappointments. He was suspended for a few months on the basis of insignificant grounds and reinstated with a reprimand of the minister. His views on the faculty were no longer heeded and to add insult to injury, the Mathematical Centre was founded in Amsterdam, which virtually got all the facilities, and more, that were promised to Brouwer as a compensation for turning down the Berlin chair in 1920. In addition he was reduced to a symbolic figurehead in his own journal the *Compositio Mathematica.*

Abroad he got the recognition that failed him in post-war Holland. He was elected foreign member of the Royal Society of London, and the University of Cambridge granted him an honorary doctorate. In 1953 he made a lecture tour through the United States and Canada. Furthermore he lectured in Finland, England, France, Belgium and South Africa. He survived his wife Lize de Holl by seven years and died on December 2, 1966, being run over by a car.

Bibliography

[1] P. Alexandroff and H. Hopf, *Topologie I*, Springer, Berlin (1935).

[2] P.S. Alexandrov, *Die Topologie in und um Holland in den Jahren 1920–1930*, Nieuw Arch. Wis. **17** (1969), 109–127.

[3] P.S. Alexandrov, *Pages from an autobiography*, Russian Math. Surveys **34** (1979), 267–302.

[4] P.S. Alexandrov, *Pages from an autobiography*, Russian Math. Surveys **35** (1980), 315–358.

[5] L.E.J. Brouwer, *On a decomposition of a continuous motion about a fixed point O of S_4 into two continuous motions about O of S_3's*, Nederl. Ak. Wetensch. Proc. **6** (1904), 716–735.

[6] L.E.J. Brouwer, *On symmetric transformation of S_4 in connection with S_r and S_l*, Nederl. Ak. Wetensch. Proc. **6** (1904), 785–787.

[7] L.E.J. Brouwer, *Algebraic deduction of the decomposability of the continuous motion about a fixed point of S_4 into those of two S_3's*, Nederl. Ak. Wetensch. Proc. **6** (1904), 832–838.

[8] L.E.J. Brouwer, *Polydimensional vectordistributions*, Nederl. Ak. Wetensch. Proc. **9** (1906), 66–78.

[9] L.E.J. Brouwer, *The force field of the non-Euclidean spaces with negative curvature*, Nederl. Ak. Wetensch. Proc. **9** (1906), 116–133.

[10] L.E.J. Brouwer, *The force field of the non-Euclidean spaces with positive curvature*, Nederl. Ak. Wetensch. Proc. **9** (1906), 250–266.

[11] L.E.J. Brouwer, *Continuous one–one transformations of surfaces in themselves*, Nederl. Ak. Wetensch. Proc. **11** (1909), 788–798.

[12] L.E.J. Brouwer, *On continuous vector distributions on surfaces*, Nederl. Ak. Wetensch. Proc. **11** (1909), 850–858.
[13] L.E.J. Brouwer, *Continuous one–one transformations of surfaces in themselves II*, Nederl. Ak. Wetensch. Proc. **12** (1909), 286–297.
[14] L.E.J. Brouwer, *Die Theorie der endlichen continuierlichen Gruppen unabhängig von den Axiomen von Lie*, *Atti IV Congr. Internat. Mat. Roma*, Vol. 2, 1909, pp. 296–303.
[15] L.E.J. Brouwer, *Die Theorie der endlichen kontinuierlichen Gruppen, unabhängig von den Axiomen von Lie I*, Math. Ann. **67** (1909), 246–267.
[16] L.E.J. Brouwer, *Het wezen der meetkunde* (1909), Openbare Les privaat docent 12.10.1909 (inaugural address).
[17] L.E.J. Brouwer, *On continuous vectordistributions on surfaces II*, Nederl. Ak. Wetensch. Proc. **12** (1910), 716–734.
[18] L.E.J. Brouwer, *On continuous vectordistributions on surfaces III*, Nederl. Ak. Wetensch. Proc. **12** (1910), 171–186.
[19] L.E.J. Brouwer, *On the structure of perfect sets of points*, Nederl. Ak. Wetensch. Proc. **12** (1910), 785–794.
[20] L.E.J. Brouwer, *Über eineindeutige, stetige Transformationen von Flächen in sich*, Math. Ann. **69** (1910), 176–180.
[21] L.E.J. Brouwer, *Die Theorie der endlichen kontinuierlichen Gruppen, unablängig von den Axiomen von Lie*, II, Math. Ann. **69** (1910), 181–203.
[22] L.E.J. Brouwer, *Beweis der Invarianz der Dimensionenzahl*, Math. Ann. **70** (1911), 161–165.
[23] L.E.J. Brouwer, *Beweis der Invarianz des n-dimensionalen Gebiets*, Math. Ann. **71** (1911), 305–313.
[24] L.E.J. Brouwer, *Beweis des Jordanschen Satzes für den n-dimensionalen Raum*, Math. Ann. **71** (1911), 314–319.
[25] L.E.J. Brouwer, *Continuous one–one transformations of surfaces in themselves III*, Nederl. Ak. Wetensch. Proc. **13** (1911), 767–777.
[26] L.E.J. Brouwer, *Continuous one–one transformations of surfaces in themselves IV*, Nederl. Ak. Wetensch. Proc. **14** (1911), 300–310.
[27] L.E.J. Brouwer, *On the structure of perfect sets of points II*, Nederl. Ak. Wetensch. Proc. **14** (1911), 137–147.
[28] L.E.J. Brouwer, *Über Abbildungen von Mannigfaltigkeiten*, Math. Ann. **71** (1911), 97–115.
[29] L.E.J. Brouwer, *Über Jordansche Mannigfaltigkeiten*, Math. Ann. **71** (1911), 320–327.
[30] L.E.J. Brouwer, *Beweis des ebenen Translationssatzes*, Math. Ann. **72** (1912), 37–54.
[31] L.E.J. Brouwer, *Zur Invarianz des n-dimensionalen Gebiets*, Math. Ann. **72** (1912), 55–56.
[32] L.E.J. Brouwer, *Beweis der Invarianz der geschlossenen Kurve*, Math. Ann. **72** (1912), 422–425.
[33] L.E.J. Brouwer, *Continuous one–one transformations of surfaces in themselves V*, Nederl. Ak. Wetensch. Proc. **15** (1912), 352–360.
[34] L.E.J. Brouwer, *Sur la notion de 'classe' de transformations d'une multiplicité*, Proc. V Internat. Congr. Math. Cambridge (1912), Vol. 2, 9–10.
[35] L.E.J. Brouwer, *Some remarks on the coherence type* η, Nederl. Ak. Wetensch. Proc. **15** (1913), 1256–1263.
[36] L.E.J. Brouwer, *Über den natürlichen Dimensionsbegriff*, J. Reine Angew. Math. **142** (1913), 146–152.
[37] L.E.J. Brouwer, *Begründung der Mengenlehre unabhängig vom logischen Satz vom ausgeschlossenen Dritten I. Allgemeine Mengenlehre*, Kon. Ned. Ak. Wet. Verhandelingen **5** (1918), 1–43.
[38] L.E.J. Brouwer, *Begründung der Mengenlehre unabhängig vom logischen Satz vom ausgeschlossenen Dritten II. Theorie der Punktmengen*, Kon. Ned. Ak. Wet. Verhandelingen **7** (1919), 1–33.
[39] L.E.J. Brouwer, *Énumération des surfaces de Riemann regulières de genre un*, Comptes Rendus **168** (1919), 677–678, 832.
[40] L.E.J. Brouwer, *Énumération des groupes finis de transformations topologiques du tore*, Comptes Rendus **168** (1919), 845–848, 1168.
[41] L.E.J. Brouwer, *Intuitionistische Mengenlehre*, Jahresber. Deutsch. Math. Ver. **28** (1920), 203–208.
[42] L.E.J. Brouwer, *Über die periodischen Transformationen der Kugel*, Math. Ann. **80** (1919), 39–41.
[43] L.E.J. Brouwer, *Über die Minimalzahl der Fixpunkte bei den Klassen von eindeutigen stetigen Transformationen der Ringflächen*, Math. Ann. **82** (1920), 94–96.
[44] L.E.J. Brouwer, *Besitzt jede reelle Zahl eine Dezimalbruch-Entwickelung?*, Math. Ann. **83** (1921), 201–210.

[45] L.E.J. Brouwer, *Begründung der Funktionenlehre unabhängig vom logischen Satz vom ausgeschlossenen Dritten I. Stetigkeit, Messbarkeit, Derivierbarkeit*, Kon. Ned. Ak. Wet. Verhandelingen **2** (1923), 1–24.
[46] L.E.J. Brouwer, *Bemerkungen zum natürlichen Dimensionsbegriff*, Nederl. Ak. Wetensch. Proc. **27** (1924), 635–638.
[47] L.E.J. Brouwer, *Mathematik, Wissenschaft und Sprache*, Monatsh. Math.-Phys. **36** (1929), 153–164.
[48] L.E.J. Brouwer, *Die Struktur des Kontinuums*, Sonderabdruck (1930).
[49] L.E.J. Brouwer, *Zum Triangulationsproblem*, Indag. Math. **1** (1939), 248–253.
[50] L.E.J. Brouwer, *An intuitionist correction of the fixed-point theorem on the sphere*, Proc. Roy. Soc. London **213** (1952), 1–2.
[51] L.E.J. Brouwer, *Collected Works I. Philosophy and Foundations of Mathematics*, A. Heyting, ed., North-Holland, Amsterdam (1975).
[52] L.E.J. Brouwer, *Collected Works II. Geometry, Analysis Topology and Mechanics*, H. Freudenthal, ed., North-Holland, Amsterdam (1976).
[53] L.E.J. Brouwer, *Intuitionismus*, D. van Dalen, ed., Bibliographisches Institut, Wissenschaftsverlag, Mannheim (1992).
[54] S.S. Cairns, *On the triangulation of regular loci*, Ann. Math. **35** (1934), 579–587.
[55] S.S. Cairns, *Triangulation of the manifold of class one*, Bull. Amer. Math. Soc. **41** (1935), 549–552.
[56] D. van Dalen (ed.), *L.E.J. Brouwer, Over de Grondslagen van de Wiskunde* (Brouwer's dissertation with correspondence and related papers), Mathematisch Centrum, Amsterdam (1981).
[57] D. van Dalen (ed.), *L.E.J. Brouwer, C.S. Adama van Scheltema. Droeve snaar, vriend van mij*, Arbeiderspers, Amsterdam (1984).
[58] D. van Dalen, *The war of the frogs and the mice, or the crisis of the Mathematische Annalen*, Math. Intelligencer **12** (1990), 17–31.
[59] D. van Dalen, *Hermann Weyl's intuitionistic mathematics*, Bull. Symb. Logic. **1** (1995), 145–169.
[60] D. van Dalen, *Mystic, Geometer, and Intuitionist: The Life of L.E.J. Brouwer*, Vol. 1: *The Dawning Revolution*, Oxford University Press, Oxford (1999).
[61] D. van Dantzig, *Gerrit Mannoury's significance for Mathematics and its Foundations*, Nieuw Arch. Wisk. **5** (1957), 1–18.
[62] J. Dieudonné, *A History of Algebraic Differential Topology, 1900–1960*, Birkhäuser, Basel (1989).
[63] J. Franks, *A new proof of the Brouwer plane translation theorem*, Ergodic Theory Dynamical Systems **12** (1992), 217–226.
[64] H. Freudenthal, *Topologie in den Niederlanden: das erste Halbjahrhundert*, Nieuw Arch. Wisk. **26** (1978), 22–40.
[65] D. Hilbert, *Über die Grundlagen der Geometrie*, Math. Ann. **56** (1902), 381–422.
[66] D.M. Johnson, *The Problem of the Invariance of Dimension in the Growth of Modern Topology I*, Arch. Hist. Exact Sci. **20** (1979), 97–188.
[67] D.M. Johnson, *The Problem of the Invariance of Dimension in the Growth of Modern Topology II*, Arch. Hist. Exact Sci. **25** (1981), 85–267.
[68] N.H. Kuiper, *A short history of triangulation and related matters*, Proc. Bicentennial Congress Wiskundig Genootschap, P.C. Baayen, D. van Dulst and J. Oosterhoff, eds, Vol. 1, Mathematisch centrum, Amsterdam (1979), 61–79.
[69] H. Lebesgue, *Sur la non-applicabilité de deux domaines appartenant respectivement des espaces à n et n + p dimensions (Extrait d'une lettre à M.O.Blumenthal)*, Math. Ann. **70** (1911), 166–168.
[70] H. Lebesgue, *Sur l'invariance du nombre de dimensions d'un espace et sur le théorème de M. Jordan relatif aux variété fermées*, Comptes Rendus **152** (1911), 841–843.
[71] H. Lebesgue, *Sur les correspondences entre les points de deux espaces*, Fund. Math. **2** (1921), 256–285.
[72] N.J. Lennes, *Curves in nonmetrical analysis situs with an application in the calculus of variations*, Amer. J. Math. **33** (1911), 287–326.
[73] G. Mannoury, *Lois cyclomatiques*, Nieuw Arch. Wisk. **2** (1898), 126–152.
[74] G. Mannoury, *Methodologisches und Philosophisches zur Elmentarmathematik*, Visser & Zn., Haarlem (1909).
[75] K. Menger, *Selected Papers in Logic and Foundations, Didactics, Economics*, Reidel, Dordrecht (1979).
[76] J. Nielsen, *Über fixpunktfreie topologische Abbildungen geschlossener Flächen*, Math. Ann. **81** (1920), 94–96.
[77] H. Poincaré, *Pourquoi l'espace a trois dimensions*, Rev. Metaph. Morale **20** (1912), 484–504.

[78] A. Schoenflies, *Entwickelung der Mengenlehre und ihrer Anwendungen I, 1. Hälfte*, Teubner, Leipzig/ Berlin (1913).
[79] W.P. van Stigt, *The rejected parts of Brouwer's dissertation on the Foundations of Mathematics*, Historia Mathematica **6** (1979), 385–404.
[80] W.P. van Stigt, *L.E.J. Brouwer: Life, Art and Mysticism*, Notre Dame J. Formal Logic **37** (1996), 381–429 (Translation of *Leven, Kunst en Mystiek* by L.E.J. Brouwer (1905)).
[81] H. Weyl, *Über die neue Grundlagenkrise der Mathematik*, Math. Zeit. **10** (1921), 39–79.

CHAPTER 36

Max Dehn

John Stillwell

Department of Mathematics, Monash University, Clayton, Victoria 3168, Australia
E-mail: john.stillwell@monash.edu.au

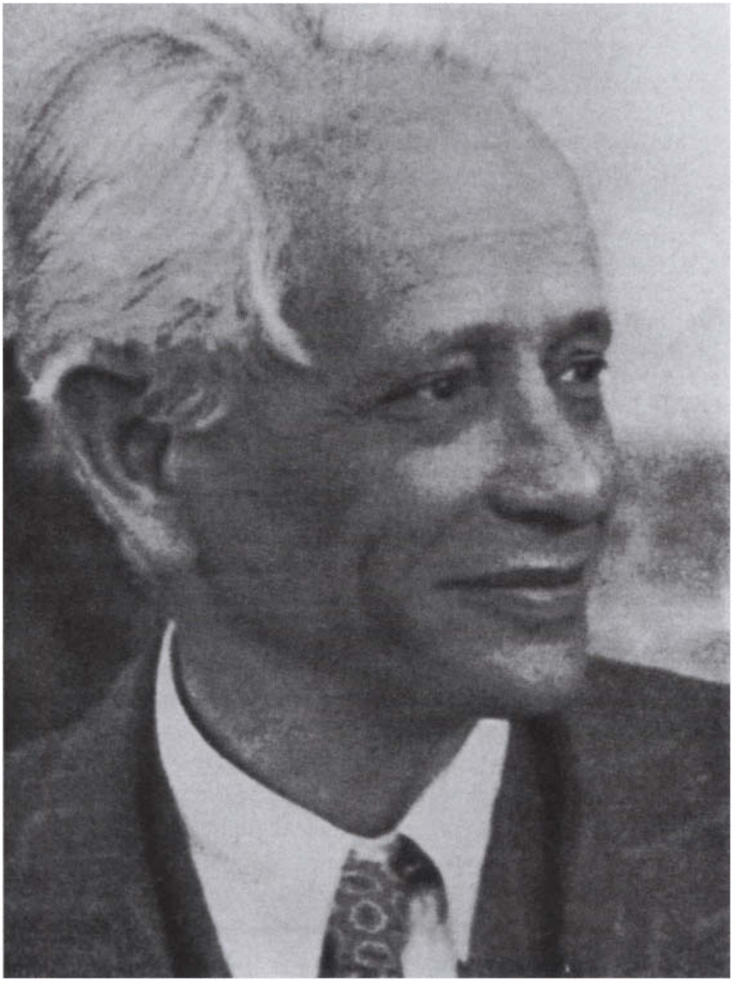
1. Introduction

Max Dehn is remembered today for many concepts in topology and the related fields of geometry and combinatorial group theory: Dehn's lemma, Dehn's algorithm, Dehn surgery, Dehn filling, Dehn twists and the Dehn invariant. Remarkably, most of these concepts were recognised and brought to maturity only after Dehn's death in 1952. One reason for this is that Dehn was often ahead of his time. He worked in topology and combinatorial group theory before they were considered important, so mainstream mathematicians did not at first follow up his ideas. Also, his work was perhaps too visual and intuitive to be respectable, and indeed this approach sometimes led him into error. However, Dehn was also influential in his lifetime through the work of his students Jakob Nielsen, Wilhelm Magnus and Ruth Moufang. He was generous with ideas, and happy for others to publish proofs of results he had discovered.

In this essay I shall try to follow those threads in the history of topology that pass through Dehn: the influence of Hilbert, Poincaré, Heegaard and Tietze on him, and Dehn's influence in turn on other topologists and on the development of topology.

Sources I have found particularly helpful are the biographical articles of Magnus (1978), Magnus and Moufang (1954) and Siegel (1965). I obtained valuable information about Kneser's discovery of the error in Dehn's lemma from Wilhelm Magnus and Martin Kneser, and on Dehn's later career from Linda Hill. Abe Shenitzer and Sanford Segal provided information about Dehn in Frankfurt, Vagn Lundsgaard Hansen and Hans Jørgen Munkholm checked the details of Dehn's life where it intersects with the lives of Nielsen and Heegaard (see also their contributions to this volume), and Walter Neumann made some valuable technical comments. Dehn's daughter, Maria Peters, very graciously shared her family reminiscences with me. Finally, more mathematical details may be found in the book Dehn (1987), which consists of my translations of, and commentary on, the most important Dehn papers.

HISTORY OF TOPOLOGY
Edited by I.M. James



Max Dehn (1878–1952)

2. Early influences

2.1. *Foundations of geometry*

Max Dehn was born in Hamburg on 13 November 1878, the fourth of eight children of a physician. Most of his brothers and sisters lived in Hamburg until the 1930s, so Hamburg remained home base even after Dehn's career took him elsewhere. As his daughter Maria recalled in a letter to me:

> Visits to Hamburg were visits to Heaven for us. The city itself seems to foster a kind of wisdom and self-confidence in its inhabitants; could this be due to the many waterways and the Alster river and lake? Instead of trucks rumbling through the streets, you have barges gliding along canals with their loads. Instead of going to work by streetcar or bus, people take the steamer across the Alster from the residential to the business district, or they walk to work along the water's edge – such a good way to start the day! But for us it was family that made Hamburg our dream town. It was the Verwandtenstadt (family town). Every day had its joyful events, but the special occasions were something else again! Then there would be table songs written and sung, chamber music would be played, quadrilles danced, skits put on to tease, and love and laughter filled the rooms.

His mathematical career began as a student of Hilbert in Göttingen in 1899. This was during the period of Hilbert's interest in the foundations of geometry, and Dehn's first original work was a proof of the Jordan curve theorem for polygons, based on Hilbert's axioms of order and incidence. It was not published at the time, perhaps because Hilbert thought it was easy – it is stated without proof in Hilbert (1899). Dehn's proof was first described in the article of Guggenheimer (1977). In 1900 Dehn completed his doctoral dissertation on the role of the Legendre angle sum theorem in axiom systems of geometry.

Also in 1900, Dehn made his greatest contribution to geometry with the solution of Hilbert's third problem. Specifically, he showed *the regular tetrahedron is not equidecomposable with the cube*, that is, it is impossible to decompose a regular tetrahedron, by planar cuts, and paste the pieces together to form a cube. This result gave a negative answer to a question that had effectively been open since the time of Euclid: can the volume of polyhedra be defined without using infinite constructions? In view of the age of the question, and the fact that Gauss and Hilbert had worked on it without success, Dehn's solution is brilliantly simple.

He associates with each polyhedron Π an invariant of cutting and pasting which we would now write as the tensor $l_1 \otimes \alpha_1 + l_2 \otimes \alpha_2 + \cdots + l_k \otimes \alpha_k$, where the l_i are the edge lengths of Π and the α_i are the corresponding dihedral angles. Elementary linear algebra suffices to reduce the question of equidecomposability to commensurability of the dihedral angles of the tetrahedron and cube, and their incommensurability is proved by some elementary number theory.

Dehn had no thought of topology when he solved Hilbert's third problem, so it is interesting that his invariant has recently been applied to 3-manifolds. A 3-manifold can be defined by identifying faces of a polyhedron, the Dehn invariant of which is a geometric invariant of the manifold. But we now know, thanks to the Mostow rigidity theorem, that geometric invariants are also *topological* invariants of hyperbolic 3-manifolds. Thus for hyperbolic manifolds the Dehn invariant is a topological invariant, as was apparently first noticed by Thurston (see Dupont and Sah (1982, p. 160)). (Technically, it is not quite as simple as this, because the plain Dehn invariant vanishes for 3-manifolds, but one can extract a nontrivial invariant from it.)

In 1900/1901 Dehn was an assistant in the Technische Hochschule at Karlsruhe. His work on Hilbert's third problem became his *Habilitationsschrift*, which earned him a position as *Privatdozent* at Münster. He held this position until 1911, during which time he came under the influence of Heegaard and Poincaré, and his interests shifted to topology and group theory.

2.2. *Foundations of topology*

Algebraic topology was founded by Poincaré in a great series of papers between 1895 and 1904. Poincaré unified and generalised the fragmentary topological ideas of Euler, Riemann, Jordan and Betti into a comprehensive theory of homology and homotopy, including the Betti and torsion numbers of manifolds of arbitrary dimension, and the fundamental group. However, he obtained his results by somewhat ad hoc methods – sometimes assuming differentiability, sometimes a triangulation, sometimes a geometric structure such as a hyperbolic metric. A single foundation for all of Poincaré's results was lacking.

In 1903 or 1904, Dehn and Heegaard met, and found they had a mutual interest in the foundations of topology. According to Johannson's obituary of Heegaard, Dehn met Heegaard at a conference in Kassel in 1903, after which they discussed foundational questions on the train between Göttingen and Hamburg. According to Dehn's widow Toni (told to Wilhelm Magnus) the conference was the International Congress of Mathematicians at Heidelberg in 1904. At any rate, the meeting led to joint work in Kiel during the summer of 1905, and their joint article on analysis situs (combinatorial topology) for the *Enzyklopädie der Mathematischen Wissenschaften*. This article gave a uniform combinatorial foundation for homology and the fundamental group. Among its many results was the first rigorous proof of the classification of compact surfaces, previously discovered by Möbius (for orientable surfaces) and Dyck.

The analysis situs article with Heegaard appeared in 1907. A footnote at the beginning says:

> Heegaard undertook the collection of literature for the article, as well as working out essential portions. Responsibility for the final form of the article is Dehn's.

It appears that the article was originally Heegaard's idea, and Heegaard later felt he received insufficient credit for it (see the biography of Heegaard in this volume). Dehn was well versed in axiomatics from his time with Hilbert, and may well have supplied an axiomatic structure to meet Heegaard's intuitive requirements. However, it should also be said that Dehn himself took an intuitive view of topology, and did not pursue axiomatics thereafter. In fact, Dehn's intuition was his greatest strength, and on occasion it exceeded his ability to give rigorous proofs.

2.3. *Homology spheres*

Dehn was particularly impressed with the Poincaré (1904) paper, which used hyperbolic geometry to study the topology of surfaces. This paper also contains the first construction of a homology sphere, which greatly interested Dehn. Poincaré's construction comes out of the blue, presenting a 3-manifold whose homology is unclear, but which turns out to be trivial after a group-theoretic calculation. The calculation also shows that the fundamental group is *not* trivial, so the manifold is not homeomorphic to the 3-sphere. Dehn sought a more insightful construction, giving a manifold with obviously trivial homology, yet obviously not homeomorphic to the 3-sphere.

His other paper of 1907, a research announcement in the yearly report of the Deutsche Mathematiker-Vereinigung, is an attempt to do this. After correcting the encyclopedia article's account of the Poincaré homology sphere, Dehn gives a new homology sphere construction which is short enough to quote in full:

> A very clear example of such a noteworthy manifold can be constructed as follows: a *knotted* torus T_2 in ordinary space [meaning the 3-sphere] divides the latter into a solid torus T_3 and a part K_3 not homeomorphic to it. Suppose that curves C, respectively, Γ are nonseparating on T_2 and bounding in T_3, respectively, K_3. One joins K_3 to a homeomorphic body K_3' which is bounded by T_2' (with the curves C' and Γ') in such a way that T_2 is identified with T_2', C' with Γ and C with Γ'. In the resulting closed M_3, each curve is bounding when taken once. However, this M_3 is not homeomorphic to ordinary space, since it is divided by the torus T_2 into parts K_3 and K_3', neither of which is homeomorphic to a solid torus.

The claim in the last sentence seems intuitively clear, but in fact it was first proved by Alexander (1924). Perhaps Dehn realised that it would be hard to formalise his intuition, and set off instead on the group-theoretic path followed in his great series of papers from 1910 to 1914. As I pointed out in Stillwell (1979), the construction above can be justified by a group-theoretic argument and Dehn's lemma. The lemma appears in the first paper of the series, Dehn (1910), though in fact *not* in connection with homology spheres, which Dehn does afresh with new constructions which reduce the difficulties to pure group theory. He succeeds in making the homology obviously trivial, but the nontriviality of the fundamental group still requires some work.

2.4. *The Poincaré conjecture*

Poincaré's homology sphere disproves a claim he made at the end of Poincaré (1900): every closed 3-manifold with trivial homology is homeomorphic to the 3-sphere. In the light of the homology sphere counterexample, Poincaré (1904) strengthened the condition to trivial homotopy in the famous *Poincaré conjecture*: every closed 3-manifold with trivial fundamental group is homeomorphic to the 3-sphere. Thanks to the research of Moritz Epple, we now know that Dehn tried to prove the Poincaré conjecture in 1908, and thought for a while he had succeeded.

He actually sent his proposed proof to Hilbert, urging him to speed up publication in case someone else got there first ("Poincaré, for example", he says). As we now know, there was no need to be so hasty! Tietze found a mistake in Dehn's proof, and the paper was withdrawn. As Volkert (1996) puts it, Dehn nearly became "the first victim of Poincaré's conjecture".

3. The major papers

3.1. *Group theory*

Dehn was not the only mathematician to be inspired by Poincaré's results. The long paper by Tietze (1908) on multidimensional manifolds took up where Poincaré left off, introducing some crucial examples and problems for group theory and low-dimensional topology: Tietze transformations, the isomorphism problem (even the statement that it is unsolvable!), mapping class groups, knot groups, proof that the trefoil is knotted, and lens spaces. Dehn does not refer to this paper until 1914, but in Dehn (1911) he refers to a 1907 report of Tietze which mentions some of its results. It seems likely that Tietze's paper first alerted Dehn to the power of group theory in topology.

Around 1910 Dehn gave a lecture course on group theory and topology, two chapters of which were eventually published in Dehn (1987). The group theory chapter introduces his "Gruppenbild" (group diagram) a generalisation of Cayley diagrams to infinite groups, also related to the tessellation pictures of infinite discontinuous groups used by Dyck, Fricke and Klein. He begins by giving presentations and diagrams for some of the alternating and symmetric groups related to regular polyhedra, showing that the diagrams follow the shape of the corresponding polyhedra. Then he studies the group with a related presentation, by generators s_2 and s_3 and relations $s_2^5 = s_3^2 = (s_3s_2)^4 = 1$. He shows that its diagram is naturally embedded in the hyperbolic plane, and hence this group is infinite.

In fact, nearly every group discussed in the course is naturally represented by a diagram on the sphere, Euclidean plane or hyperbolic plane, and Dehn's diagram is just the dual of the tessellation picture of the group. It is nevertheless a somewhat clearer picture, since generators correspond to edges and relations to closed paths (instead of closed chains of cells). And it is better suited to topology, where elements of the fundamental group are also represented by paths. This is the origin of Dehn's interest in the word and conjugacy problems for finitely presented groups, which he was apparently the first to state. Solving the word problem amounts to deciding whether a path can be contracted to a point, and solving the conjugacy problem amounts to deciding when one path can be deformed to another.

The chapter on surface topology presents several of the results about fundamental groups of surfaces that were eventually published in Dehn (1911) and (1912), though with different proofs. In particular, he gives a geometric algorithm for the conjugacy problem similar to an algorithm used by Poincaré (1904) to detect whether a closed curve on a surface of genus > 1 is homotopic to a simple curve. He lifts the curve to the universal cover (the hyperbolic plane), where its free homotopy class (and hence the conjugacy class of the corresponding element of the fundamental group) is determined by its "ends" at infinity.

3.2. *Knots and groups*

Some of the ideas from Dehn's lectures saw the light, in suddenly matured form, in Dehn's (1910) paper on the topology of 3-dimensional space. This very rich paper includes a discussion of word and conjugacy problems for arbitrary groups, Dehn's lemma, presentation of knot groups, and the construction of homology spheres by Dehn surgery.

Dehn saw that the fundamental group translated many topological problems into problems of combinatorial group theory – for example, free homotopy of curves was equivalent to the conjugacy problem, and triviality of knots was equivalent to commutativity of the knot group – but at the same time he saw that the group theory problems were hard, and perhaps best solved by further use of topology and geometry. A substantial part of the paper is in fact a topological attack on the knot problem, with Dehn's lemma used to prove that a nontrivial knot has a noncommutative group. Unfortunately, Kneser (1929) discovered a gap in Dehn's proof, and the lemma was not proved until 1957 (by Papakyriakopoulos).

Dehn's lemma was indeed deeper than Dehn realised, and the gap in Dehn's proof held up the development of 3-manifold topology until the 1950s. Nevertheless, Dehn's "switchover" (Umschaltung) technique from his unsuccessful proof was an important tool. Kneser (1929) was the first to recognise this, and he used the technique to prove the main result of his paper. When 3-manifold theory finally took off, Kneser's result was seen retrospectively as a breakthrough, and a slew of new results was obtained by the same technique.

The 1910 paper was more successful in the construction of homology spheres. Thanks to his understanding of group diagrams, Dehn was able to give constructions which were rigorous, clear, and also highly original. Improving on the idea of his (1907) announcement, he constructs them by removing a solid torus from the 3-sphere and "sewing it back differently" – what is now called *Dehn surgery*. He actually constructs infinitely many different homology spheres, the simplest of which is homeomorphic to Poincaré's, though this was first shown by Weber and Seifert (1933). This homology sphere is still the only one known with finite fundamental group.

3.3. *Dehn's algorithm*

From 1911 to 1913 Dehn was *Extraordinarius* (associate professor) in Kiel. He prepared three more major papers during this time, the first two (1911 and 1912) bringing his study of the word and conjugacy problems for surface groups to an elegant conclusion with *Dehn's algorithm.*

Dehn (1911) is a remarkable blend of topology, algebra and geometry. A problem motivated by topology (deciding whether curves on a surface are free homotopic) is translated into a problem of algebra (the conjugacy problem for the fundamental group $\pi_1(S)$ of the surface S), and a combinatorial algorithm for its solution is justified by appeal to geometry (using the hyperbolic plane as universal cover for a surface S of genus > 1, and hence as the location of the group diagram of $\pi_1(S)$).

The algorithm and its proof can be seen as a transitional stage between the group theory notes of 1910, where the algorithm *and* its proof are both geometric, and Dehn (1912), where the algorithm and proof are combinatorial. However, this stage includes a new idea which was important for the Dehn–Nielsen theorem of the 1920's, and the geometric group theory of today – the *word metric.* Dehn defines the distance between elements g_1 and g_2 of $\pi_1(S)$ as the minimum length of a word for $g_1 g_2^{-1}$, that is, the minimum number of edges in a path connecting the vertices g_1 and g_2 in the group diagram of $\pi_1(S)$. He shows that this distance lies within constant bounds of the hyperbolic distance, and hence can be used as a substitute for it.

Dehn (1912) contains his last word on the conjugacy problem for surface groups, as well as a very simple solution of the word problem. The solutions are by what is now known as *Dehn's algorithm,* a decisive improvement on the 1911 paper both in concept and computational efficiency. In 1911, Dehn needed the existence of the hyperbolic metric but computed only with the combinatorial structure of the group diagram. His 1912 algorithm needs the existence of the group diagram but computes only with words. He shows that if a word for an element of the fundamental group is equal to 1 then it can be reduced monotonically, by replacing any subword that is more than half the defining relator by its complement, or by trivial cancellations. (For the conjugacy problem, the same operations suffice, but the word is regarded as circular.)

It would seem that Dehn here is returning to the goal of a purely combinatorial topology, the ideal of the 1907 Dehn and Heegaard encyclopedia article. His next paper however, was perhaps his most brilliant application of hyperbolic geometry.

3.4. *The two trefoil knots*

Dehn (1914) answers one of the most intuitively appealing questions in topology, by showing that the left and right trefoil knots are not equivalent. Dehn reduced the problem to finding the automorphisms of the trefoil knot group, but these were far from easy to find with the presentation and group diagram he used, taken straight from his 1910 paper. The diagram lies in the product of the hyperbolic plane with the real line, and Dehn needs all his skill in hyperbolic geometry find the automorphisms. Schreier (1924) showed that the automorphisms can be found more easily, and purely algebraically, using a presentation of the trefoil knot group with generators A, B and defining relation $A^2 = B^3$. However, no really elementary method for distinguishing the two trefoil knots was known until 1984, with the discovery of the Jones polynomial.

Towards the end of his term in Kiel, Dehn met Jakob Nielsen and took over the supervision of Nielsen's thesis after the death of Landsberg, Nielsen's original supervisor. The last section of Nielsen's thesis (1913) deals with a problem on torus mappings suggested by Dehn. This was the beginning of Nielsen's lifelong interest in surface mappings and related questions on automorphisms of groups. Dehn too seems to have begun investigating automorphisms around 1913. His first published results are in his 1914 paper on the two trefoil knots, but there he also raised the general problem: given a presentation of a group by generators and relations, find a presentation of its automorphism group of the same type.

4. Dehn's career between the wars

4.1. *Breslau*

From 1913 to 1921 Dehn was *Ordinarius* (full professor) at the Technische Hochschule, Breslau. After the trefoil knot paper his term was interrupted by army service (1915–1918), and, in fact, he produced no more topology papers until the 1930s. However, there were important developments immediately after the war, when he really began to influence the development of topology.

In 1920 and 1921 Dehn was joined by Jakob Nielsen in Breslau. The Dehn–Nielsen theorem, that every outer automorphism of $\pi_1(S)$ is induced by a homeomorphism of S, probably dates from this time. In Breslau on the 9th and 11th of March 1921, Nielsen gave a proof of the theorem for a surface of genus 2. Notes of these talks were later discovered by Fenchel and an English translation was published in Nielsen's *Collected Mathematical Papers* (1986). The notes credit the theorem to Dehn and say that Dehn's proof uses the idea from Poincaré (1904) of lifting curves to the universal cover (the hyperbolic plane) and using their ends at infinity. Nielsen himself took up this idea later, and greatly expanded its scope.

Dehn's own work on surface mappings was only partially published in his lifetime. His earliest work on the subject, "On curve systems on two-sided surfaces, with application to the mapping problem", is now available in Dehn (1987). This paper is based on a lecture by Dehn to the Breslau mathematics colloquium on 11 February, 1922. It was not published at the time but evidently circulated among a few mathematicians. Followup papers were published by Baer in 1927 and 1928, Goeritz in 1933, and Dehn himself in 1938. In Section I of the 1922 paper Dehn stakes his claim to the Dehn–Nielsen theorem and says that his proof uses essentially topological properties of the diagram of the fundamental group. As mentioned above, Dehn (1911) discovered that the advantages of the hyperbolic metric could be recaptured in a topological setting by the word metric, and this idea indeed leads to a natural proof of the Dehn–Nielsen theorem (see Appendix to Dehn (1987)). The first published proof of the theorem was given by Nielsen (1927) and, like most of Nielsen's work, it makes full use of hyperbolic geometry.

4.2. *Frankfurt*

In 1922 Dehn was called to Frankfurt to succeed Bieberbach, and stayed there until 1935. This happy period in Dehn's career has been described by Siegel (1965) and Magnus (1974). Magnus compares Frankfurt in those days to the empire of the caliphs in the time of Harun al-Rashed, described by Goethe as follows:

> Proverbially, it was a time when, in a particular locality, all human endeavours interacted in such a fortunate way that the recurrence of a similar period could be expected only after many years and in very different places under exceptionally favourable circumstances.

The heart of these fortunate endeavours was the history of mathematics seminar, founded by Dehn in 1922, and continuing under his leadership until 1935. Siegel (1965) says:

> Dehn was in a sense our spiritual leader, and we always followed his advice in choosing topics for each semester. As I look back now, those communal hours in the seminar are some of the happiest of my life. Even then I enjoyed the activity which brought us together each Thursday afternoon from four to six. And later, when we had been scattered over the globe, I learned through disillusioning experiences elsewhere what rare good fortune it is to have academic colleagues working unselfishly together without thought of personal ambition, instead of just issuing directives from their lofty positions.

It is typical of Dehn's lack of personal ambition that he readily allowed others to find proofs of theorems, and get the credit for them, instead of publishing his own proofs. We have already seen how this was the case with the Dehn–Nielsen theorem, and Magnus (1978) gives other examples. One was the Nielsen–Schreier theorem that every subgroup of a free group is free, another was the "Freiheitssatz" which Dehn set Magnus as a thesis topic in 1928. Magnus succeeded in proving it, but not by the topological method Dehn apparently had in mind.

4.3. *Setbacks and new directions*

On 22 April, 1929, Hellmuth Kneser wrote to Dehn informing him of a mistake in the Dehn (1910) proof of Dehn's lemma. Some interesting correspondence ensued (now in the Archives of American Mathematics in Austin Texas), including Kneser's beautiful representation of the Poincaré homology sphere as a dodecahedron with opposite sides identified. However, they were unable to repair the proof of the lemma. This did not greatly affect Dehn, who was no longer working on 3-manifold topology, but it was a blow to Kneser, who had to drop plans for a book on the subject based on Dehn's lemma. (I owe this information to a letter from Kneser's son Martin, who also suggests that the collapse of Dehn's proof may have been influential in his father's switch from topology to several complex variables.)

In 1930 Ruth Moufang completed a thesis on the algebra of projective planes, supervised by Dehn. The problem goes back to Hilbert (1899), who showed that the algebra of segments in a projective plane is commutative if the theorem of Pappus holds, and it is associative if Desargues' theorem holds. Moufang found a weaker theorem which implies only that the algebra is alternative (for example, the octonions). At the time, this topic did not seem to be related to topology, but it became so in the 1950s, when the existence of division algebras with various properties was found to be controlled by topology, and in particular by the properties of the spheres $\mathbb{S}^1$, $\mathbb{S}^3$ and $\mathbb{S}^7$.

In the summer of 1935, Dehn was dismissed from his position at Frankfurt, "pensioned because of Non Aryan legislation" as he later wrote on his resumé. Siegel (1965) conjectures that this was

> the result of an act of revenge by an influential ministry official in Berlin. This man had published a rather inferior mathematics text some 30 years earlier which Dehn had reviewed unfavorably.

Dehn, who was of Jewish descent (though in fact a Lutheran since the age of 18), had kept his position up to this time because of his war service, along with other Jewish colleagues at Frankfurt, Epstein and Hellinger. But by the fall of 1935 they had all lost their positions with the passage of new laws at the Nuremberg party congress. Dehn used his forced retirement to write up some topology he had been thinking about since the early 1920s.

Dehn (1936) contains a miscellany of results on combinatorial topology, mostly in areas where group theory is not applicable. This little known paper is noteworthy for the solution of a problem posed by Gauss, the "crossing problem". Read and Rosenstiehl (1976) drew attention to the efficiency of Dehn's solution.

4.4. *Mapping class groups*

The long and complicated paper Dehn (1938) was also overlooked for a long time, until its main result was rediscovered by Lickorish (1962): the mappings of a closed orientable surface are composed (up to isotopy) of twist mappings. Dehn was already working towards this result in his unpublished 1922 paper, and perhaps delayed publication because of the extreme complexity of his proof – an induction on the complexity of the surface and the mapping, with a base step involving the study of spheres with up to five holes. The proof by Lickorish is similar in concept but much simpler. Amazingly, Lickorish also rediscovered

Dehn surgery at the same time, and had the new and fruitful idea of *combining the two*, to show that all 3-manifolds may be constructed by Dehn surgery on the 3-sphere.

While the work of Lickorish enables us to skip many of the details of Dehn (1938), it does not replace the whole paper. Dehn had another idea which went unnoticed even longer, until it was rediscovered by Thurston in 1976. This was the idea of studying the mapping class group by its action on the space of simple curve systems. Dehn took this idea far enough to represent mapping classes by transformations of the $(6g-6)$-dimensional space of integer vectors, but not far enough to understand the geometric meaning of this space.

Thurston interpreted the simple curve systems as "rational points" in the boundary of the Teichmüller space of surfaces of genus p. This space is defined as the space of hyperbolic structures on a surface of genus p, and, hence, demands a return to the hyperbolic geometry Dehn had abandoned in his quest for purely combinatorial proofs. What Dehn did not know is that mapping classes are easier to understand when they act on the whole Teichmüller space, rather than on the rational points in its boundary, because the whole space is, in fact, a topological ball.

The results of Dehn and Lickorish show that the mapping class group is finitely generated, but they do not give a set of defining relations. Undoubtedly, Dehn would have liked to do so, as part of his program from 1914 of finding the automorphism groups of finitely presented groups, but this was first done by McCool (1975).

5. Dehn in America

5.1. *Escape from Germany*

Dehn stayed in Germany until 1939. As Siegel (1965) explains

> Despite the increased oppression in Germany, older Jews often decided not to emigrate. They would have had to surrender their savings in accordance with the strict new regulations, leaving them with only ten Marks to try to start a new life abroad. In the first few years after 1933, so many university-educated emigrants had gone to America that it would have been almost impossible for an older professor to begin a new existence there; at the same time, European countries would grant a foreigner resident status only if he were capable of supporting himself and had brought his fortune with him.

Nevertheless, Dehn explored possibilities outside Germany, and sent his children to study in England. The way this came about was quite remarkable. In the early 1930s Dehn sent his daughter Maria to study at a school near Ulm called the Herrlingen school. This was on the advice of one of his former students, Ado Prag, who taught Greek, Latin and mathematics there. Dehn became quite involved in school affairs, including plans to move the school to another country. After much searching for a suitable location, it was moved to England in September 1933, and became the New Herrlingen school near Favesham in Kent. Maria moved with it, and was later joined by her younger sister Eva. Dehn himself taught mathematics at the school between January and April 1938. This little known episode in his life is mentioned in his resumé, which is in a collection of Dehn documents maintained by Linda Hill at Idaho State University.

Dehn returned to Germany, and Siegel relates what happened to him in November 1938.

> The real terror in Germany began in earnest on November 10, 1938, when the government organized an anti-Jewish pogrom: synagogues were burned, many Jewish businesses destroyed, and all the concentration camps in existence at the time were filled to overflowing with Jews who had been dragged from their homes. Hitler's minions went after Dehn, Epstein and Hellinger to cart them off as well. After an initial period of arrest, however, Dehn was released by the police because there was no more room in Frankfurt to keep prisoners under lock and key. To avoid being taken again the following day, Dehn and his wife left for Bad Homburg, where they found asylum with our friend and colleague Willy Hartner.

This lucky escape is described more vividly by Maria in her letter to me:

> I was told that when the Nazis came for my father he was very deliberate, insisting on fetching his hat to go out. When they got him down to the station, they were barked at "We told you not to bring in anyone after 7" – so my father had to go home again! Mother was still standing, clueless, at the top of the stairs (as he had left her), when he came back! He then spent several days riding all over the railroads till their safe haven with the Hartners in Hamburg was secured.

It was there that Siegel met him, after initially seeking Dehn at home to congratulate him on his 60th birthday. A few days later, Dehn was smuggled on to a train to Hamburg with the help of Wilhelm Magnus. In Hamburg he again met Siegel and, together with a Danish colleague and former student of Dehn's (could this have been Nielsen?) discussed the possibility of emigrating to Scandinavia.

For a while, this plan was successful. Dehn and his wife left Germany for Copenhagen in January 1939, later moving to Trondheim in Norway, where Dehn took the place of Viggo Brun at the Technical University until early 1940. But when Siegel visited him there in March 1940 more trouble was on the horizon. Ships carrying German flags were loitering in the harbour, ostensibly with engine trouble. In reality, they had a more sinister purpose, as Siegel explains:

> They were filled with war material for the German soldiers who suddenly occupied Trondheim on the day of the invasion of Norway. They were followed by the Gestapo and the national-socialist party organisation. ... Dehn escaped to a farmer's house in the early days of the German occupation, but returned to Trondheim when, at least at first, no further acts of violence or arrests had occurred. In the course of the next few months, Hellinger and a few of his other friends were laying the groundwork for Dehn's second emigration.

5.2. *The years in America*

With the help of Hellinger and Clare Haas, a family friend who had fled earlier and found work at the State Hospital South in Blackfoot, Idaho, Dehn obtained a position at Idaho Southern University (now Idaho State) in Pocatello. A wealth of information from this period exists in the Dehn documents at Idaho State. The local newspaper reported Dehn's arrival and Dehn related the story of his escape from Europe in an address to the university faculty. He and his wife Toni left Norway in October 1940, and travelled to America via Stockholm, Moscow, Siberia and Japan – in the belief that it was then more dangerous to cross the Atlantic. During their 10 day rail journey across Siberia the temperature dropped to around -50° and Dehn contracted what the doctor in Irkutsk cheerfully told him was a "touch" of pneumonia. At this stage the Dehns were forced to wash with eau d'cologne, it being the only liquid that did not appear to freeze.

On arrival in Vladivostok, Dehn managed to contact Dean Nichols of Idaho Southern University to tell him they were safe. Dehn was apparently worried about arriving early enough to prepare his lectures, as Nichols wrote to him in Tokyo to assure him there was plenty of time. Dehn "took a deep breath of recovery" on arrival in Japan, basking in the warm climate, feasting his eyes on the flowering plants and colorfully dressed women, and admiring the skill and precision of Japanese craftsmanship. From Japan the Dehns travelled by ship to San Francisco, arriving on 1 January, 1941.

They stayed in Pocatello for about a year, with Dehn teaching Freshman Algebra, History of Ancient Philosophy and History of Modern Philosophy. Toni Dehn had visited America 30 years earlier, and spoke excellent English, so the culture shock was not as severe as it might have been. The beauty of the surrounding countryside and opportunities for hiking also helped. In the summer of 1941 it was a little like old times in the Taunus mountains near Frankfurt, when Dehn and his wife were visited by Siegel and Hellinger.

The Dehns were well liked at the University of Idaho, but it was not a permanent job, nor was it academically challenging. In 1942 they moved to Chicago, where Dehn was given a job at the Illinois Institute of Technology. The pay was higher, but otherwise IIT was less pleasant than Pocatello. As Siegel (1965) recalls, Dehn disliked the turbulence of the big city and

> One semester he had to deliver the same lecture [course] on differential calculus twice – once for the new students and once again for the students who had understood nothing of it from the previous semester. He told me that this latter group had greeted him at the beginning of the first lecture with: "Hello Professor, we're the dumb ones".

In 1943/1944 Dehn moved to St. John's College in Annapolis, Maryland, and spent a particularly unhappy year. The school had an impossibly ambitious program of "great books in the original languages", which Dehn was required to teach to teenagers who, in some cases, had not even mastered English.

In 1945 Dehn finally found peaceful and productive conditions in an unlikely place, Black Mountain College in North Carolina. This period of his life has been reported by Sher (1994). Black Mountain College was a small, liberal arts college founded in 1933, with strong emphasis on the creative arts. There were only about 100 members, including faculty, and Dehn was the only mathematician. Nevertheless, he was happy at Black Mountain, teaching not only mathematics but also philosophy, Latin and Greek. He died there on 27 June, 1952.

Bibliography

Alexander, J.W. (1924), *On the subdivision of 3-space by a polyhedron*, Proc. Nat. Acad. Sci. **10**, 6–8.

Dehn, M. (1907), *Berichtigender Zusatz zu III AB3 Analysis situs*, Jahresber. Deutsch. Math.-Verein. **16**, 573.

Dehn, M. (1910), *Über die Topologie des dreidimensionalen Raumes*, Math. Ann. **69**, 137–168.

Dehn, M. (1911), *Über unendliche diskontinuierliche Gruppen*, Math. Ann. **71**, 116-144.

Dehn, M. (1912), *Transformation der Kurven auf zweiseitigen Flächen*, Math. Ann. **72**, 413–421.

Dehn, M. (1914), *Die beiden Kleeblattschlingen*, Math. Ann. **75**, 402–413.

Dehn, M. (1936), *Über kombinatorische Topologie*, Acta Math. **67**, 123–168.

Dehn, M. (1938), *Die Gruppe der Abbildungsklassen*, Acta Math. **69**, 135–206.

Dehn, M. (1987), *Papers on Group Theory and Topology*, Springer, Berlin.

Dehn, M. and Heegaard, P. (1907), *Analysis situs*, Enzyklopädie der Mathematischen Wissenschaften, III AB3, 153–220.

Dupont, J.L. and Sah, C.-H. (1982), *Scissors congruences, II*, J. Pure Appl. Algebra **25**, 159–195.

Guggenheimer, H. (1977), *The Jordan curve theorem and an unpublished manuscript of Max Dehn*, Arch. Hist. Exact Sci. **17**, 193–200.

Hilbert, D. (1899), *Grundlagen der Geometrie*, Teubner, Leipzig.

Kneser, H. (1929), *Geschlossene Flächen in dreidimensionalen Mannigfaltigkeiten*, Jahresber. Deutsch. Math.-Verein. **38**, 248–260.

Lickorish, W.B.R. (1962), *A representation of closed orientable* 3-*manifolds*, Ann. Math. **76**, 531–540.

Magnus, W. (1974), *Vignette of a cultural episode*, Studies in Numerical Analysis, B.K. Scaife, ed., Academic Press, New York, 7–13.

Magnus, W. (1978), *Max Dehn*, Math. Intelligencer **1**, 132–143.

Magnus, W. and Moufang, R. (1954), *Max Dehn zum Gedächtnis*, Math. Ann. **127**, 215–227.

McCool, J. (1975), *Some finitely presented subgroups of the automorphism group of a free group*, J. Algebra **35**, 205–213.

Nielsen, J. (1913), *Kurvennetze auf Flächen*, Inaugural-Dissertation, Kiel.

Nielsen, J. (1927), *Untersuchungen zur Topologie der geschlossenen zweiseitigen Flächen, I*, Acta Math. **50**, 189–358.

Nielsen, J. (1986), *Collected Mathematical Papers*, Birkhäuser, Basel.

Papakyriakopoulos, C.D. (1957), *On Dehn's lemma and the asphericity of knots*, Ann. Math. **66**, 1–26.

Poincaré, H. (1900), 2^{d} *complément à l'Analysis situs*, Proc. London Math. Soc. **32**, 277–308.

Poincaré, H. (1904), *Cinquième complément à l'Analysis situs*, Rend. Circ. Mat. Palermo **18**, 45–110.

Read, R.C. and Rosenstiehl, P. (1976), *On the Gauss crossing problem*, Colloq. Math. Soc. János Bolyai **18**, 843–876.

Schreier, O. (1924), *Über die Gruppen* $A^a B^b = 1$, Abh. Math. Sem. Univ. Hamburg **3**, 167–169.

Sher, R.B. (1994), *Max Dehn and Black Mountain College*, Math. Intelligencer **16**, 54–55.

Siegel, C.L. (1965), *Zur Geschichte des Frankfurter mathematischen Seminars*, Gesammelte Abhandlungen, Vol. 3, Springer, Berlin. English translation: Math. Intelligencer **1**, 223–230.

Stillwell, J.C. (1979), *Letter to the Editor*, Math. Intelligencer **1**, 192.

Tietze, H. (1908), *Über die topologischen Invarianten mehrdimensionaler Mannigfaltigkeiten*, Monatsh. Math. Phys. **19**, 1–118.

Volkert, K. (1996), *The early history of Poincaré's conjecture*, Henri Poincaré, Science and Philosophy, J.-L. Greffe, G. Heinzmann and K. Lorenz, eds, Akademie-Verlag, Albert Blanchard, 241–250.

Weber, C. and Seifert, H. (1933), *Die beiden Dodekaederräume*, Math. Z. **37**, 237–253.

CHAPTER 37

Jakob Nielsen and His Contributions to Topology

Vagn Lundsgaard Hansen

Department of Mathematics, Technical University of Denmark, DK-2800 Lyngby, Denmark
E-mail: V.L.Hansen@mat.dtu.dk

Two of the central figures in Danish mathematics in the first half of this century were Harald Bohr (1887–1951) and Jakob Nielsen (1890–1959). Both of them won international recognition, but the immediate impact of their work in Denmark was not of the same magnitude. Harald Bohr quickly established a strong school of analysis in Copenhagen through his pioneering work on almost periodic functions. The deep work in group theory and the topology of surface transformations of Jakob Nielsen did not immediately attract many students and maybe the time was, in fact, not quite ripe for it, when Nielsen was at his height as a mathematician. A few years after the death of Nielsen, a Danish school in algebraic topology was, however, founded in Aarhus by Leif Kristensen, student of Nielsen's close collaborator and friend Werner Fenchel in Copenhagen and Saunders MacLane in Chicago. Internationally, the impact of the work of Jakob Nielsen has never been stronger than now towards the end of the 20th century.

The biography of Jakob Nielsen and the description of his mathematical work given below is a modified version of my paper [9] supplied with new information on Nielsen's relations to other mathematicians. The memorial paper by Werner Fenchel [6], reprinted in [15, Vol. 1], contains exact references to the works of Jakob Nielsen mentioned in the following.

I am indebted to Dirk van Dalen, Erik Bent Hansen, Kurt Ramskov, Asmus Schmidt and Christian Siebeneicher for supplying valuable pieces of information in connection with the present biography of Jakob Nielsen.

1. Childhood and school years

Jakob Nielsen was born in the small village of Mjels on the island of Als in North Schleswig on October 15, 1890 as the youngest of four children. North Schleswig, called Sønderjylland in Denmark, was part of Germany in the period 1864–1920 but has since 1920 again been the southern part of Denmark. Like many people of Danish origin from this region, Jakob Nielsen felt a strong association with his birthplace throughout his life.

HISTORY OF TOPOLOGY
Edited by I.M. James

Jacob Nielsen (1890–1959)

At the age of 3 he lost his mother. In the following years, an aunt, who was a teacher at Rendsburg, followed his progress closely. She noticed early on that he was an unusually gifted child, and in the year 1900 he moved to live with her in Rendsburg, as this town offered a far better educational system. Here he attended the so-called Realgymnasium, where the teaching of Latin carried considerable weight in the curriculum. Throughout his life he retained a deep passion for Roman poetry. After a few years, the relations between him and his aunt deteriorated because they were both rather uncompromising characters, and at the age of 14 he left her. For the remainder of his schooldays, and later during all his years of study, he earned his living by tutoring pupils in a variety of subjects – even Norwegian, as he once put it.

In December 1907, he was expelled from the Realgymnasium because he and a few of his schoolmates had founded a pupils' club which, though quite harmless, was against the rules. Full of confidence, however, he continued studies on his own and matriculated at the University of Kiel in the Spring of 1908, also obtaining his school-leaving certificate privately at Flensburg in the Autumn of 1909.

2. Years of study in Kiel and doctorate

Jakob Nielsen spent all his years of study in Kiel, with the exception of the summer term 1910, which he spent at the University of Berlin. In the first years of study, he attended lectures in mathematics, physics, chemistry, geology, biology, literature, and philosophy. Only slowly did mathematics begin to play the central role in his studies, but philosophy also was a subject close to his heart.

Among his teachers in Kiel, Jakob Nielsen valued in particular the mathematician Georg Landsberg, known among other things for his work on algebraic functions. It was Landsberg who encouraged Nielsen to study the mathematical problems underlying his doctoral dissertation of 1913. In a short autobiography appended to the thesis, he expresses his devotion to Landsberg, who had died shortly before.

Of the outmost importance to Nielsen for his start as a research mathematician, however, was Max Dehn (1878–1952), who was attached to the university in Kiel at the end of the year 1911. Dehn was already considered an eminent mathematician, and through him Nielsen came into contact with the most recent advances in topology and group theory. It was precisely to these fields that Nielsen devoted most of his research.

In itself, Nielsen's thesis was not epochmaking. It contains, however, the germs of his pioneering work on surface transformations which we shall describe in some detail. In addition, the group theoretical papers of Nielsen, including a very important paper in Danish from 1921 on free groups, have roots going back to his years of study in Kiel, and, in particular, to the inspiration from Dehn, with whom he developed a life-long friendship.

3. Military service and marriage

Immediately upon receiving his Ph.D. in 1913, Jakob Nielsen was drafted to do military service in the German Imperial Navy and was assigned to coastal defense. After the outbreak of the war in 1914, he was sent first to Belgium and then, in April 1915, to Constantinople as one of the German officers assigned to advise the Turkish government on the defense of the entrance to the Black Sea through the Bosporus and the Dardanelles. At the end of the First World War in 1918, he finished his military service. During these years there was, of course, not much time for mathematics, but somehow Jakob Nielsen found time to write a couple of important short group theoretical papers published in Mathematische Annalen and a paper on a subject from ballistics. The home journey from Turkey went through Russia and Poland in November 1918, and during this journey he kept a diary, which was published in the Danish newspaper 'Politiken' on the tenth anniversary of Armistice Day. The whole period made a strong impression on him and probably contributed to his complete open-mindedness throughout his life towards people with a background different from his own. In the spring of 1919, Jakob Nielsen married the German medical doctor Carola von Pieverling, whom he had met in Hamburg shortly after the war. They lived a happy family life together and had one son and two daughters.

4. The early career of Jakob Nielsen

In the summer term 1919, Jakob Nielsen stayed at Göttingen, which was at that time the undisputed centre for mathematical research in the world. Here, he was especially attracted by the algebraist Erich Hecke (1887–1947), and when Hecke the same year received nomination to a chair in mathematics at the recently established university in Hamburg, Jakob Nielsen followed Hecke as his assistant with the title of 'Privatdozent'. Already in 1920, however, Nielsen himself was named to a professorship at the Technical University of Breslau. Here he could resume close contact with Max Dehn, who had been a professor at the University of Breslau for some years. In two inaugural lectures in Breslau in 1921, Jakob Nielsen formulated clearly that circle of problems concerning surface transformations upon which he was so strongly engaged for the rest of his life. Handwritten notes from these lectures were translated into English and published for the first time in 1986 in connection with the publication of his collected mathematical papers [15].

The most spectacular mathematical work of Jakob Nielsen from his early career is his work in combinatorial group theory, which was just beginning at the time with emphasis on finding descriptions of groups by generators and relations. For that purpose free groups play a decisive role.

In a very important paper in Danish published in Matematisk Tidsskrift in 1921, Nielsen proved that every subgroup of a free group is itself free. (This fundamental paper is included in his collected works in an English translation by Anne W. Neumann, which was first published in The Mathematical Scientist 60 years after the original paper.) Nielsen assumed the free group to be finitely generated, but 5 years later, Otto Schreier proved that this assumption is not necessary, so that the result is true in complete generality. The theorem, which is now known in the mathematical literature as the *Nielsen–Schreier theorem*, is important among others when dealing with the relations in a group. Though this result is extremely significant, the main goal of Nielsen was, however, to describe the automorphism group of a free group, that is the group of isomorphisms of the free group onto itself. For that purpose, he introduced some basic automorphisms, now known as *Nielsen transformations*. In this context, a paper in Mathematische Annalen from 1918 should be mentioned, in which it is shown that the automorphism group of the free group on n generators is generated by $n + 1$ automorphisms. The corresponding relations for the automorphism group are determined in a paper which is very difficult to read, published in Mathematische Annalen in 1924.

Although his work in combinatorial group theory quickly won recognition, Nielsen did not take it up again in major scientific publications before his very last research paper from 1955.

The theory of automorphisms of free groups has only recently approached a definitive stage with works of Culler and Vogtman [5] and Bestvina and Handel [1].

5. Jakob Nielsen at his height

Jakob Nielsen's stay in Breslau was not to last long, for at the reunion of North Schleswig with Denmark in 1920, he opted for Denmark, and in 1921 he took over the vacant position as lecturer in mathematics at the Royal Veterinary and Agricultural University in Copenhagen. In 1925, he succeeded Christian Juel as professor of rational mechanics at the Technical University, where mathematicians held this chair. Following tradition, the name of the chair only determined the teaching duties of the professor.

For some years, Jakob Nielsen based his teaching of mechanics on Juel's textbook, but gradually it became clear that a revision was needed. Nielsen plunged into this work with great energy, and the new textbook in rational mechanics was published in two volumes in 1933/1934. The book was to a large extent original pedagogical work on an advanced level for its time, and in his exposition, Nielsen made extensive use of mathematical tools such as vectors and matrices, which were then relatively new concepts in textbooks. The text is not very easy, and Jakob Nielsen's lectures were rather demanding on the part of the students. He was, however, well known for his ability to express himself with great clarity and intensity.

The lectures were also for mathematics students at the University of Copenhagen, and at a 25 years anniversary meeting in 1985, my old mathematics teacher at secondary school ('gymnasium') told me that he had attended one of the very first courses Jakob Nielsen gave on rational mechanics in Copenhagen. He was immensely impressed by Jakob Nielsen and also very proud that he received a top mark in the course.

In 1935, the textbook was translated into German by Werner Fenchel, who by invitation of Harald Bohr had come to Copenhagen shortly before as a refugee from Nazi Germany; it was reprinted by Springer-Verlag in 1985. In 1941, the teaching of aerodynamics was introduced at the Technical University of Denmark and put into the hands of Jakob Nielsen. The more theoretical parts of the lecture notes in this connection were published in 1952 as the third volume of his textbook on theoretical mechanics. The book is remarkable for its clear distinction between the empirical foundation and the mathematical theory.

As soon as the work with the first edition of the textbook was completed, Jakob Nielsen returned to his earlier studies of topology and group theory and thus over a span of more than 30 years he published his pioneering work on surface transformations, marked in particular by four long memoirs in Acta Mathematica (1927, 1929, 1932, 1942) and a memoir in 'Meddelelser fra det Kongelige Danske Videnskabernes Selskab' from 1944.

About the life of mathematicians in Denmark in the twenties and thirties, Harald Bohr has given a vivid description in his retrospective lecture "Et Tilbageblik" delivered on the occasion of his 60th birthday. In this lecture ([3, p. xxxi]), Bohr said:

> Bonnesen, Jakob Nielsen and I followed each other's work with keen interest during those years, and many Tuesday evenings Bonnesen and I walked to Hellerup to visit Jakob Nielsen, who was distinguished among us by being in possession of a blackboard, and in a cosy atmosphere we told each other what was on our minds.

It also belongs to the picture of the friendly atmosphere that Jakob Nielsen in 1919 had bought a little house on Als and that Harald Bohr a few years later followed his example. In [6, p. xii], Werner Fenchel writes:

> Year after year, in the summer vacation, a group of mathematicians, young and old, Danish and foreign, gathered about those two (Bohr and Nielsen). Apart from normal holiday activities, the study of mathematics was pursued. Not a few advances and discoveries were presented in Bohr's little half-timbered house, in the study remarkable for its blackboard – unforgettable experiences which are remembered with gratitude by all who had the privilege of attending.

Of course, not everything was idyllic. Also Jakob Nielsen had his quarrels. I have been told that at a certain occasion, professor Richard Petersen, a mathematics colleague, asked Jakob Nielsen that they stopped being informal – which they had earlier decided to be after solemn agreement – and returned to addressing each other formally (like German 'Sie') since they had so many disagreements about teaching.

When it really came down to details, Jakob Nielsen worked, however, very much on his own on the problems on surface transformations. His work was highly respected by his contemporaries, but it was not in the main stream of the then burgeoning field of algebraic topology. There does exist a small correspondence between Jakob Nielsen and L.E.J. Brouwer who handled two short papers of Nielsen on fixed points for surface transformations published in Mathematische Annalen in 1920. And it is clear that Brouwer found the work of Nielsen interesting. There is no evidence of any correspondence with Solomon Lefschetz on fixed point theory, as could have been expected, in the Nielsen archives discovered by Sigurd Elkjær at Mathematical Institute, University of Copenhagen.

At the beginning of the German occupation of Denmark 1940–1945 during the Second World War, some attempts were made to bring Jakob Nielsen to America, since it was feared that he might be assaulted by the Nazis. In a letter from Oswald Veblen at Princeton University to Gustav Hedlund at the University of Virginia dated May 23, 1940 (Veblen Papers, Library of Congress, Washington, D.C.), it is said:

> Nielsen has no things of Jewish blood, but he was born on the island of Als which is in the part of Denmark which was ceded by Germany to Denmark after the last war. He was at one time a professor in Germany (Breslau), but he elected to become a Danish citizen and to take a chair in Copenhagen. He is a great friend of Harald Bohr and has done a great deal to help refugees from Germany. It is pretty sure that he is well known to the German secret police.

Later in the same letter it is said:

> You, of course, know Nielsen's scientific work at least as well as I do, but if it is of any use to you you may quote my (Veblen) opinion that he (Nielsen) is one of the leading topologists of the world. I have been particularly impressed with the fact that he has gone after the simple hard problems, rather than the showy generalizations.

As it turned out, Nielsen stayed in Denmark during the war and he was not assaulted by the Nazis.

6. The later career of Jakob Nielsen

Jakob Nielsen often lectured on topology and group theory to small groups of interested younger mathematicians at the University of Copenhagen. Of particular importance is a series of lectures in the year 1938–1939 on discontinuous groups of isometries in the hyperbolic plane. Inspired by these lectures, in 1942 Svend Lauritzen wrote his thesis: "En Indledning til en gruppeteoretisk Behandling af de ikke orienterbare Flader" (An introduction to a group theoretical study of the non-orientable surfaces). It contains no mention of Nielsen's lectures in which the corresponding study for orientable surfaces was presented; and the thesis remained published in Danish only.

It quickly proved desirable to take up studies of discontinuous groups in their full generality, and Jakob Nielsen began with Werner Fenchel (1905–1988) to prepare a manuscript for a monograph on this subject. Even though he was heavily engaged in this project, Nielsen could only devote a limited part of his time to it because after the Second World War 1939/1945 he became more and more involved in international work, in particular in UNESCO, where he was a member of the executive board from 1952 to 1958. In this context also, he was highly esteemed for his personal integrity.

After the death of Harald Bohr in 1951, Jakob Nielsen was nominated as his successor as professor of mathematics at the University of Copenhagen. Already in 1955 he resigned from the chair, however, since he no longer felt that he could carry out his work as a professor fully due to his many international obligations. A first version of the manuscript just mentioned was completed, but both Jakob Nielsen and Werner Fenchel felt that it needed a thorough revision. The revision was not finished when Nielsen died in 1959, and later the original of the manuscript was stolen from a parked car, much to the embarrassment of Fenchel, who was extremely careful in all matters. Various copies have, however, circulated among specialists, and in several cases, other mathematicians have found alternative proofs of the most important results in the manuscript. Major parts of the theory, now known as the *Fenchel–Nielsen theory*, have therefore gradually become known among the researchers in the field.

Over the years, the manuscript, which Jakob Nielsen arranged for publication in the Princeton Mathematical Series in the late 1940s, has gained quite a lot of fame. As time passed without the necessary revision being completed, Princeton University Press faded out as a publisher of the book, and Werner Fenchel was not particularly happy to be reminded of the unpublished manuscript, which he felt had several shortcomings. He found that they could be resolved by considering trigonometric formulas resulting from hyperbolic geometry and wrote up notes to prepare a typed form of his ideas. He finished this work in 1986 and decided to have it published as a separate book "Elementary Geometry in Hyperbolic Space", published posthumously in 1989 by de Gruyter. Undoubtedly, Fenchel saw his book as an introduction to the first chapter of the Fenchel–Nielsen manuscript. In his last years, Fenchel almost completed the revision of the total manuscript, and after further work, particularly by Christian Siebeneicher in Bielefeld and Asmus Schmidt in Copenhagen, it will be published by de Gruyter in the near future. By then, the Fenchel–Nielsen manuscript will represent an important piece of history of mathematics.

Jakob Nielsen was elected member of the Royal Danish Academy of Sciences and Letters in 1926, and in his last years he lived in the Academy's honorary residence close to the castle of Hamlet in the town of Elsinore.

In January 1959 Jakob Nielsen was stricken by the illness which led to his death on August 3, 1959.

7. The work of Jakob Nielsen on the mapping class group of a surface

An important method in the investigation of a geometrical object is to study its degree of symmetry. If you are interested only in the topological properties, you need also to consider "qualitative symmetries", where certain distortions are allowed. Nielsen's investigations deal with topological transformations ("qualitative symmetries") of surfaces. His four long memoirs in Acta Mathematica and the memoir in 'Meddelelser fra det Kongelige Danske Videnskabernes Selskab' have already been mentioned. When Nielsen began his studies of transformations (homeomorphisms) of surfaces, topology was still at a formative stage with its roots particularly in work by Poincaré at the end of the previous century. Concerning the study of manifolds, the subject had not yet come very far, but it was known that one can realize every closed, orientable surface in space by adding handles to a sphere; the number of handles being the *genus* of the surface. As Nielsen writes in his first long memoir in Acta Mathematica from 1927, "the 2-dimensional manifolds (i.e. the surfaces) have thereby prematurely offered themselves for deeper study", and he gives almost no further motivation to embark on his detailed study of surface transformations.

Let φ denote a closed, orientable surface of genus $p \geqslant 1$, and let $\mathcal{M}(\varphi)$ denote the group of isotopy classes of orientation preserving homeomorphisms of φ onto itself. Two homeomorphisms belong to the same *mapping class* if they can be continuously deformed into each other through homeomorphisms. By a result of Baer from 1928, it is sufficient to require that the two homeomorphisms be homotopic. The group $\mathcal{M}(\varphi)$ is called the *mapping class group* of φ. Already in his thesis from 1913, Nielsen had proved that the mapping class group of the torus (a surface of genus 1) is nothing but the so-called *elliptic modular group* $SL(2, \mathbb{Z})$ of integral (2×2)-matrices with determinant 1; a group closely associated with the theory of doubly periodic algebraic functions.

Now consider a closed, orientable surface of genus $p \geqslant 2$. As the keystone in his investigations of surface transformations, Jakob Nielsen in 1942 succeeded in proving for a surface of genus $p \geqslant 2$ that if the nth iterate of a homeomorphism of the surface can be deformed into the identity homeomorphism, then the homeomorphism itself can be deformed into another homeomorphism for which the nth iterate is exactly the identity homeomorphism. This main result on surface transformations can be given the following formulation: Every cyclic subgroup of $\mathcal{M}(\varphi)$ can be represented by a cyclic subgroup of homeomorphisms of φ. In this formulation, the problem can be generalized. In 1948, Fenchel proved that every finite solvable subgroup of $\mathcal{M}(\varphi)$ can be represented by a subgroup of homeomorphisms of φ. In 1981, Kerckhoff proved this result in complete generality for an arbitrary finite subgroup of $\mathcal{M}(\varphi)$, thereby solving what by then had become known as the *Nielsen realization problem*. Heiner Zieschang pointed out in 1976 that one of Nielsen's arguments in his memoir of 1942 is not correct in that his proof does not cover all cases. The first complete proof of Nielsen's theorem is, therefore, contained in Fenchel's paper of 1948, where other methods are used. Nielsen's general description of surface transformations in terms of primitive homeomorphisms, which is perhaps even more important and to which we shall return at the end of this paper, is completely correct, however.

8. On Nielsen fixed point theory

Major parts of Nielsen's research on surface transformations have to do with the study of fixed points of homeomorphisms of an orientable closed surface φ of genus $p \geqslant 2$ onto itself. We shall describe some of the investigations, which have led to the development of a theory now known as *Nielsen fixed point theory*.

The main tools in Nielsen's investigations are the notions of fundamental group and universal covering space. In the case of an orientable surface φ of genus $p \geqslant 2$, the universal covering space Φ can be identified with the interior of the unit disc in the complex plane. As Poincaré has shown, Φ can be equipped with a non-Euclidean metric, thereby providing a model of *the hyperbolic plane*. Accordingly, the fundamental group F of ϕ (the group of covering space transformations) can be identified with a group generated by $2p$ *hyperbolic translations* (special Möbius transformations) in Φ. Nielsen made extensive use of this non-Euclidean setting and he worked as comfortably in the hyperbolic as in the Euclidean plane.

Every homeomorphism $\tau : \varphi \to \varphi$ can (in many ways) be lifted to a homeomorphism $t : \Phi \to \Phi$,

$$\begin{array}{ccc} \Phi & \xrightarrow{t} & \Phi \\ \downarrow & & \downarrow \\ \varphi & \xrightarrow[\tau]{} & \varphi \end{array} .$$

By a clever argument Nielsen shows that $t : \Phi \to \Phi$ can be extended to the closed unit disc $\overline{\Phi}$, so that t defines a homeomorphism $t|E : E \to E$ of the unit circle E onto itself. It is by a close examination of the homeomorphism $t|E$ that Nielsen gets his results about $\tau : \varphi \to \varphi$.

In the 1927 memoir in Acta Mathematica, it is shown to begin with that every automorphism of the fundamental group of φ, that is, of the group F, can be realized by a homeomorphism of φ onto itself. This theorem is due to Dehn, but the first proof of it in print is in Nielsen's memoir. Nielsen later always gave full credit to Dehn, and the theorem is now known as the *Dehn–Nielsen theorem*.

The main bulk of the work is devoted to an analysis of the fixed point set of $\tau : \varphi \to \varphi$. A fixed point $\tilde{x} \in \Phi$ for a lift $t : \Phi \to \Phi$ of $\tau : \varphi \to \varphi$, that is, $t(\tilde{x}) = \tilde{x}$, is projected onto a fixed point $x \in \varphi$ for τ, that is, $\tau(x) = x$. Two lifts $t,\ t' : \Phi \to \Phi$ of the same homeomorphism $\tau : \varphi \to \varphi$ have the same fixed point projections onto φ if and only if they are conjugate under F, that is, $t' = TtT^{-1}$ for a hyperbolic translation $T \in F$. Every fixed point $x \in \varphi$ for $\tau : \varphi \to \varphi$ is the projection of a fixed point $\tilde{x} \in \Phi$ for some lift $t : \Phi \to \Phi$ of τ. The collection of fixed points for τ, which are the projections of all the fixed points for the lifts t of τ in a conjugacy class of lifts, is called a *fixed point class*. The index of an (isolated) fixed point x of τ measures the twisting of τ about x and can be identified with the winding number, as in complex analysis, of $\mathbf{1} - \tau$, where $\mathbf{1}$ denotes the identity map. After deforming the map τ to have only finitely many fixed points, the *index* of a fixed point class is defined to be the sum of the indices of its members. The number of fixed point classes with index $\neq 0$ is now called the *Nielsen number* of τ and is denoted by $N(\tau)$. It was this number that Nielsen tried to determine. Clearly, the Nielsen number $N(\tau)$ provides a lower bound for the number of fixed points of τ.

It can be proved that two fixed points $x_1, x_2 \in \varphi$ for $\tau : \varphi \to \varphi$ belong to the same fixed point class if and only if x_1 and x_2 can be connected by a curve C in φ such that C is homotopic to $\tau(C)$ by a homotopy keeping x_1 and x_2 fixed. The Nielsen number $N(\tau)$ is again the number of "essential" fixed point classes, that is, those with index $\neq 0$. In this formulation, the notion of fixed point classes and indexes for these can be generalized to mappings $f : X \to X$ between more general types of spaces than surfaces, for example, polyhedra and manifolds. There is an extensive literature on the subject.

Homotopic maps $f, g : X \to X$ have the same Nielsen number, that is, $N(f) = N(g)$. The following question about the Nielsen number $N(f)$ for a map $f : X \to X$ is therefore interesting: Does there exist a map $g : X \to X$ homotopic to $f : X \to X$, such that the number of fixed points for g is exactly $N(g) = N(f)$? In other words: Can the Nielsen number be realized? In 1942, it was proved by Wecken that for a fairly large class of finite polyhedra, containing among others all triangulable manifolds of dimension $\geqslant 3$, such a minimality theorem holds.

Nielsen conjectured in his 1927 memoir that Nielsen numbers of maps of surfaces can be realized. For homeomorphisms, the answer is correct, as Nielsen himself proved in part (see also [16]), though the proof was completed only recently [13]. However, for continuous maps, in [10, 11], Boju Jiang produced examples which show that it is not always possible to realize the Nielsen number on surfaces. In fact, for any surface of negative Euler characteristic, Jiang has recently proved that there is a map such that the gap between its Nielsen number and the minimum number of fixed points of all maps homotopic to it is arbitrarily large [12]. This would most certainly have come as a surprise to Jakob Nielsen. The survey paper by R.F. Brown [4] discusses the realization problem further.

9. The synthesis of Nielsen's work on surface transformations

As the synthesis of his work on homeomorphisms of a closed, orientable surface φ of genus $p \geqslant 2$ – for obvious reasons also called a *hyperbolic surface* – Jakob Nielsen gained the deep insight that up to isotopy, and possibly after a finite iteration, every homeomorphism of a hyperbolic surface can be written as a composition of certain primitive homeomorphisms defined essentially on disjoint subsurfaces. The work of Nielsen was based, as indicated above, on a thorough analysis of the fixed point sets for the homeomorphisms. Nielsen found that the primitive homeomorphisms of a hyperbolic surface were of two types, one type consisting only of periodic homeomorphisms. The second type of primitive homeomorphisms was not clearly identified before the end of the 1970s, where, by completely different methods, William Thurston found that they are nonperiodic and that they preserve a pair of transverse measured foliations by geodesic lines. They are now called pseudo-Anosov homeomorphisms. These homeomorphisms are very important both in the theory of 3-dimensional manifolds and in the study of iterations of mappings. The study of the connections between the work of Nielsen and the work of Thurston has been the subject of several papers, among which we mention the papers by Gilman [7], Miller [14], and Handel [8]. See also the paper by Thurston [16] and the book by Bleiler and Casson [2].

Bibliography

[1] M. Bestvina and M. Handel, *Train tracks and automorphisms of free groups*, Ann. of Math. **135** (1992), 1–51.

[2] S. Bleiler and A. Casson, *Automorphisms of Surfaces after Nielsen and Thurston*, London Math. Soc. Student Texts vol. 9, Cambridge Univ. Press, Cambridge (1988).

[3] H. Bohr, *Et Tilbageblik*, Matematisk Tidsskrift A (1947), 1–27. English translation: *Looking backward*, Harald Bohr: Collected Mathematical Works, Vol. 1, Danish Mathematical Society (1952), xiii–xxxiv.

[4] R.F. Brown, *Wecken properties for manifolds*, Nielsen Theory and Dynamical Systems, C.K. McCord, ed., Contemp. Math. vol. 152, Amer. Math. Soc., Providence, RI (1993), 9–21.

[5] M. Culler and K. Vogtman, *Moduli of graphs and automorphisms of free groups*, Invent. Math. **84** (1986), 91–119.

[6] W. Fenchel, *Jakob Nielsen in memoriam*, Acta Math. **103** (1960), vii–xix.

[7] J. Gilman, *On the Nielsen type and the classification of the mapping-class group*, Adv. Math. **40** (1981), 68–96.

[8] M. Handel, *New proofs of some results of Nielsen*, Adv. Math. **56** (1985), 173–191.

[9] V.L. Hansen, *Jakob Nielsen (1890–1959)*, Math. Intelligencer **15**(4) (1993), 44–53.

[10] B. Jiang, *Fixed points and braids I*, Invent. Math. **75** (1984), 69–74.

[11] B. Jiang, *Fixed points and braids II*, Math. Ann. **272** (1985), 249–256.

[12] B. Jiang, *Commutativity and Wecken properties for fixed points of surfaces and* 3*-manifolds*, Topology Appl. **53** (1993), 221–228.

[13] B.J. Jiang and J.H. Guo, *Fixed points of surface diffeomorphisms*, Pacific J. Math. **160** (1993), 67–89.

[14] R.T Miller, *Geodesic laminations from Nielsen's viewpoint*, Adv. Math. **45** (1982), 189–212.

[15] J. Nielsen, *Collected Mathematical Papers*, V.L. Hansen, ed., Vol. 1 (1913–1932), Vol. 2 (1932–1955), Birkhäuser, Boston, MA (1986).

[16] W.P. Thurston, *On the geometry and dynamics of diffeomorphisms of surfaces*, Bull. Amer. Math. Soc. **19** (1988), 417–431.

This Page Intentionally Left Blank

CHAPTER 38

Heinz Hopf

Günther Frei

Département de mathématiques, Université Laval, Québec, QC, G1K 7P4, Canada
E-mail: gfrei@mat.ulaval.ca

Urs Stammbach

Mathematik, ETH-Zentrum, CH-8092 Zürich, Switzerland
E-mail: stammb@math.ethz.ch

Introduction

Without doubt, Heinz Hopf (1894–1971) was one of the most distinguished mathematicians of the twentieth century. His work is closely linked with the emergence of algebraic topology; it is most decisively thanks to his early works that this area established itself as a new and important branch of mathematics. His œuvre has influenced profoundly the evolution not only of topology but of a large part of mathematics. But Heinz Hopf was not only a gifted researcher: he was also an excellent teacher and a personality of the highest integrity. At the same time, he effervesced with charm and subtle humour. In the obituary that appeared in the organ of the IMU, Henri Cartan describes Heinz Hopf:[1]

> Ceux qui l'ont connu n'oublieront jamais sa finesse et sa douceur, alliées à une grande fermeté du caractère. Ils n'oublieront non plus le professeur ou le conférencier: Hopf n'avait pas besoin d'élever la voix pour se faire écouter; la précision de son langage ne l'empêchait pas, bien au contraire, d'éveiller l'intuition chez son auditeur; à partir de quelques constatations simples, de caractère élémentaire, il posait des problèmes neufs et les regardait sous leur différents aspects: analytique, géométrique, algébrique.

Youth (1894–1913)

Heinz Hopf's ancestors belonged to a respected and prosperous family of hop traders in Nuremberg.[2] The great-grandfather, Löb Hopf, came from Ühlfeld, a small town in Upper Franconia. He moved with his family to Nuremberg in 1852, where he was one of the first Jews to be able to acquire citizenship. The grandfather, Stephan Hopf (1826–1893),

HISTORY OF TOPOLOGY
Edited by I.M. James

Heinz Hopf (1894–1971)

became respectably wealthy as a hop wholesaler and played politically important roles in Nuremberg as *Kommerzienrat*, *Magistratsrat*, and *Landrat*. The father, Wilhelm Hopf (1861–1942), first learned brewing in Flensburg and then, in 1887, moved to Breslau after he had quarreled with his father and had had his inheritance paid out. There he joined Heinrich Kirchner's brewery. Only one year later, thanks to his considerable inheritance, he became sole owner. On May 28th, 1892, he married Elisabeth Kirchner, the elder one of Heinrich Kirchner's two daughters. In 1895, Wilhelm Hopf adopted his wife's Protestantism. In their happy marriage they had two children, Hedwig Hopf (1893–1953) and Heinz Hopf.

Heinz Hopf was born on November 19th, 1894, in Gräbschen near Breslau. Father Hopf had had a villa-like house built in the Gründerzeit[3] style, surrounded by a large garden. In 1901, Heinz entered Dr. Karl Mittelhaus' higher boys' school and, from 1904, went to the König-Wilhelm-Gymnasium in Breslau. His mathematical talent was soon recognized and found an active supporter in his teacher Bruchmann. His Abitur certificate from May 13th, 1913, says:[4]

> Mathematik: Er hat für den Gegenstand, besonders nach der algebraischen Seite hin, eine nicht gewöhnliche Begabung gezeigt.*

In the other subjects his marks were not as good. It is possible that Hopf neglected his homework at times and preferred doing sports. Throughout his life he loved sports activities. Though not tall, he was of a tough and strong constitution. In his childhood his favourite sports were swimming and tennis. Later he made regular swimming outings, rambling and mountaineering, in winter often with skis. An extended daily walk was a necessity to him, as to his father before him.

Student period (1913–1925)

In April 1913, after his Abitur, Hopf matriculated for mathematical studies at the Silesian Friedrich Wilhelms University in Breslau.[5] He attended lectures by *Adolf Kneser*, *Erhard Schmidt*, and *Rudolf Sturm* as well as by *Max Dehn* and *Ernst Steinitz* who worked at the Breslau Polytechnic at that time. Besides mathematics, Hopf also attended lectures in physics, philosophy, and psychology, subjects in which he was also interested after his studies.

One year later already, the outbreak of World War I interrupted Hopf's studies. Hopf, following the common war enthusiasm at that time, volunteered for military service. For a long time during the next four years, he fought on the Western Front as lieutenant of reserves. He was wounded twice during the war, and in 1918 he was awarded the Iron Cross (first class).

A short holiday from service in June 1917 was, according to Hopf himself, the decisive turning-point in his mathematical career. During this holiday, he attended a lecture course on set theory by Erhard Schmidt. At that time, Schmidt was treating Brouwer's theorem on the invariance of dimension under topological maps and presented the proof Brouwer had given in 1911 using the mapping degree. Hopf tells in his memoirs:[6]

> Ich war fasziniert; diese Faszination – durch die Kraft der Methode des Abbildungsgrades – hat mich nicht wieder verlassen, sondern große Teile meiner Produktion entscheidend beeinflusst. Und wenn ich heute den Gründen für diese Wirkung

* Mathematics: He has shown an extraordinary gift in this topic, especially in the algebraic direction.

> nachgehe, so sehe ich besonders zweierlei: erstens die Eindringlichkeit und mitreißende Begeisterung des Vortrages von Erhard Schmidt, und zweitens meine eigene gesteigerte Aufnahmefähigkeit während einer vierzehntägigen Unterbrechung eines langjährigen Militärdienstes.*

After the end of the war, in December 1918, Hopf was discharged from military service and resumed his interrupted studies at the University of Breslau. However, in the meantime, Erhard Schmidt had been appointed the successor of Hermann Amandus Schwarz in Berlin. This may have been the reason for Hopf not continuing his studies in Breslau. In autumn 1919, he changed to the University of Heidelberg. The reason for this choice can be simply assumed to be due to his sister who had begun studying law there a year already. Besides lectures on philosophy and psychology, Hopf attended only a few mathematical lecture courses by *Oskar Perron* and *Paul Gustav Stäckel* and furthermore a mathematical seminar.

Already in autumn 1920, Hopf decided to follow his teacher Erhard Schmidt from Breslau and to continue his studies in Berlin. This step was extraordinarily significant for his development. Since the time of Kummer, Kronecker, and Weierstrass, Berlin had been one of the leading universities in mathematics in Germany. In his scientific interests, he followed mainly Erhard Schmidt, whom he owed many ideas. Their personal relationship was based on high mutual esteem; however it was part of Schmidt's nature to maintain a certain distance. Scientifically as well as personally, Hopf was also close to the algebraist Issai Schur. Hopf attended lectures on set theory, on differential equations, and on complex analysis by Schmidt, and on number theory, elliptic functions, and invariant theory by Schur. Hopf learned much about the newest developments in topology from Schmidt, in particular about Brouwer's work and about Schmidt's own work on the Jordan curve theorem. Schmidt also made his assistant Feigl give a lecture course on Poincaré's work on *Analysis situs*.

In Schmidt's seminar, Hopf gave talks on the Clifford surface and the Clifford–Klein space problem in the winter semester of 1921/22. The topic he treated for his dissertation under Schmidt's supervision during the following years was in this area. In the first part of his doctoral thesis, Hopf proved the theorem that a simply connected complete Riemannian 3-manifold of constant curvature is globally isometric to either the Euclidean, the spherical or the hyperbolic space. The connection between local and global phenomena that emerges here also preoccupied Hopf in many of his later works. In the second part Hopf treated the relation between the *curvatura integra* of closed hyper-surfaces M in $\mathbb{R}^{n+1}$, defined as the degree of the Gauss normal map, and the indices of the zeroes of tangent vector fields on M. Hopf proved that, independently of the vector field, the sum s of the indices of the zeroes is nought for n odd, and twice the *curvatura integra* for n even, therefore in particular even. Hopf published the results of his dissertation in two separate papers in the *Mathematische Annalen* [1], [2]. He obtained his degree in February 1925. Schmidt closed his report with the remark:[7]

> Die Kühnheit der Fragestellungen verdient ebensoviel Bewunderung wie die überraschenden Ergebnisse der Lösungen. Das Schönste der Arbeit bildet aber doch die

* I was fascinated; this fascination – of the power of the method of the mapping degree – has not left me since, but has influenced great parts of my production. And when I look for the cause of this effect, I see particularly two things: firstly, Schmidt's vividness and enthusiasm in his talk, and secondly my own increased receptivity during a fortnight off many years of military service.

> Methode der Beweisführung, die, was bei Arbeiten in diesem Gebiet besonders selten ist, abstrakt und in jedem Schritte kontrollierbar vorgeht und kraft der Abstraktion in gleich hohem Maße Reichtum der anschaulich-geometrischen Fantasie an den Tag legt.*

For the thesis Schmidt pleaded for the rare predicate *eximium*. In the final result – Hopf was examined in mathematics by Schmidt and Bieberbach, in physics by Planck, and in philosophy by Wertheimer – Hopf got the predicate *summa cum laude*.

The period as a Privatdozent (1925–1931)

Immediately after his doctorate, Hopf turned to his *Habilitation*. On Schmidt's advice he intensively studied Brouwer's publications – in his memoirs he remarks:[8] *that was tough work* – and Hadamard's paper *Note sur quelques applications sur l'indice de Kronecker*. From this emerged the two papers *Abbildungsklassen n-dimensionaler Mannigfaltigkeiten* [5] (mapping classes of n-dimensional manifolds) and *Vektorfelder n-dimensionaler Mannigfaltigkeiten* [6] (vector fields on n-dimensional manifolds), which Hopf submitted as his Habilitation thesis. Hopf could already talk on these results during the annual conference of the *Deutsche Mathematiker-Vereinigung* in September 1925, only half a year after his doctorate. In the second of these papers, the famous theorem appears which says that the sum of the indices of the singularities of a vector field on a closed orientable manifold is an invariant of the manifold, namely, the Euler characteristic. The first proof of this had been given by Lefschetz a short time before; Hopf presented a new proof based on a complicated induction argument on the dimension. His Habilitation took place in autumn 1926. In his reference[9], Schmidt stated that according to him, Hopf should be seen as *already standing in the first rank of German mathematicians*.

Hopf spent the academic year which lies between Doctorate and Habilitation in Göttingen. The University of Göttingen was a most active centre of mathematical research of international prestige at that time. Besides *David Hilbert*, *Richard Courant*, *Carl Runge* and others, also a number of prominent Privatdozenten worked there, among them *Paul Bernays* and *Emmy Noether*, and many important mathematicians from all over the world came as long- or short-term guests.

In his memoirs, Hopf begins his description of the Göttingen year as follows:[10]

> Mein wichtigstes Erlebnis in Göttingen war es, dass ich dort Paul Alexandroff traf. Aus diesem Zusammentreffen wurde bald eine Freundschaft; nicht nur Topologie, und nicht nur Mathematik wurden diskutiert; es war eine glückliche und auch eine sehr fröhliche Zeit, die nicht auf Göttingen beschränkt war, sondern sich auf vielen gemeinsamen Reisen fort setzte.†

A deep friendship started here, lasting until Hopf's death.

---

* The boldness of the questions deserves as much admiration as the surprising results of the solutions. But the most beautiful thing in his thesis is the method of proving, which is, particularly rarely found in works in that area, abstract and comprehensible in every step, and which, due to the abstractness, shows equally clearly the richness of the concrete geometrical imagination.

† My most important experience in Göttingen was to meet Paul Alexandrov. This meeting soon became friendship; not only topology, not only mathematics was discussed; it was a fortunate and also a very happy time, not restricted to Göttingen but continued on many joint journeys.

In every year since 1923 Alexandrov had been a guest in Göttingen. Though a little younger than Hopf, he was already regarded as one of the leaders in point-set topology. Just at that time he began to apply algebraic methods to set-theoretic questions. One of the tools developed for that purpose was to associate with a covering of a topological space its nerve, i.e. a simplicial complex describing the combinatorics of the covering. The nerve can be viewed as an abstract algebraic approximation of the space, and by means of the notions of algebraic topology, results on the topological space itself can be derived. In the following years this notion would be applied extensively and led to a great number of new and interesting results in point-set topology.

Hopf commented on Alexandrov's idea of nerves in his memoirs:[11]

> Sie war der erste erfolgreiche Versuch, algebraische Betrachtungen in die mengentheoretische Topologie einzuführen – sehr zum Missfallen mancher Verfechter der "Reinheit der Methode". [...] Mich selbst hat damals die Erkenntnis, eine wie große Rolle die Algebra in den topologischen Problemen spielt, in entscheidender und bleibender Weise beeinflusst.*

Their common interests brought Alexandrov and Hopf together from the beginning, and thanks to Alexandrov, Hopf was warmly received in the Göttingen circle around Courant, Hilbert, and Emmy Noether.

Another important idea concerning the link between topology and algebra emerging then for the first time was due to Emmy Noether. Alexandrov tells in his autobiographic notes[12] how Emmy Noether explained the idea of Betti groups of a complex after a dinner at Brouwer's house in Blaricum in December 1925. She suggested introducing the factor group of cycles modulo boundaries and replacing the complicated numerical study of Betti numbers by the algebraic investigation of these groups. The idea was adopted at once, in particular by Vietoris, Alexandrov, and Hopf, and soon became popular in algebraic topology. It not only made it possible to give concise and simple definitions of the basic notions of algebraic topology but also prompted a wholly new view of the methods of algebraic topology. This shows up very clearly in the example of Hopf's paper *Eine Verallgemeinerung der Euler–Poincaréschen Formel* [12] which appeared in 1928. Here, for the first time Hopf explicitly uses homology groups. He shows how the Euler–Poincaré formula, interpreted in this new framework, can be generalised easily to yield a simple and lucid proof of the Lefschetz fixed point formula.

Alexandrov and Hopf, soon later joined by Otto Neugebauer, formed a closely linked group of friends in Göttingen, and they called themselves a *two-dimensional simplex*. They spent a lot of their spare time together on walks or in the attendant Klie's swimming pool at the river Leine. That was where often the whole mathematical department of Göttingen met, together with the guests who were present. Alexandrov writes in his Memories of Heinz Hopf:[13] *many a discussion, mathematical and nonmathematical, took place* there, and *many mathematical ideas were born there*. In the semester vacation, Alexandrov, Hopf, and Neugebauer made several major journeys, for example, to Brittany in France, to the Pyrenees and to Corsica after the end of the summer semester 1926, and later, in May 1927, to Upper Bavaria, after the summer semester 1927 to the Dauphiné, to Cassis near Marseille and to Portofino on the Italian Riviera.

---

* This was the first successful attempt to introduce the algebraic study of point-set topology – much to the dislike of some supporters of the "Pureness of method". [...] I was influenced in a decisive and persistent manner by the insight into the importance of the role of algebra in topological problems.

Following the first journey to France, Hopf returned to Berlin in the autumn of 1926. During the following winter semester he gave a course on *combinatorial topology* which encompassed many of the most recent results. Hopf's student Erika Pannwitz compiled this into a script. He regularly informed Alexandrov in Moscow about the contents of his lectures, who in turn discussed the new results in his circle. During that time, Hopf was thoroughly occupied with the analysis of the mapping degree and the question of how far the mapping degree determines the homotopy class of a map between manifolds. The two resulting papers [11], [14] appeared in the Mathematische Annalen in 1928 and 1929, respectively.

Hopf and Alexandrov spent the academic year 1927/28 together at Princeton University on a Rockefeller fellowship. In his final report on this stay Hopf says that he went to lectures by Lefschetz and Alexander on *Analysis Situs* and that, on the Princeton mathematicians' request, he also gave a number of talks himself on his own works and those of other European mathematicians. He continues:[14]

> Jedoch erblicke ich in diesen äußeren Ereignissen keineswegs den wichtigsten Teil meines Princetoner Aufenthaltes. Diesen sehe ich vielmehr in den häufigen zwanglosen Gesprächen mit [den] Professor[en] Alexander, Lefschetz und Veblen, sowie mit Professor P. Alexandroff aus Moskau, mit dem ich in Princeton täglich zusammen war und alle frisch empfangenen wissenschaftlichen Eindrücke und Gedanken sofort gründlich durchsprach.*

During his time at Princeton, Hopf worked primarily on the homology of manifolds. The discovery of the intersection ring of a manifold goes back to that time. Hopf showed that the homology of a manifold becomes a ring when one views the intersection of two cycles as a product. This *intersection ring* behaves contravariantly – this was completely surprising to Hopf – in that a map between manifolds corresponds to the so-called inverse homomorphism between the intersection rings. Only a few years later could this contravariant behaviour be explained completely with the introduction of cohomology: the intersection ring can be identified with the cohomology ring of a manifold by means of Poincaré duality. Hopf's paper [16], where he expands the theory of the intersection ring, appeared in the Journal für reine und angewandte Mathematik in 1930.

Having returned from Princeton, Hopf and Alexandrov again spent the summer of 1928 in Göttingen. During that time, Courant proposed that they should write a book on topology for the Springer-Verlag series *Grundlehren der mathematischen Wissenschaften.* They agreed but did not suspect that this joint work should take up so much of their time during the following seven years. They outlined a comprehensive exposition of the whole area of point-set and algebraic topology. For this extended programme a single volume would certainly not suffice, as they soon realized. They planned a second and later even a third volume, but only the first one was finished. It was published in 1935. The difficulties of that time and eventually the outbreak of World War II contributed to the discontinuation of the project. Also it is clear that the very rapid development of algebraic topology in the 30s would have made the task very difficult, even in ideal circumstances.

---

* But I do not regard these circumstantial events as the most important part of my Princeton stay, but much more the frequent informal talks with Professors Alexander, Lefschetz, and Veblen, as well as with Professor P. Alexandrov from Moscow whom I met daily in Princeton and with whom I discussed all the freshly absorbed scientific impressions and thought thoroughly.

In October 1928, Hopf married *Anja von Mickwitz* (1891–1967). Anja von Mickwitz came from a German–Baltic family of pastors blessed with many children.[15] She had trained in St. Petersburg to become a teacher. After the First World War she moved to Northern Germany and later worked as a private teacher in Berlin. After the wedding the couple spent a few days in Hopf's parents' holiday home in Hain in the Sudeten Mountains. It was there where Hopf often retired for rambling and skiing.

In December 1929 Hopf was offered by Princeton University an assistant professorship, but he turned it down. In the autumn of the following year, the Eidgenössische Technische Hochschule in Zürich asked in a diplomatically worded letter whether Hopf would accept an offer to succeed Hermann Weyl. This inquiry was in part induced by a statement by Issai Schur who had written about Hopf to Zürich:[16]

> Hopf ist ein ganz vorzüglicher Dozent, ein Mathematiker von starkem Temperament und starker Wirkung, ein Muster seiner Disziplin, der auch auf anderen Gebieten vorzüglich geschult ist. [...] Was seine Art, seine Bildung und liebenswürdiges Wesen betrifft, wünsche ich Ihnen keinen besseren Kollegen.*

After a short consultation with Courant, Hopf replied:[17]

> [...] eine Berufung in die Schweiz nach der schönen Stadt Zürich würde mich sehr locken und ehren, zumal auf einen so berühmten Lehrstuhl. Ich erkläre mich daher grundsätzlich bereit, eine eventuelle Wahl anzunehmen.†

While Hopf was waiting for a reply from Zürich, he received another offer from Freiburg i. Br., where Lothar Heffter's chair was vacant. But Hopf maintained his decision for Zürich, and before the end of the year he was elected Full Professor for Mathematics at the Eidgenössische Technische Hochschule. In the beginning of April 1931 he took up his new position.

Zürich before World War II (1931–1939)

Only a few days before he wrote his acceptance in the autumn of 1930, Hopf had finished his manuscript *Über die Abbildungen der dreidimensionalen Sphäre auf die Kugelfläche* (*On the maps from a three-dimensional sphere to the two-dimensional sphere*) in his parents' holiday home in the Sudeten Mountains and submitted it to the Mathematische Annalen. We will take a deeper look into this work because it is particularly typical for Hopf's methods of working and thinking; it is illustrated beautifully by Eckmann's words:[18]

> [Hopf hat] mit sicherem Instinkt tiefe Probleme ausgewählt und reifen lassen, um dann jeweils in einem Wurf eine Lösung zu geben, in der neue Gedanken und Methoden zu Tage traten.‡

Since Brouwer, the theory of the mapping degree had developed from the theory of maps between spheres of the same dimension. In 1925, Hopf was able to prove that the

* Hopf is an excellent lecturer, a mathematician of strong temperament and strong influence, a leading example in his discipline, and he is also well-educated in other subjects. [...] I cannot wish you a better colleague in respect to his manners, his education and his sympathetic nature.

† [...] A call to Switzerland, to the beautiful city of Zürich, could indeed tempt and honour me, particularly to such a famous chair. I therefore declare that I am in principle willing to accept such an offer.

‡ Hopf selected deep problems with an unerring instinct and let them mature. Then he presented in one piece a solution that showed new thoughts and methods.

homotopy class is characterized by the mapping degree. Along these lines it seemed natural also to study maps between spheres of different dimensions. At that time nothing was known about this apart from the simple fact that all continuous maps $f : \mathbb{S}^n \to \mathbb{S}^m$ with $n < m$ are contractible to a point. Since all maps between spheres of different dimensions induce the zero homomorphism in the homology groups, they cannot be distinguished homologically. Hopf considered the simplest case of maps from the three-dimensional to the two-dimensional sphere. To tackle this problem, Hopf introduced a new invariant which was later named after him. Hopf defined it as the linking number of the pre-images of two different points of $\mathbb{S}^2$ in $\mathbb{S}^3$. In an involved proof with the aid of simplicial approximation, he could show that this linking number is independent of the choice of the two points and that it is an invariant of the homotopy class. To exhibit a *topologically essential* map it was therefore sufficient to construct a map with a nonvanishing Hopf invariant. Due to his knowledge of classical projective geometry, Hopf could describe such a map – nowadays known as the Hopf fibration: he embeds $\mathbb{S}^3$ as the unit sphere into the four-dimensional space $\mathbb{R}^4$. Then he regards $\mathbb{R}^4$ as $\mathbb{C}^2$, maps a point P of $\mathbb{S}^3$ to the line that joins P with 0 and interprets it as a point in $P^1(\mathbb{C})$. Finally, he uses that $P^1(\mathbb{C})$ is homeomorphic to $\mathbb{S}^2$. This map can be described in a simple and completely explicit way using coordinates. But it is more difficult to prove that it is essential, i.e. not homotopic to the trivial map. Hopf derived this using his invariant.

Looking at the explicit form of this map, it can easily be inferred that the pre-image of any point in $\mathbb{S}^2$ is a great circle in $\mathbb{S}^3$. Hopf explains the fact that the linking number of any two of these great circles is ± 1 as follows – this quotation at the same time illustrates Hopf's graphic and clear formulation:[19]

> Eine dreidimensionale und eine zweidimensionale Ebene durch den Mittelpunkt der $\mathbb{S}^3$ schneiden sich, wenn die letztere nicht ganz in der ersteren liegt, in einer Geraden durch den Mittelpunkt; dies bedeutet, wenn man zu den Schnitten mit der $\mathbb{S}^3$ übergeht: eine zweidimensionale Großkugel und ein Großkreis schneiden sich, wenn der Kreis nicht auf der Kugel verläuft, in zwei zueinander diametralen Punkten; folglich wird die Hälfte H einer Großkugel von jedem Großkreis, der fremd zu dem Rand von H ist und daher nicht auf der Großkugel verläuft, stets in genau einem Punkt geschnitten; da es zu jedem Großkreis (unendlich viele) von ihm berandete Hälften von Großkugeln gibt, folgt hieraus: je zwei zueinander fremde Großkreise der $\mathbb{S}^3$ sind miteinander verschlungen, und zwar ist ihre Verschlingungszahl ± 1.*

The last statement follows from the fact that the linking number of the two great circles is equal to the intersection number of one great-circle with the great hemisphere bounded by the other.

Hopf would generalise the methods and results of this work to maps between spheres of higher dimensions a few years later (in 1935). Surprisingly, a connection to the theory of real algebras showed up here.

---

* A three-dimensional and a two-dimensional plane through the center of $\mathbb{S}^3$ intersect in a line through the center unless the latter lies completely in the former; this means when passing to the intersection with $\mathbb{S}^3$: a two-dimensional great sphere and a great circle intersect in two antipodal points unless the circle lies inside the sphere; therefore the hemisphere H of the great sphere intersects every great circle which is disjoint from the boundary of H and which is therefore not part of the great sphere, in precisely one point; since for every great circle there are (infinitly many) great hemispheres bounded by it, it follows: any two disjoint great circles in $\mathbb{S}^3$ are intertwined; their linking number is ± 1.

The result described above invoked several important lines of development in algebraic topology; it stimulated algebraic topology frequently and for years, and prompted further developments. Examples that should be mentioned are the homotopy groups (Hurewicz 1935), in particular those of spheres, the notion of fibration (Seifert 1932), the conclusion of the study of the Hopf invariant one maps (Adams 1958/60), and their various relations to the theory of real algebras.

Almost at the same time as the work on maps of spheres, he wrote with his Berlin student Willi Rinow the joint work *Über den Begriff der vollständigen differentialgeometrischen Fläche* (On the notion of complete differential geometric surfaces), which appeared in Commentarii Mathematici Helvetici in 1931 [20]. Here they prove the equivalence of different definitions of completeness. In particular it is proved that completeness in the sense of point-set topology is equivalent to the property that on a geodesic ray starting at any point one can go arbitrarily far (*auf jedem geodätischen Strahl [...] [kann man] von jedem Punkt aus jede Strecke abtragen*). Here again the fascination is apparent which Hopf felt for the link between local and global properties.

From the 4th to 12th of September, the International Congress of Mathematicians took place in Zürich. Hopf was one of the organizers, being a member of an executive committee of five. At the congress itself Hopf talked about results he had achieved together with his Berlin student Erika Pannwitz. Soon later the paper *Über stetige Deformationen von Komplexen in sich* (On continuous deformations of complexes into themselves) appeared in the Mathematische Annalen [25]. The question was here which complexes can be deformed into proper subcomplexes of themselves.

Alexandrov took the occasion of the International Congress for a longer stay in Zürich. This provided a welcome opportunity to pursue the book project further in direct cooperation. Until now, Hopf and Alexandrov had been posting each other the manuscripts for correction and criticism. Now much could simply be settled in direct discussions. They did not have another such opportunity before September 1935, when Hopf participated in the *Erste Internationale Konferenz über Topologie* in Moscow, run by Alexandrov.[20] Almost all important topologists of that time were present. In the talks, a number of new ideas and results were presented for the first time. Alexander, Gordon, and Kolmogorov, for example, talked about their independently obtained results on cohomology. A surprising fact – also for Hopf – was that a product could be defined for cohomology classes of arbitrary complexes and spaces, which gave cohomology a ring structure. Hopf had thought that such a product structure – as he had given for homology in his definition of the intersection ring – could only exist for manifolds, due to the local Euclidicity.

In Moscow, Hopf himself reported on his student Eduard Stiefel's results on the question of whether there are m continuous vector fields on an n-dimensional manifold. Stiefel had introduced *characteristic classes* in his work, which could be used to answer the question. After Hopf's talk, Whitney remarked in a discussion that a part of Stiefel's results were also contained in his note *Sphere Spaces* that had just appeared. Subsequently, it became common in algebraic topology to name the characteristic classes after Stiefel *and* Whitney.

Hopf had travelled to Moscow together with his wife; their plan was to spent several weeks with Alexandrov and Kolmogorov on the Crimean Peninsula after the congress. During this stay in Gaspra near Jalta, where Alexandrov had been several times in his holidays, the joint book was completed; they read the last corrections and finally edited the preface.

After his return to Switzerland, Hopf took part in a conference on topology in Geneva. Elie Cartan talked on his result that the homology of the classical compact simple Lie groups is the homology of a product of spheres of odd dimensions. Afterwards Cartan posed the question of whether this is also true for the exceptional groups and, hence, because of the structure theorems, in general for all compact Lie groups.

Hopf was able to solve this problem in an utterly new way in the course of the following years. The resulting paper appeared in 1941 in the *Annals of Mathematics* [40]. It had been submitted to the *Compositio Mathematica* in August 1939, but because of the war this journal had to be discontinued. Like Elie Cartan, Hopf was not satisfied by a proof by direct verification because such a proof[21] *contained no general reasons for the truth of the theorem.* He therefore tried to determine the homology of a compact Lie group using only general properties. For this goal he introduced so-called Γ-manifolds; these are manifolds on which a continuous but not necessarily associative product is defined. So, group spaces are particular examples of Γ-manifolds. Hopf then showed that the intersection ring of a Γ-manifold is isomorphic to a product of intersection rings of spheres of odd dimension. It is essential for his proof that the product structure of the manifold induces a coproduct in homology via the inverse homomorphism. The intersection ring therefore becomes – as it is called nowadays – a Hopf algebra. The result then follows because the algebra structure of a Hopf algebra is very restricted. In the case of a Γ-manifold one gets an exterior algebra with generators in odd dimensions.

In that way, Hopf solved the problem in unexpectedly great generality. At that time, no further examples of Γ-manifolds were known other than Lie groups and spheres of odd dimension, but Hopf had recognized the pivotal role this concept plays in the study of Lie groups. Starting with Hopf's work, the theory of H-spaces was intensively developed in algebraic topology in the following years. The insight that the existence of a coproduct in an algebra posed severe restrictions on its structure is the beginning of the theory of Hopf algebras which, as should be shown later, plays an important role not only in algebraic topology but also in many other areas.

At the end of his paper, Hopf briefly referred to the algebra structure in the homology of a Lie group introduced by Pontryagin a short time before and conjectured that the proof could also be done using the Pontryagin algebra. This was soon later proved by Hopf's student Hans Samelson.

Instead of the Hopf intersection ring, one today considers the cohomology ring which carries a Hopf algebra structure due to the product on the manifold. This point of view was already well known at that time. But Hopf preferred – here as well as in other works – to use homology; apparently the more geometric cycles were nearer to his way of thinking than the cocycles which are better suited for computations.[22] During the first ten years of his Zürich time Hopf published, besides the voluminous book with Alexandrov, about twenty papers; several among these papers have influenced the further development of algebraic topology in a pioneering way. He achieved this in addition to all the duties of his professorship at the ETH. From the very beginning, Hopf devoted himself to extensive lecturing, comprising various areas of mathematics and also many elementary courses. His lectures were regarded – as before in Berlin – as excellent by his students, and they were known to be extraordinarily clear and gripping. He always succeeded in making his audience ask, think, and work together with him. Therefore it is not astonishing that he attracted a number of excellent students who wanted to work for a Diploma or Doctorate under his supervision. In particular his PhD students always found him attentive – discussions took place in his house in Zollikon where tea and cakes were served afterwards – and he supported them generously with ideas and consultation.

Zürich during World War II (1939–1945)

In addition to the great demands of his work, Hopf was also under great psychological stress during his first years in Zürich, due to the political situation in Germany. His parents still lived in Breslau. Being a Jew, his father was exposed to increasing pressure by the Nazis. Until 1939, Hopf could visit his parents regularly and get his own impression of the situation.[23] This made him try to get them an immigration permit for Switzerland. Although his application was approved, the planned journey had to be deferred because the father became seriously ill. The outbreak of the war made it impossible to pursue the plans later. Hopf's father died in Breslau in 1942.

In Zürich and at his place in Zollikon, Hopf, together with his wife, tried to provide aid for persecuted people from Germany. His cousin Ludwig Hopf was a regular guest in Zollikon. Ludwig Hopf had been professor at the Technische Hochschule Aachen and lost his position in 1934 because of the Nazi laws. In 1938 he managed to flee from Germany. He became lecturer at Trinity College, Dublin, but died only a few months later. After the loss of his position, Issai Schur, Hopf's former teacher in Berlin, spent some time at the ETH with a teaching post before he could emigrate to Palestine in 1939.[24] Hopf tried to help many other persecuted people financially or by supporting their cause outside Germany. For his student Hans Samelson he managed to organize a position in Princeton in July 1940, when Switzerland was already almost surrounded by the Axis occupied territory.

In Princeton, people were worrying about Heinz Hopf's future fate, and Solomon Lefschetz sent him an invitation to Princeton in November 1940. Hopf replied in his letter from January 1st, 1941:[25]

> Das ist sehr nett von Ihnen, und ich bin Ihnen für diese Anfrage sehr dankbar. [...] Aber [wir halten] es aus prinzipiellen Gründen für richtiger, das Schiff nicht zu verlassen, solange trotz des Sturmes doch noch eine Möglichkeit besteht, dass es nicht untergeht.*

Two years later, circumstances had deteriorated for Hopf in such a way that he had to apply for the Swiss citizenship.[26] Until then, he had not planned to take this step before the end of the war in order not to be considered an opportunist. But in March 1943, he received a notice that his property had been confiscated by the German authorities. Soon afterwards, the German consulate general in Zürich refused to extend his *Heimatschein*, and he was threatened with the loss of his German citizenship unless he moved back to the area of the German Reich. Hopf's plea for Swiss citizenship was approved by the *Bürgergemeindeversammlung* of Zollikon in the same year.

Whereas until the outbreak of the war there were some, albeit censored, connections with Germany, they were disrupted completely when the war began. Scientific contacts with France, Great Britain, and America were strongly restricted and even the formerly frequent correspondence with Alexandrov in Moscow ended around Christmas 1940. In spite of this isolation which Hopf found oppressing, he and his students continued to publish works of the highest standards.

* That is very kind, and I am very grateful that you offered this. [...] However, for reasons of principle we consider it better not to leave the ship as long as despite of the tempest, there is a possibility that it will not sink.

A compilation of works from the school of algebraic topology Hopf had founded in Zürich can be found in his report *Bericht über einige neue Ergebnisse in der algebraischen Topologie* [42], which was meant to be a contribution to the Festschrift for Brouwer's sixtieth birthday in 1941 but then could only be published in 1946 because of the war. In that paper Hopf first reports on Eduard Stiefel's study of the existence of systems of continuous tangent vector fields on real projective spaces, in particular of the parallelizability of these spaces (1941); then on Beno Eckmann's results on the homotopic properties of fibred spaces from which statements about the parallelizability of spheres follow, among others (1941). He continued with an account of Werner Gysin's work on the homology of fibred spaces with fibre a sphere (1941), of Hans Samelson's work on the homology of spaces on which Lie groups act, from which a general reason for the special structure of the homology of compact Lie groups could be derived (1941), and finally of Alexandre Preissmann's results on the fundamental group of closed Riemannian manifolds of negative curvature (1942/43). His own contribution was about the question in how far the fundamental group of a connected complex determines the second Betti group. Using ideas which also appeared in the works of Samelson and Preissmann in other contexts, Hopf considered the quotient of the second Betti group of a complex with fundamental group G modulo homology classes whose cap-products with arbitrary one-dimensional cohomology classes vanish. He showed that this quotient only depended on the fundamental group G. As a conclusion he obtained that the second Betti group is completely determined by the fundamental group if every image of a two-dimensional sphere is null-homologous.[27]

In the sequel, these considerations led to the important paper *Fundamentalgruppe und zweite Bettische Gruppe* [44]. Here Hopf weakened the preconditions and investigated the quotient of the second Betti group modulo the homology classes which contain continuous images of spheres. Hopf showed that also this quotient depends only on the fundamental group. In particular, the theorem followed that the second Betti group is completely determined by the fundamental group if every image of a two-dimensional sphere is null-homotopic. Starting with a free presentation F/R of the fundamental group G, Hopf gave an explicit description of this group, namely $[F, F] \cap R/[F, R]$. The abelian group associated with G by this formula had already arisen in works of Issai Schur's on projective representations which had appeared just after the turn of the century. But Hopf does not seem to have noted this connection with the Schur multiplier in the beginning.

The paper *Fundamentalgruppe und zweite Bettische Gruppe* is legitimately regarded to be the beginning of homological algebra. It opened the way for the definition of the homology and cohomology of a group. This step was made independently at different places shortly after the paper had become known: in the USA in the circle around Samuel Eilenberg and Saunders MacLane, in Switzerland by Heinz Hopf and Beno Eckmann and in the Netherlands by Hopf's former student Hans Freudenthal. Hopf's own paper on this topic *Über die Bettischen Gruppen, die zu einer beliebigen Gruppe gehören* [49] appeared in 1944/45. Following his work mentioned above, he had conjectured that its main result could be generalised to higher dimensions. Hurewicz had shown in the thirties that the homology groups of an aspherical connected space are completely determined by the fundamental group G. Hopf's first work contained the algebraic details of this proposition for the second homology group. In the comprehensive sequel he now showed how one can treat higher dimensions similarly. From today's point of view, one can describe his purely algebraic construction as a G-free resolution of $\mathbb{Z}$. For Hopf, it arose as the algebraic analogue of the complex of the universal covering $\widetilde{X}$ of an aspherical space X with fundamental group G (whose existence was proved by Eilenberg and MacLane at the same time and independently of Hopf). The Betti groups were then defined as the homology groups of the complex which resulted from the free resolution by trivializing the G-action (tensor product with $\mathbb{Z}$ over G). Hurewicz's result mentioned above corresponds in this context to the fact that the Betti groups do not depend on the choice of a particular free resolution of $\mathbb{Z}$. By his procedure, Hopf assigned Betti groups to a given group in a purely algebraic way; so the basis for the (co-)homology theory of groups and of homological algebra was established. In the following years, this theory earned broad appreciation only slowly, possibly due to the necessary complex algebraic machinery. But gradually it became an indispensable tool in quite a large range of mathematical areas.

At the same time, Hopf occupied himself with the theory of ends of open spaces already developed in 1931 by his student Hans Freudenthal. Hopf considered spaces which are regular coverings of a compact space. He showed that there are only three possibilities: either the number of ends is one or two, or the set of ends has the cardinality of the continuum. If the (finitely generated) group G is realized as the group of deck transformations of the covering $\widetilde{X}$ of a compact space X, then the number of ends of $\widetilde{X}$ is an invariant of G. Hopf posed the question about the group-theoretic significance of the number of ends and solved the case of two ends completely: a group G has two ends if and only if it contains an infinite cyclic subgroup of finite index. Hopf did not succeed in characterising the other cases completely. The theory of ends was taken up again soon later by Hopf's student Ernst Specker. In the end of the sixties, the theories of ends of a group played a key role in Stallings' solution for the problem of groups of cohomological dimension one.[28]

Zürich after World War II (1945–1971)

After the end of the war, the interrupted scientific relations were gradually reestablished. First, Hopf tried to contact Alexandrov. The latter had come through the war safe and sound. His house in the vicinity of Moscow was slightly damaged by grenade splinters but he was able to spend the time in safety east of Moscow, although in rather primitive conditions.

In the period just after the war, Hopf tried to help his relatives and friends on the other side of the Swiss border to the best of his ability. On the one hand, the support consisted of the bare necessities of life, for the terrible shortage could only be alleviated by food parcels from foreign countries. But Hopf's assistance was also aimed at helping reestablishing mathematical life in Germany. Already in August 1946, Hopf was guest at the Mathematical Research Institute at Oberwolfach in the Black Forest, which was founded by Wilhelm Süss after the war.

In the period from October 1946 until March 1947, Hopf went to America.[29] On the journey he first went to Paris where he spent a few days with Jean Leray and participated in a meeting of the Académie Française. Then he boarded a ship in Le Havre for New York. After arrival he spent a few weeks in New York, the remaining time primarily in Princeton. He met many old friends for the first time after a long period, Courant, Friedrichs, Stoker, Neugebauer in New York, Veblen, Alexander, Lefschetz in Princeton. In Princeton he shared a flat with J.H.C. Whitehead.

At New York University, Hopf gave talks on *Selected Topics in Geometry* with much success. In his audience there were some young mathematicians who would become well-known later, e.g., Louis Nirenberg, Peter Lax, and Anneli Leopold (later Lax). Peter Lax worked up his notes of these lectures; they were published posthumously in volume 1000 of the Springer Lecture Notes in Mathematics, together with a lecture *Differential Geometry in the Large* Hopf gave in Stanford in 1956 written up by John W. Gray.

During his stay in Princeton he received several calls and offers from American universities, among them Harvard University, the California Institute of Technology in Pasadena, and Princeton University. According to the reports Hopf sent home, Courant, whom Hopf consulted, had a very high opinion of the offer from Harvard: *only few mathematicians have such good positions*. In the sequel, Hopf could slightly improve his position at the ETH, and after thorough consideration he decided to stay in Zürich.

On the occasion of the bicentenary of the University of Princeton he was awarded with the title of honorary doctor. In a letter to his wife he tells:

> Dinge gehen vor im Mond.* [...] So werden Illusionen zerstört. Was habe ich mir doch als unschuldiger Jüngling unter einem Ehrendoktor, noch dazu von Princeton, für einen klugen und weisen Mann vorgestellt. Aber es freut mich natürlich gewaltig.†

He also tells in detail about the celebration itself in his letter and adds humorously:

> Mein 'gown' war mir zwar zu weit, aber glücklicherweise nicht zu lang, so dass ich nicht daraufgetreten bin.‡

Towards the end of his American visit Hopf made some major journeys. He visited Harvard University, Brown University and the University of North Carolina at Chapel Hill. Finally he undertook an extended journey with lectures at Toronto, Chicago, Bloomington, and Ann Arbor. In the beginning of April, he returned from New York to Zürich by aeroplane.

As before the war, Hopf made it possible for many, especially younger mathematicians, to stay in Zürich. For example, in 1948 the young Hirzebruch was hospitably welcomed at Hopf's home in Zollikon. Also Tits and Nirenberg both spent several months at the ETH in Zürich as post-doctoral visitors.

Towards the end of the forties, mathematics in Europe came to life again. Hopf was now invited frequently, often as the principal speaker at congresses and conferences. In 1947 he travelled to Paris, in 1949 to a major conference on topology in Oberwolfach, in 1950 to Brussels. On the occasion of Severi's seventieth birthday in Rome in the same year, he met Paul Alexandrov for the first time after a long period. In 1953, he was Henry Whitehead's guest in Oxford while the conference for "Young Topologists" was held there.

In the winter semester 1955/56, Hopf again went to America, this time together with his wife. The ship voyage to New York was followed by an excellently organized lecture trip across the whole American country, and during a longer stay at Stanford University, Hopf gave a lecture course on *Differential Geometry in the Large* which appeared posthumously in the volume of the Lecture Notes in Mathematics mentioned above.

Thanks to his high scientific and personal reputation, Hopf was elected President of the International Mathematical Union from 1955 until 1958. Since Alexandrov worked in

---

* *Dinge gehen vor im Mond / die das Kalb selbst nicht gewohnt / ...* (things happen in the moon that even the calf is not used to) is the beginning of the humorous poem "Mondendinge" by C. Morgenstern.

† This is how you lose your illusions. What an intelligent and wise man I imagined a honorary doctor, and in particular one from Princeton, must be when I was an innocent youth. But of course it makes me tremendously happy.

‡ Although my gown was too wide, it was fortunately not too long so that I did not step on it.

the executive committee too, the two old friends now met more frequently, at the International Congresses of Mathematicians in Amsterdam (1954), in Edinburgh (1958), in Stockholm (1962), and finally in Moscow (1966). When René Thom received the Fields Medal in Edinburgh, Hopf was asked to give the laudatory speech. Now the honours accumulated: the Princeton honorary doctorate was followed by five more: Freiburg i. Br. (1957), Manchester (1958), Sorbonne at Paris (1964), Brussels (1964), and Lausanne (1965). From the University of Göttingen he received the Gauß–Weber Medal in 1955, and from the Academy of Sciences of the USSR in Moscow the Lobachevsky award in 1967. In 1958, he became member of the Deutsche Akademie der Naturforscher Leopoldina in Halle. Furthermore, he was corresponding member of the Heidelberg Akademie der Wissenschaften (1949) and of the Akademie der Wissenschaften in Göttingen (1966), honorary member of the London Mathematical Society (1956), of the Schweizerische Mathematische Gesellschaft (1957), of the American Academy of Arts and Sciences (1961) as well as foreign member of the National Academy of Sciences of the USA (1957) and the Accademia Nazionale dei Lincei (1962).

On the occasion of his seventieth birthday in 1964, the *Selecta Heinz Hopf* appeared, in which the 19 most important of his over 70 articles were published and in that way made accessible to the mathematical world more easily. One year later, on 6th July 1965, Hopf gave his retirement lecture at the ETH as part of a major celebration.

In personal life he could not escape sorrows. In the year 1959, he had to be operated on for a stomach ulcer and had to recover at home for an extended period. Around the middle of the sixties, his wife Anja fell very ill. They had planned a journey together to the International Congress of 1966 in Moscow and to Alexandrov, but Hopf had to go alone. Anja died in February 1967. Hopf did not recover from this blow. Symptoms of a geriatric disease appeared which confined him to his house. He died in hospital on 3rd June 1971.

Sources

Heinz Hopf's scientific papers are in the Wissenschaftshistorische Sammlungen der ETH-Bibliothek Zürich, under the reference Hs. 620–622. In the same collection, under the reference Hs. 160, there are copies of fifty letters by Heinz Hopf to Paul Alexandrov; the originals are kept in the Russian Academy of Sciences in Moscow. A comprehensive collection of Heinz Hopf's letters to his wife during his America visit in 1946/47 is property of Dr. Elisabeth Ettlinger-Lachmann, Heinz Hopf's niece.

Several obituaries on Heinz Hopf have appeared. We want to mention in particular:

Alexandrov, P. (1976), *Einige Erinnerungen an Heinz Hopf*, Jber. Dt. Math.-Verein. **78**, 113–125.

Behnke, H. and Hirzebruch, F. (1972), *In memoriam Heinz Hopf*, Math. Ann. **196**, 1–7.

Cartan, H. (1972), *Heinz Hopf (1894–1971)*, International Mathematical Union (IMU), 7–10.

Eckmann, B. (1971), *Zum Gedenken an Heinz Hopf*, Neue Züricher Zeitung, June 18th, reprint in L'Enseignement Mathématique **18** (1972), 105–112.

Hilton, P.J. (1972), *Heinz Hopf*, Bull. London Math. Soc. **4**, 202–217.

Samelson, H. (1976), *Zum wissenschaftlichen Werk von Heinz Hopf*, Jber. Dt. Math.-Verein. **78**, 126–146.

Acknowledgements

We wish to thank the employees of the Wissenschaftshistorische Sammlungen der ETH-Bibliothek Zürich, particularly the former director, Dr. Beat Glaus, who supported us continually by word and deed; Dr. Elisabeth Ettlinger-Lachmann for many informative talks on Heinz Hopf and for allowing us to look at the letters he wrote to his wife from America; and Professor Beno Eckmann for retelling many personal memories of Heinz Hopf and for numerous comments about the development of mathematics as he had experienced it as one of Hopf's students. Finally we thank Tilman Bauer for the excellent translation of our German text into English.

Notes

References to Heinz Hopf's publications are marked by square brackets; the numbering corresponds to the one in *Selecta Heinz Hopf*, Springer-Verlag, 1964.

[1] International Mathematical Union (IMU), 1972, pp. 7–10.

[2] The details of the Hopfs' family tree are drawn from Arnd Müller: Geschichte der Juden in Nürnberg 1146–1945, Selbstverlag der Stadtbibliothek Nürnberg 1968. Together with much more information on the circumstances of Heinz Hopf's life, they were kindly put at our disposal by Elisabeth Ettlinger-Lachmann, Heinz Hopf's niece.

[3] In German history the period from the 1870/71 war between France and Germany until about the end of the century is known as the "Gründerzeit". These early years of German unification were marked by periods of exceptional economic growth. The architecture of the time was designed to show to the world the newly accumulated wealth of the owner.

[4] Wissenschaftshistorische Sammlungen der ETH-Bibliothek Zürich, Hs. 622:7.

[5] Wissenschaftshistorische Sammlungen der ETH-Bibliothek Zürich, Hs. 622:9–10.

[6] Hopf, H., Einige persönliche Erinnerungen aus der Vorgeschichte der heutigen Topologie. CBRM Bruxelles (1966), 9–20. From these autobiographic notes we also extract many more remarks on Hopf's personal and mathematical development.

[7] Promotion Heinz Hopf, Gutachten und Prüfungsprotokoll. Cited in Biermann, K.-R., Die Mathematik und ihre Dozenten an der Berliner Universität (1810–1920), Akademie-Verlag, Berlin (1973), pp. 335–338.

[8] Hopf, H., Einige persönliche Erinnerungen aus der Vorgeschichte der heutigen Topologie, see note 6.

[9] Reference by Erhard Schmidt on Heinz Hopf's Habilitation thesis, partially quoted in Biermann, K.-R., Die Mathematik und ihre Dozenten an der Berliner Universität (1810–1920), Akademie-Verlag, Berlin (1973), p. 205.

[10] Hopf, H., Einige persönliche Erinnerungen aus der Vorgeschichte der heutigen Topologie, see note 6. – There is much information on Paul Alexandrov and Heinz Hopf, in: Alexandroff, P., Einige Erinnerungen an Heinz Hopf, Jahresbericht der Deutschen Mathematiker-Vereinigung 78 (1976), 113–125 as well as in Alexandrov, P.S., Pages from an autobiography, Russian Mathematical Surveys 35 (1980), 315–358. See also Frei, G. and Stammbach, U., Correspondence between Alexandrov and Hopf, in: Proceedings of the International Topology Conference, dedicated to P.S. Alexandrov's 100th birthday, Phasis Publishing House, Moscow (1996), pp. xxiii–xxxviii.

[11] Hopf, H., Einige persönliche Erinnerungen aus der Vorgeschichte der heutigen Topologie, see note 6.

[12] Alexandrov, P.S., Pages from an autobiography, Russian Math. Surveys 34:6 (1979), p. 324.

[13] Alexandroff, P., Einige Erinnerungen an Heinz Hopf, Jber. Dt. Math.-Verein. 78 (1976), p. 113.

[14] Bericht über die Zeit meines "Fellowships" der Internationalen Education Board vom 1. Oktober 1927 bis 1. Juni 1928, Sketch. Wissenschaftshistorische Sammlungen der ETH-Bibliothek Zürich, Hs. 622:47.

[15] The information on Anja von Mickwitz was taken from a typoscript which Leopold Ettlinger put generously at our disposal.

[16] Letter by Issai Schur to George Polya from June 30th, 1930. Archiv des Schweizerischen Schulrates, Korrespondenz des Schweizerischen Schulrates, Akten. Wissenschaftshistorische Sammlungen der ETH-Bibliothek, Zürich.

[17] Letter by Heinz Hopf to Schulratspräsident Rohn from September 30th, 1930. Archiv des Schweizerischen Schulrates, Korrespondenz des Schweizerischen Schulrates, Akten. Wissenschaftshistorische Sammlungen der ETH-Bibliothek, Zürich.

[18]Eckmann, B., Zum Gedenken an Heinz Hopf, Neue Zürcher Zeitung, June 18th (1971); reprint in: L'Enseignement Mathématique 18 (1972), 105–112.

[19]The quotation is taken from [18], see also Selecta Heinz Hopf, p. 54.

[20]The source for the information on the *Erste Internationale Konferenz über Topologie* is Hopf, H., Einige persönliche Erinnerungen aus der Vorgeschichte der heutigen Topologie, see note 6; Alexandroff, P.: Einige Erinnerungen an Heinz Hopf, Jber. Dt. Math.-Verein. 78 (1976), 113–146, as well as the correspondence between P. Alexandrov and H. Hopf, Wissenschaftshistorische Sammlungen der ETH-Bibliothek, Zürich, Hs. 621:15–146 and Hs. 160.

[21]The quotation is from the work [40], see also Selecta Heinz Hopf, p. 125.

[22]Cf. Hilton, P. J., A brief, subjective history of homology and homotopy theory in this century, Math. Magazine 61 (1988), 282–291.

[23]The journeys are mentioned in Hopf's letters to Alexandrov. For Hopf's correspondence with the Swiss authorities in that respect, consult Wissenschaftshistorische Sammlungen der ETH-Bibliothek Zürich, Hs. 622:43–44.

[24]Brauer, A., Gedenkrede auf Issai Schur, in: Issai Schur Gesammelte Abhandlungen, Vol. I, Springer (1973), pp. v–xiv.

[25]Draft of a letter to Solomon Lefschetz, dated 1/41. Wissenschaftshistorische Sammlungen der ETH-Bibliothek Zürich, Hs. 92:289.

[26]Cf. the correspondence in Wissenschaftshistorische Sammlungen der ETH-Bibliothek Zürich, Hs. 622:45–75.

[27]Hopf had summarized these results in a note *Relations between the fundamental group and the second Betti group* in 1940 and sent it to America where they were thoroughly studied by Eilenberg and MacLane on the occasion of a topology conference at the University of Michigan in Ann Arbor. See MacLane, S., Group extensions for 45 years, Math. Intelligencer 10 (1988), No. 2, pp. 29–35.

[28]Stallings, J., On torsion free groups with infinitely many ends, Ann. Math. 88 (1968), 312–334.

[29]The information on Hopf's time in America stems from letters Hopf wrote to his wife from America; the three following quotations are from letters written on December 20th, 1946, January 13th, and February 22nd, 1947. We thank Elisabeth Ettlinger for allowing us to read these letters.

CHAPTER 39

Hans Freudenthal

17 September 1905 – 13 October 1990

W.T. van Est

Aert van Neslaan 628, 2341 HV Oegstgeest, The Netherlands

Writing a short biography of Hans Freudenthal in a Handbook of Topology creates something of a predicament.

Of course it is quite appropriate that he should be recorded here for his contributions to Topology. On the other hand he was a man of erudition and widespread interests both in the mathematical sphere as well as in other fields like literature, philosophy, history, mathematics education. In practical life there was likewise a broad spectrum of activities he was involved in, and one had the impression that any one of these activities – be it administrative, scientific, literary or one of the many miscellaneous ones – was equally important to him.

In the diploma of one of the honorary doctorates that were bestowed on him, this broadness of activities and interests was duly acknowledged.[1]

In the last part of his life he concentrated on mathematics education – a point of interest since his student days – and he devoted his energies to this up to his last day.

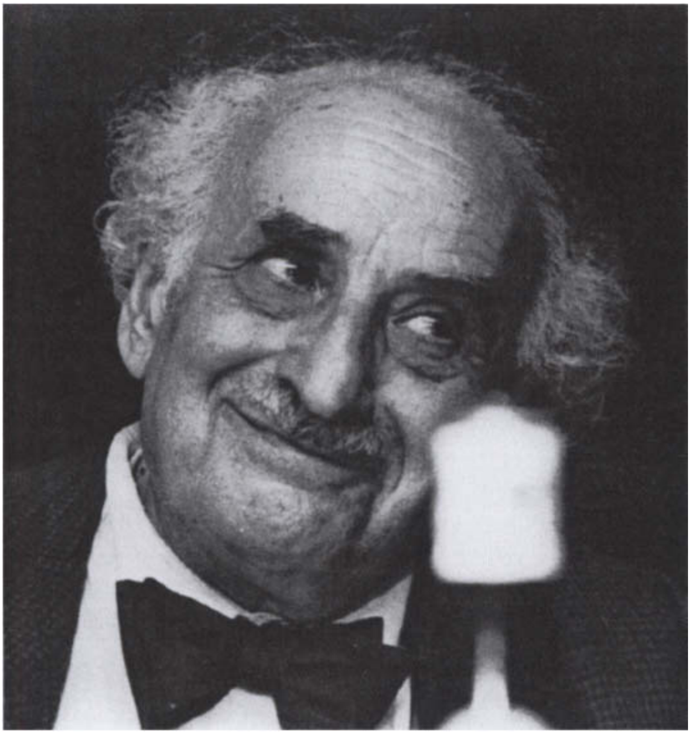
Hans Freudenthal was born on 17 September 1905 as the son of the teacher *Joseph Freudenthal* and his wife *Elsbeth* née *Ehmann*, at Luckenwalde, a small city some forty miles south of Berlin.

Having passed through the Reformrealgymnasium of his hometown, he registered as a student at the university of Berlin in 1923.[2] Here he found an extremely stimulating atmosphere with *L. Bieberbach, R. von Mises, E. Schmidt* and *I. Schur* as professors and among the staff of 'Privatdozenten' the young *H. Hopf, K. Löwner* and *J. von Neumann.* In addition there was the active club MAPHA (Mathematisch Physikalische Arbeitsgemeinschaft), in which *G. Feigl*, an assistant of Schmidt, played a leading role. Later on

[1] The diploma of the Humboldt Universität at Berlin (November 1960) states that apart from the outstanding work in mathematics the honorary doctorate has been conferred "... sowie in Anerkennung seiner vielseitigen Bemühungen Probleme der modernen Kultur mathematisch zu durchdringen...".

[2] *Berlin 1923–1930. Studienerinnerungen von Hans Freudenthal*, DMV-Tagung, 21–25 September 1987, Sonderdruck, W. de Gruyter, Berlin.

HISTORY OF TOPOLOGY
Edited by I.M. James



Hans Freudenthal (1905–1990)

Freudenthal expressed more than once his indebtedness to Feigl, who had been to him and many a student a wise counselor.

In physics the names of *Max Planck* and *Albert Einstein* (among others) lent lustre to the faculty, and of course the young student picked his share of their lectures too.[3]

Among the visitors that passed by, in particular *L.E.J. Brouwer* should be mentioned, who from January to March 1927 gave a series of lectures on Intuitionism.[4] This brought Freudenthal into closer contact with Brouwer and Intuitionism – some familiarity with these ideas he had already gained from an analysis course by Löwner. This first contact turned out to be decisive for his later career.

In any case for the time being Freudenthal continued his studies. He spent the summer semester of 1927 in Paris, where he heard among others *J. Hadamard, G. Julia* and *E. Picard*. Once returned to Berlin, where he got a small job as 'Hilfsassistent', he set out to work on a thesis.

The choice of topology as a subject was due to the overall influence of Erhard Schmidt and the stimulating lectures of Heinz Hopf. As a matter of fact Schmidt has been one of the first people[5] who had studied and understood papers by Brouwer – hard as they were to read! – and who had been disseminating over the years the knowledge of these. For example, Heinz Hopf recounts [Hopf 1964] that his first contact with topology came about by a lecture of Schmidt on Brouwer's proof of the invariance of dimension at the university of Breslau in 1917. So in Berlin, where Schmidt had moved to in 1920, he stimulated the interest in topology among the younger people, and Hopf, who had moved over with Schmidt, all the more contributed to this interest.

In any case when Freudenthal made his first steps, he found Hopf as a guide, and ever since he always considered himself a student of Hopf. From that period on dates their friendship which was maintained by a lifelong correspondence when, a few years later, their ways parted.

Hopf at the time was pondering about the construction of a $S^3 \to S^2$. The 'natural candidate' – the Hopf map – was already at his disposal, but the proof of noncontractibility was lacking.[6] For a while Freudenthal too was involved in the problem, but then he turned to the theme of the ends a space, and in particular of a topological group. In 1930 he defended a thesis on the subject [HF 2].[7] Very soon after he was attracted by Brouwer to Amsterdam as an assistent, and a new period in his career began.

In Amsterdam Freudenthal came to know *W. Hurewicz*, another assistant of Brouwer.[8] Although Brouwer at the time was not so actively interested in topology anymore, the fact of being his assistant was for both a mighty spur in their research and teaching.[9]

As for Hurewicz, his Amsterdam period gave birth, apart from various interesting papers, to his theory of homotopy groups [Hurewicz, 1935, 1936]. His emigration to the US in

3 One should recall that at the time students had a great freedom to take the courses they liked. Mathematics students often took physics courses as well.

4 A radical constructivistic conception of mathematics; see [Brouwer, 1913].

5 Another one was J.W. Alexander.

6 In April 1928, when at Princeton, Hopf in a letter to Freudenthal sketched a proof of the noncontractibility.

7 The doctoral degree was formally conferred in October 1931.

8 Hurewicz came to Amsterdam on a Rockefeller grant for the year 1927–1928, and became an assistant to Brouwer thereafter.

9 Both had been appointed 'privaatdocent', a formal (nonremunerated) position that gave the right to lecture on topics of their own choice.

1936 was of course a great loss for Amsterdam.[10]

Freudenthal published on various subjects. Making a rough incomplete clustering up to 1941, when he was suspended from duties (like so many others), one may distinguish three main categories, to wit: Topological and Lie groups [HF 13,14,42,44], Algebraic Topology proper [HF 9,10,25,30,31,32] and Linear Analysis [HF 16,17,18,19,20].

On the whole the period from 1930 up to 1941 was a relatively happy one for Freudenthal. He got settled in the Netherlands and in 1932 he married *Susanna J.C. Lutter*. The marriage was blessed with four children.

In the sphere of professional activities he became de facto the first managing editor of *Compositio Mathematica*, the journal founded in 1934 on Brouwer's initiative. Freudenthal always felt much attached to the journal, and after the war he took action to resuscitate the journal.

Furthermore, when in 1937 the chairs of geometry and analysis fell vacant, geometry and algebra were assigned to *A. Heyting* in the position of reader, whereas Freudenthal was made responsible for the courses in analysis, pro forma in a position of curator (of the library and the cabinet of mathematical models). Of course this was not a regular teaching position, but the economic crisis of the thirties still enforced budget cuts on all levels.

To both mathematicians this gave room to introduce certain innovations. For example, for second year students Heyting taught a course on 'Modern Algebra' based on the book by van der Waerden, which at the time was still recent.

Freudenthal set up a five year course in analysis. The first two years treated calculus, more variables, differential forms and Stokes' theorem, ending up with Lebesgue integration. Comparing it to the usual analysis courses in the first two years, the subject matter in itself was not new, except perhaps for differential forms and Lebesgue integration. But it was rather the treatment of the subject with an extensive use of metric topology and linear algebra that made the difference.

The subsequent courses were on 'Functions of One Complex Variable', including Riemann surfaces and elliptic functions, 'Differential Equations' (ordinary and partial) inspired by the book of *Frank* and *von Mises* (a modernized version of the classical *Riemann–Weber*), and finally 'Linear Analysis' with the basics of Banach spaces, spectral theory in Hilbert space, unitary semi-groups (Stone–von Neumann), ergodic theorems, and almost periodic functions on groups. Of course not all chapters were taught every year, the later ones were taught alternately.

At the time this was an utterly modern integrated analysis course, and as such it made quite an impression on the students.

Apart from the 'assigned' courses both Heyting and Freudenthal found time to teach special topics such as 'Intuitionism' (Heyting) and 'Combinatorial Topology' (Freudenthal). All in all this added up for both of them to a considerable teaching load, but on the other hand, classes being much smaller than nowadays, it was bearable.

Freudenthal's style of lecturing was quite vivid, perhaps not always easy to follow. But one saw how mathematical ideas emerged, and this was certainly not the least merit of his lectures.

The relatively care-free period came abruptly to an end by the German occupation. Being a Jew, Freudenthal, like so many others, was suspended from duties in 1941. The future

---

[10] In retrospect one might say that his emigration spared him the ordeal that would have been in store for him during the war.

looked ominous, and even more so when the systematic deportation of Jews began. However, since Mrs. Freudenthal was non-jewish, there was, for the time being, no immediate danger.

Despite the harassing circumstances in the period 1941–1945 – occasional arrests by the German police, work camp detention, a period of hiding out – he managed to pursue his mathematical and other interests whenever circumstances permitted. Among the rare gratifying events, if any, were the theses of two young mathematicians at the university of Groningen that complemented some of his work [de Groot, 1942; van Heemert, 1943].

At irregular intervals the correspondence with Hopf was maintained. Fruits of this difficult period are [HF 47] and [HF 48]; both papers deal with questions addressed by Hopf earlier.

Furthermore he came out with some literary work and a deepened interest in the history of mathematics. His impressive inaugural address at the university of Utrecht [HF 50] and some later articles bear witness to this.

In this Utrecht period (1946–1980) his interest changed gradually to questions of relations between geometry and Lie groups, in particular to geometries associated to the exceptional simple Lie groups. Only in a few papers [HF 77,78,80,96] he came back to questions of topology, [HF 77] and [HF 96] being a completion of earlier results.

Apart from these, many of his postwar papers address a broader public. Elementary texts on statistics, logic, natural philosophy [HF 83,263], short articles on philosophical questions, history of mathematics and mathematics education belong to this category. Somewhat apart from these, although related to his interest in logic, stands his book *LINCOS, design of a language for cosmic intercourse* [HF 190].[11]

In the last years of his career his interest focused practically entirely on questions of mathematics education from the primary up to the secondary level. As we mentioned before, these questions were already a longtime interest of Freudenthal. Very soon after the war, (or maybe even earlier) he joined the *Werkgemeenschap voor Vernieuwing van Opvoeding en Onderwijs*, where an active group of mathematics teachers in secondary education strove for innovation in the traditional methods of teaching mathematics. It was natural that Freudenthal, by his active interest, was gladly accepted as president. One of the leading ideas in his efforts was to bring together observations on phenomena in everyday life and mathematical activity. In 1971 Freudenthal managed to found an institute for developing new methods of teaching mathematics, the *IOWO*. And with full impetus he gathered a group of young people around him who set out to develop, under his guidance, mathematics courses at various levels. These new courses and teaching methods were tried out at various schools and resulted here and there into changes of teaching methods at a national level.

The ideas attracted attention abroad too, and for some time there was a co-operation project between the Utrecht group and a group of the university of Wisconsin.

In the field of mathematics education Freudenthal left a 'school' in the sense of a group of younger people working more or less along the lines that had been traced by the master. In mathematics proper the 23 theses under his patronage were quite diverse, various of these (due to special circumstances) somewhat off the lines he had been working on himself.

[11] As the reviewer in MR **22** (1961) #9378 observed, the application of the ideas of LINCOS is perhaps rather in the communication between humans and systems with artificial intelligence.

As mentioned before, the range of Freudenthal's interests and activities was a broad one. Let us just mention the many columns and articles in non-mathematical journals and his active participation in national and international committees of various kinds. In particular the *Wiskundig Genootschap* (the Dutch mathematical society), which he served twice as president, owes him a great deal because of some of his initiatives that changed life within the Genootschap radically.

Death gently surprised Freudenthal on 13 October 1990, thereby putting an end to an immensely active life.

* * *

In conclusion we shall try to put into context and perspective some of Freudenthal's papers dealing with topological questions.

Ends and topological groups [HF 2,13,14,47,77]

All spaces to be considered in this section will be supposed to belong to the category $\mathcal{F}$ of connected, locally connected, locally compact, second countable Hausdorff spaces, unless mentioned otherwise.

In [HF 2] it is shown that any space X in $\mathcal{F}$ admits a compactification $\overline{X} \supset X$ such that

(i) X is open and dense in $\overline{X}$,

(ii) the set $E = \overline{X} - X$ of *endpoints* (or *ends*) is zero-dimensional,

(iii) X is locally connected in $\overline{X}$, i.e. any endpoint $e \in E$ admits arbitrarily small neighbourhoods U_e such that $U_e \cap X$ is connected.

The conditions (i)–(iii) guarantee the essential uniqueness of $\overline{X}$. If X is obtained from a compact space Y by leaving out a closed not necessarily connected subset A, then $\overline{X}$ is essentially obtained from Y by collapsing to a single point a_i every component A_i of A.

The points $e \in E$ are defined by descending chains of noncompact connected opens with compact boundary. A similar construction in a more general context was later used by Fox [1954].

In the case of a manifold X a construction of endpoints in terms of diverging sequences of points was given by Hopf in [1943–1944]. The Hopf paper called for an intrinsic construction of the ends of a finitely generated abstract group, and this was done in [HF 47]. For further developments see [Specker, 1950; Stallings, 1971] and [Peschke, 1990].

If X is in addition the underlying space of a topological group, then it turns out that *card* $E \leqslant 2$.

At the time this must have been a surprising result, we think. The general notion of 'topological group' had been introduced not so long ago by Leja [1927], but nothing much was known about the topology of these objects in general, apart from some scarce results on locally Euclidean groups (hence Lie groups as we now know) [Schreier, 1928; Cartan, 1928]. Furthermore there was the result by Brouwer [1910] and Leja [1928], that the more than once punctured plane was not a group manifold, a result that also followed directly from Schreiers' result on the commutativity of the fundamental group. Since a more than once punctured noncompact manifold has at least three ends, the Freudenthal result gives another very general reason for the result of Brouwer and Leja.

[HF 13] collects some simple, by now standard, facts to be used in [HF 14]. It essentially points out that the canonical factorization of a homomorphism in the category of abstract groups should in the category of (general!) topological groups take the form of a factorization into an open continuous surjection and a continuous injection (of course the terminology is somewhat different). Furthermore it proves that in the category of complete second countable Hausdorff groups any continuous surjection is open. Second countability is essential, hence the result is not quite an extension of the similar result for Banach spaces.

[HF 14] proves that any locally compact connected group with sufficiently many almost periodic functions is the direct product of some $\mathbf{R}^n$ and a compact connected group, and the latter factor is the inverse limit of a sequence of compact connected Lie groups.

A nice application of [HF 14] is [HF 77] where it is proved that a group in the category $\mathcal{F}$ with two ends is the direct product of $\mathbf{R}$ and a compact group, thereby completing the results of [HF 2] in a satisfactory way.

The results of [HF 14] were so to speak in the air; the introduction discusses related results by contemporaries.

A (not quite dispassionate) survey of various results in the theory of topological groups up to 1936 is to be found in the review paper [HF 42].

Limits and algebraic topology proper [HF 9,25,32,48]

We mention [HF 9] only as an example of a situation where the adequate tools are lacking to establish the 'good result' predicted by a correct intuition.

Recall that a classical theorem of Hopf states essentially that the homotopy classes of maps $P \to S^n$, P being an n-dimensional polyhedron, are in natural 1–1 correspondence with the elements of $H^n(P, \mathbf{Z})$. Now $H^n(P, \mathbf{Z})$ has a natural group structure, whereas for the homotopy classes there is no natural composition. However, in the cases $n = 1, 3$, S^n has a group structure and this induces a group structure on the set of maps $P \to S^n$, and thereby a composition for homotopy classes that corresponds to the addition in cohomology [Bruschlinsky, 1934]. The paper imposes by brute force a 'group structure with singularities' on S^n and sets out to show that it actually induces a 'good' composition for the homotopy classes.

In retrospect we now see that the advent of $K(\pi, n)$-spaces with group structure had to be awaited before a fully satisfactory homotopical definition of cohomology could be given.

Inverse limits of groups had occurred earlier in mathematics (e.g., [Brouwer, 1910; van Dantzig, 1930; Herbrand, 1933]), and of course examples abound in p-adics. Direct sequences had occurred. These matters are discussed in the introduction of [HF 25], and then the paper sets out to define the notions of inverse and direct limit, albeit sequential limits, and, working with a different terminology, it examines the properties of these limit procedures.

It is, we think, by this paper that the notions of inverse and direct limit acquired their formal status in mathematics, although nowadays we see them as special cases of a more general limit notion from category theory.

Whereas in the case of [HF 14] and [HF 25] one might say that the results were about to crystallize anyway, if not by Freudenthal as mediator then by somebody else, it is a different matter with the suspension paper [HF 32] establishing

$$\pi_{n+k}(S^n) \cong \pi_{n+k+1}(S^{n+1})$$

whenever $k \leqslant n - 2$, and giving results on the kernel of the suspension homomorphism, the so-called 'crude' and 'delicate' suspension theorems (J.H.C. Whitehead). Of course [Hopf, 1935] (extending [Hopf, 1931]) had just been published, and the Hurewicz homotopy theory had just been created next door so to speak, and these certainly spurred the interest in these questions. But the idea to consider, for any map $f : S^{n+k} \to S^n$ its 'suspension' $Ef : S^{n+k+1} \to S^{n+1}$ induced from $f \times 1 : S^{n+k} \times I \to S^n \times I$ by passing to the quotient spaces S^{n+k+1} of $S^{n+k} \times I$ and S^{n+1} of $S^n \times I$ gotten by identifying the ends of the cylinders to points, and to study the homomorphism of homotopy groups $E : \pi_{n+k} S^n \to \pi_{n+k+1} S^{n+1}$ resulting from this construction, was really a stroke of genius. The letter E, which is frequently employed to denote the suspension homomorphism (as in the 'EHP-sequence', see [G.W. Whitehead 1978]), derives from the German 'Einhängung' used by Freudenthal in his famous paper [HF 32].

Taking into account the utterly nontrivial results (at the time) and the hard geometric approach (bare handed so to speak) the paper reminds one of the Brouwerian papers.[12] Of course today we are much wiser and experts know how to get at the results by suitable general machinery (see [Whitehead, 1978]) or Morse theory (see, e.g., [Milnor, 1963]); the delicate results, however, are still a shade more difficult than the crude ones.

The paper is a historic landmark in that here for the first time the phenomenon of stability makes its appearance in topology.

It stood by itself quite a while,[13] untill roughly around the late forties and early fifties things began to move again in the homotopy theory of spheres.

As we mentioned before [HF 48] was written in war time, practically in 'splendid isolation'. It is concerned with the problem in what way the homology of an aspherical space is determined by the fundamental group – again a problem raised by the Hurewicz papers. The results are similar to those of [Hopf, 1944–1945], but in addition it proves that, and makes clear how, the multiplicative structure of cohomology is entirely determined by the fundamental group. The papers by Eilenberg and MacLane [1945,1947] and Eckmann [1945–1946] put rather more emphasis on cohomology. But perhaps the main difference in comparison with Hopf–Freudenthal is that the latter authors make use of the 'standard complex' (the homogeneous or inhomogeneous one) associated to a group G, supplemented by the equivariant chain homotopy theorem for maps from a free G-complex to an acyclic G-complex.

Miscellany [HF 16,20,44]

[HF 16] deals with vector lattices and more in particular with Riesz spaces. The subject goes back of course to Riesz [1928], and in 1936 it was taken up practically simultaneously by Kantorovich and Freudenthal. The main result of [HF 16] is an integral representation of the elements of a Riesz space in terms of 'idempotents', quite analogous to the spectral representation of selfadjoint operators in Hilbert space. For the further history see [Luxemburg and Zaanen, 1971].

---

[12] This goes for the readability as well; even Hopf once wrote in a letter to Freudenthal that he found it hard reading.

[13] A sequel to the paper was submitted to Compositio, and, since Compositio collapsed, was subsequently forwarded to the Annals, but it was withdrawn because it contained a mistake.

The fact that a densely defined semi-bounded Hermitian operator in Hilbert space admits a selfadjoint extension, was established by Friedrichs [1934]. [HF 20] points out a very elegant shortcut in Friedrichs' proof, and it is now to be found in the textbooks, see, e.g., Sz.-Nagy [1967].

E. Cartan [1930] and independently van der Waerden [1933], established that an abstract group isomorphism $G_1 \to G_2$ of Lie groups, with G_1 compact and simple, is a Lie isomorphism. The theorem is false for G_1 noncompact simple, as is evidenced by $Sl(2, \mathbf{C})$, where discontinuous automorphisms of $\mathbf{C}$ induce discontinuous automorphisms of $Sl(2, \mathbf{C})$. [HF 44] shows that the theorem still holds true if G_1 is a real form of a complex simple group. The result remained isolated quite a while, till Borel and Tits [1968, 1973] took up the question of abstract group isomorphisms in the framework of algebraic groups.

Acknowledgement

We acknowledge the co-operation of the *Rijksarchief* in Noord Holland Haarlem, and in particular the assistance of Drs P.J.M. Velthuys-Bechtold for making accessible to us the personal archive of Hans Freudenthal, more specifically the correspondence Hopf–Freudenthal.

Bibliography

HF-References

Taken from the full list of publications by H. Freudenthal

2 1931 *Über die Enden topologischer Räume und Gruppen*, Math. Z. **33**, 692–713.
9 1935 *Die Hopfsche Gruppe, eine topologische Begründung kombinatorischer Begriffe*, Comp. Math. **2**, 134–162.
10 1935 *Über die topologische Invarianz kombinatorischer Eigenschaften des Aussenraumes abgeschlossener Mengen*, Comp. Math. **2**, 163–176.
13 1936 *Einige Sätze über topologische Gruppen*, Ann. of Math. **37**, 46–56.
14 1936 *Topologische Gruppen mit genügend vielen fastperiodischen Funktionen*, Ann. of Math. **37**, 57–77.
16 1936 *Teilweise geordnete Moduln*, Proc. Akad. Amsterdam **39**, 641–651.
17 1936 *Eine Klasse von Ringen im Hilbertschen Raum*, Proc. Akad. Amsterdam **39**, 738–741.
18 1936 *Zur Abstraktion des Integralbegriffs*, Proc. Akad. Amsterdam **39**, 741–745.
19 1936 *Ortsoperatoren in konkreten Hilbertschen Räumen*, Proc. Akad. Amsterdam **39**, 828–831.
20 1936 *Über die Friedrichschen Fortsetzung halbbeschränkter Hermitescher Operatoren*, Proc. Akad. Amsterdam **39**, 832–833.
25 1937 *Entwicklungen von Räumen und ihren Gruppen*, Comp. Math. **4**, 145–235.
30 1937 *Alexanderscher und Gordonscher Ring und ihre Isomorphie*, Ann. of Math. **38**, 647–655.
31 1937 *Zum Hopfschen Umkehrhomomorphismus*, Ann. of Math. **38**, 837–853.
32 1937 *Über die Klassen der Sphärenabbildungen I*, Comp. Math. **5**, 299–314.
42 1940 Book review: L. Pontrjagin, *Topological Groups*, Nieuw Archief voor Wiskunde **20**, 311–316.
44 1941 *Die Topologie der Lieschen Gruppen als algebraisches Phänomen I*, Ann. of Math. **42**, 1051–1074.
47 1945 *Über die Enden diskreter Räume und Gruppen*, Comm. Math. Helvetici **17**, 1–38.
48 1946 *Der Einfluss der Fundamentalgruppe auf die Bettischen Gruppen*, Ann. of Math. **47**, 274–316.
50 1946 *5000 Jaren Internationale Wetenschap*, Inaugural address, Utrecht, Groningen, 1946.
77 1951 *La structure des groupes à deux bouts et des groupes triplements transitifs*, Proc. KNAW* A **54**, 288–294.

* Proc. KNAW = Proceedings Koninklijke Nederlands Akademie van Wetenschappen.

78 1951 *Ein Kompaktheitskriterium*, Proc. KNAW A **54**, 294–296.
80 1951 *Trennung durch stetige Funktionen in topologischen Räumen*, Proc. KNAW A **54**, 359–368.
83 1952 *Inleiding tot het Denken van A. Einstein*, Hoofdfiguren van het menselijk denken, Born, Assen, 56 p.
96 1952 *Enden und Primenden*, Fund. Math. **39**, 189–210.
190 1960 *LINCOS, Design of a Language for Cosmic Intercourse*, North-Holland, Amsterdam.
262 1965 Fifth revised and enlarged edition of 83.

Other references

Borel, A. and Tits, J. (1968), *On abstract homomorphisms of simple algebraic groups*, Proc. Coll. on Alg. Geom., Tata Inst. (1969), 75–82.

Borel, A. and Tits, J. (1973), *Homomorphismes abstraits de groupes algébriques simples*, Ann. of Math. **97**, 499–571.

Brouwer, L.E.J. (1910), *On the structure of perfect sets of points*, Proc. Akad. Amsterdam **12**, 758–794.

Brouwer, L.E.J. (1910), *Endliche kontinuierliche Gruppen II*, Math. Ann. **69**, 181–203.

Brouwer, L.E.J. (1913), *Intuitionism and formalism*, Bull. Amer. Math. Soc. **20**, 81–96.

Bruschlinsky, N. (1935), *Stetige Abbildungen und Bettische Gruppen der Dimensionszahlen 1 und 3*, Math. Ann. **109**, 525–537.

Cartan, E. (1928), *Sur les nombres de Betti des espaces de groupes*, CRAS **187**, 196–198.

Dantzig, D. van (1930), *Über topologisch homogene Kontinua*, Fund. Math. **14**, 102–125.

Eckmann, B. (1945–1946), *Der Cohomologiering einer beliebigen Gruppe*, Comm. Math. Helvetici **18**, 232–282.

Eilenberg, S. and MacLane, S. (1945), *Relations between homology and homotopy groups of spaces*, Ann. of Math. **46**, 480–509.

Eilenberg, S. and MacLane, S. (1947), *Cohomology theory in abstract groups I*, Ann. of Math. **48**, 51–78.

Fox, R.H. (1954), *Covering spaces with singularities*, Algebraic Geometry and Topology, A Symposium in honour of S. Lefschetz, Princeton Math. Series Vol. 12, Princeton Univ. Press, 1956.

Friedrichs, K. (1934), *Spektraltheorie halbbeschränkter Hermitescher Operatoren*, Math. Ann. **109**, 465–487.

Groot, J. de (1942), *Topologische Studiën*, Ph.D. Thesis, Groningen.

Heemert, A. van (1943), *De R_n-adische voortbrenging van Algemene Topologische Ruimten met toepassing op de Constructie van niet-splitsbare Continua*, Ph.D. Thesis, Groningen.

Herbrand, J. (1933), *Théorie arithmétique des corps de nombres de degré infini*, Math. Ann. **108**, 699–717.

Hopf, H. (1931), *Über die Abbildungen der dreidimensionalen Sphäre auf die Kugelfläche*, Math. Ann. **104**, 637–665.

Hopf, H. (1935), *Über die Abbildungen von Sphären auf Sphären niedriger Dimension*, Fund. Math. **25**, 427–440.

Hopf, H. (1943–1944), *Enden offener Räume und unendliche diskontinuierliche Gruppen*, Comm. Math. Helvetici **16**, 81–100.

Hopf, H. (1944–1945), *Über die Bettischen Gruppen die zu einer beliebigen Gruppe gehören*, Comm. Math. Helvetici **17**, 39–79.

Hopf, H. (1964), *Colloque de Topologie Bruxelles*, CBRM. Librairie Universitaire, Louvain, Gauthier-Villars, Paris, 1966.

Hurewicz, W. (1935), *Beiträge zur Topologie der Deformationen I, II*, Proc. Akad. Amsterdam **38**, 112–119, 521–528.

Hurewicz, W. (1936), *Beiträge zur Topologie der Deformationen III, IV*, Proc. Akad. Amsterdam **39**, 117–125, 215–224.

Kantorovich, L.V. (1936), *Sur les propriétés des espaces semi-ordonnés linéaires*, C. R. Acad. Sci. Paris **202**, 813–816.

Leja, F. (1927), *Sur la notion de groupe abstrait topologique*, Fund. Math. **9**, 37–44.

Leja, F. (1928), *Un lemme topologique et son application dans la théorie des groupes abstraits*, Fund. Math. **10**, 421–426.

Luxemburg, W.A.J. and Zaanen, A.C. (1971), *Riesz Spaces I*, North-Holland, Amsterdam.

Milnor, J. (1963), *Morse Theory*, Ann. of Math. Studies vol. 51, Princeton Univ. Press, 119.

Peschke, G. (1990), *The theory of ends*, Nieuw Archief voor Wiskunde **8**, 9–12.

Riesz, F. (1928), *Sur la décomposition des opérations fonctionnelles*, Atti Congreso Bologna **3**, 143–148.

Specker, E. (1950), *Endenverbände von Räumen und Gruppen*, Math. Ann. **122**, 167–174.

Schreier, O. (1926), *Abstrakte kontinuierliche Gruppen*, Abhn. Math.-Sem. Hamburg **4**, 15–32.
Stallings, J. (1971), *Group Theory and Three-Dimensional Manifolds*, Yale Mathematical Monographs vol. 4, Yale Univ. Press, New Haven.
Sz.-Nagy, B. (1967), *Spektraldarstellungen Linearer Transformationen des Hilbertschen Raumes*, Erg. d. Math. u. ihrer Grenzgeb, Bd. 39, Springer, Berlin.
Waerden, B.L. van der (1933), *Stetigkeitssätze für halbeinfache Liesche Gruppen*, Math. Z. **36**, 780–796.
Whitehead, G.W. (1978), *Elements of Homotopy Theory*, Graduate Texts vol. 61, Springer, Berlin, 369.

This Page Intentionally Left Blank

CHAPTER 40

Herbert Seifert*

May 27, 1907 – October 1, 1996

Dieter Puppe

Mathematisches Institut, Universität Heidelberg, Im Neuenheimer Feld 288, D-69120, Germany
E-mail: puppe@sun0.urz.uni-heidelberg.de

Herbert Seifert died October 1, 1996 in his 90-th year. He was one of the great pioneers in the field of topology which developed enormously during his lifetime. His work was an essential part of this development. He was of particular importance for the University of Heidelberg because he taught at this university as a professor for 40 years with an interruption during the Second World War. After the war he was temporarily the only one responsible for the department of mathematics and it was his influence and his policy which led to what is now the Faculty of Mathematics at Heidelberg.

Karl Johannes Herbert Seifert was born May 27, 1907 at Bernstadt in Saxony. His father was, at the end of his career, Justizamtmann, i.e. a court official of medium rank. After the family had moved to Bautzen (also in Saxony) it was there that Herbert Seifert attended the Knabenbürgerschule (primary school) and later the Oberrealschule (secondary school). His grades were good but in the beginning – even in mathematics – not outstanding. However, a mathematical exercise book from his last school years already shows not only his mathematical talent but also the precise and extremely clear style which distinguishes all his writings and lectures.

In the spring of 1926 Seifert finished secondary school with the Abitur, and in the summer term of the same year he began to study mathematics and physics at the Technische Hochschule (Technical University of) Dresden. In 1927 he attended a course on topology taught by William Threlfall, and this determined the direction of his whole life. Threlfall was the son of a niece of Robert Koch and of an Englishman, and he was a private scholar. Most of the time he taught as an unsalaried lecturer at the Technical University of Dresden until he became a professor at the University of Frankfurt am Main in 1938 as the successor of Carl Ludwig Siegel, who had emigrated to the United States of America. Threlfall owned a spacious beautiful house in Dresden, which nowadays does not exist any more. A friendship started to develop between Threlfall and Seifert (who was 20 years younger) which

*This is essentially a translation of an obituary written by the author in German and published in "Jahrbuch der Heidelberger Akademie der Wissenschaften für 1997", Heidelberg, 1998.

HISTORY OF TOPOLOGY
Edited by I.M. James

Herbert Seifert (1907–1996)

established a close personal relationship for the rest of their lives and was scientifically very fruitful. But first, in 1928/1929, Seifert went to Göttingen which at that time was the world center of mathematics and where he met some of the most important mathematicians of this century, among them David Hilbert and the topologists Paul Alexandrov and Heinz Hopf.

For the summer term of 1929 Seifert came back to Dresden and moved into Threlfall's house. On July 17, 1930 he passed the examination which entitled him to become a teacher of mathematics at a secondary school (the usual way to finish university studies in mathematics at that time), and only 4 weeks later on August 13, 1930 – he was 23 years old – he received his first doctorate (Dr. rer. techn., Doktor der Technischen Wissenschaften).

The title of his dissertation was "Konstruktion dreidimensionaler geschlossener Räume", in English translation: "Construction of 3-dimensional closed spaces", in today's terminology: ... closed manifolds". At that time the problem of classifying 3-dimensional closed manifolds with respect to homeomorphism was considered as one of the most important problems of topology. To the present day it is still unsolved. Seifert made important contributions to it, even more so in his second dissertation, to which we come back below, but the first one contains something else in addition, namely a theorem allowing him to calculate the fundamental group of a space from the corresponding groups of certain subspaces. In the literature it was for a long time often called van Kampen's theorem, although van Kampen's paper to which this refers appeared in 1933 and does not contain the usual formulation of the theorem. It is true that also Seifert formulates the theorem differently from what is usual today. So it is fair to speak of the "Theorem of Seifert and van Kampen" which is being more and more accepted.

Having completed his first doctorate Seifert was granted a scholarship from the Technical University of Dresden to enable him to continue his studies at a different university. Seifert used it to go to Leipzig in the summer of 1931. However, he usually stayed there for only part of the week. Each weekend he returned to Dresden to work with Threlfall and they regularly spent their vacation time together. As early as February 1, 1932 Seifert submitted his paper on "Topologie dreidimensionaler gefaserter Räume" (Topology of 3-dimensional fibred spaces) as a dissertation at Leipzig. Nowadays the notions of "fibred space" or "fibre space" or "fibre bundle" or "fibration" are among the most important ones in topology. The meaning of the word "fibred" has changed a little but it goes back to this second dissertation of Seifert's. Van der Waerden was the official supervisor but in reality Seifert did not need supervision. The paper was almost finished before he went to Leipzig. On March 3, 1932 he passed the oral examination for the second doctorate (Dr. phil.).

At that time Seifert and Threlfall had been working already on their textbook on topology ("Lehrbuch der Topologie", Teubner, 1934). Threlfall had introduced Seifert to topology but before long the younger one became the leader in the team. Certainly, Seifert would deny this but then Threlfall would confirm it vigorously. The preface of the book begins as follows:

> Den ersten Anlaß zur Abfassung des vorliegenden Lehrbuches gab eine Vorlesung, die der eine von uns (Threlfall) an der Technischen Hochschule Dresden gehalten hat. Aber nur ein Teil der Vorlesung ist in das Buch übernommen worden. Der Hauptinhalt ist in der Folgezeit in engem täglichem Gedankenaustausch zwischen beiden Verfassern entstanden.

(Translation: The first step towards writing this textbook was a course which one of us (Threlfall) taught at the Technical University of Dresden. But only part of the course was included in the book. The main part of its contents originated later from daily discussions between the two authors.)

This formulation is a compromise on which Seifert insisted. Originally Threlfall wanted to write (according to his diary):

> Das Buch ist aus Vorlesungen hervorgegangen, die der eine von uns dem anderen im Jahre 1927 an der Technischen Hochschule Dresden gehalten hat. Bald hat aber der Hörer so wesentlich neue Gedanken zur Ausarbeitung beigetragen und sie so von Grund auf umgestaltet, daß eher als sein Name der des ursprünglichen Verfassers auf dem Titelblatte fehlen dürfte.

(Translation: This textbook arose from a course which one of us gave to the other at the Technical University of Dresden. But soon the student contributed new ideas to such an extent and changed the presentation so fundamentally that it would be more justifiable to omit on the title page the name of the original author than his.)

The book gives an excellent account of what was known in topology at that time. It was superior in contents and in ways of presentation to other books in the field not only when it appeared but for a long time to come. It was translated into several languages, and generations of topologists in all countries of the world studied it. Even now, more than 60 years later, it is worth reading because of its lucid style and because, for some special problems, it is still the best source of information, in particular if you look not only into the main text but also into the "Anmerkungen" (Remarks) at the end.

On January 22, 1934 Seifert's Habilitation (right to teach at the university level) at the Technical University of Dresden went into effect based on his paper entitled "Verschlingungsinvarianten" (Linking invariants) and on a test lecture on "Stetige Vektorfelder" (Continuous vector fields).

At that time Seifert was already well known among German mathematicians and among topologists all over the world. So it is not surprising that in spite of his young age he got several more or less official offers from other German universities. He entered into serious negotiations with the University of Greifswald, but in the end, since the conditions of the offer were not satisfactory for him, he turned it down on September 1, 1934. At the same time he received the title of "Außerordentlicher Professor" (comparable to associate professor) at the Technical University of Dresden.

On November 5, 1935 a telegram from the Ministry of Education of the German Reich in Berlin arrived at Dresden by which Seifert was summoned to go immediately to Heidelberg and take over the duties of a full professorship in mathematics at Heidelberg University. Threlfall, in his diary, refers to this as an order. Seifert complied, arrived at Heidelberg 2 days later, and it was only afterwards that he learned the details of the situation.

Until September 30, 1935 the mathematics department of Heidelberg University was run by the (full) professors Heinrich Liebmann and Artur Rosenthal. Both of them were Jewish. During the summer term of 1935 the national-socialist student association organized a boycott of their courses. They tried to resist, Rosenthal with more energy than Liebmann. As they did not get any support from the university administration or the ministry of education they applied for premature retirement, Liebmann giving his bad health as a reason, Rosenthal expressing protest. The retirement became effective with the formal end of the summer term on September 30, 1935. In a letter of January 3, 1936 the rector of Heidelberg University stated that as a consequence of a certain law (Reichsbürgergesetz of September 15, 1935) Rosenthal lost his right to teach (Lehrbefugnis) and his status as a (retired) faculty member, both with the end of the year 1935. Although it seems that the legal situation for Liebmann was exactly the same, in his case the consequences were never drawn. On the contrary, the rector confirmed in 2 letters, the second one written after Liebmann's (natural) death on June 12, 1939, that Liebmann continued to be a (retired) faculty member of Heidelberg University. Liebmann's daughter needed this confirmation in order to obtain a scholarship. Rosenthal emigrated to USA in 1936 where he started a new academic career in 1940. His last position was that of a (full) professor at Purdue University. He died September 15, 1959. Five years earlier and retroactively effective April 1, 1949 his privileges as a retired full professor of Heidelberg University had been restored.

In November 1935, immediately after his arrival at Heidelberg, Seifert took charge of Liebmann's chair. Rosenthal was succeeded by Udo Wegner starting from the winter term 1936/37. Wegner went along with the policy of the Nazi regime, but Seifert kept his distance as much as possible. The rector once told him that this fact delayed his appointment as a professor. It was not before August 1936 that he received the official offer. This was during a stay in a hospital in Oslo, where Seifert had taken part in the International Mathematical Congress and became ill with poliomyelitis. Then it took until July 1937 before his appointment became effective. Meanwhile Liebmann's chair in the budgetary sense had been used for somebody else in a field different from mathematics. Therefore Seifert's position was only an Extraordinariat (associate professorship), although he had the personal privileges of an Ordinarius (full professor). It was not until after World War II, from 1946 on, that he filled an Ordinariat in every respect.

Under these circumstances Seifert's possibilities for working at Heidelberg were limited before the war. When the war started he was sure to be drafted to some kind of war service. In order to avoid the worst possibilities, he volunteered to work in the Luftfahrtforschungsanstalt at Braunschweig, a research institution of the German Air Force, and there in particular in the Institut für Gasdynamik (Institute for Dynamics of Compressible Fluids), whose director was Adolf Busemann. Seifert was accepted and appointed head of a department of this institute. He was on leave from Heidelberg University from the winter term 1939/40 through the winter term 1944/45.

From 1936 to 1939 Seifert and Threlfall continued their cooperation by exchanging many letters and by getting together as often as their professional duties would allow – Threlfall worked at the University of Halle in 1937 and at the University of Frankfurt/Main from 1938. As before they used to meet in Threlfall's house at Dresden and they undertook many joint holiday trips. During this time they wrote their second book, which appeared in 1938. The title is "Variationsrechnung im Großen" (Variational calculus in the large) with the subtitle "Theorie von Marston Morse". Again, as in the case of the "Lehrbuch der Topologie", they made a new part of mathematics much better accessible than it had been before. The book has a motto quoted from Kepler's "Astronomia nova" and beginning as follows: "Durissima est hodie condicio scribendi libros mathematicos". (Translation: Today it is very hard to write mathematical books.) The editor of the book series, Wilhelm Blaschke, understood this as a political allusion, which it certainly was, and wanted to remove it. But the authors insisted that the motto be printed and so it was.

As mentioned before, shortly after the beginning of World War II Seifert became head of a department of the Luftfahrtforschungsanstalt at Braunschweig. Soon he succeeded in getting Threlfall also into this department. The whole institution was considered to be important for the German war effort, but Seifert's department worked only on basic theoretical problems. The only condition was that they should be related to the dynamics of compressible fluids. Thus Seifert wrote a series of papers on differential equations, among them one on "Periodische Bewegungen mechanischer Systeme" (Periodical movements of mechanical systems), which became one of the roots of today's theory of periodic solutions of Hamiltonian systems. In Busemann's institute Seifert also had the opportunity to give courses of lectures. One of them was on general relativity and cosmology and had an influence on Seifert's teaching after the war at Heidelberg.

In 1944, because Braunschweig suffered more and more from air raids, Busemann's institute was moved to "Schloß Rust" (Rust castle) on the upper Rhine not far from the Black Forest and hence not far from Oberwolfach. The foundations of what is now the Mathematical Research Institute Oberwolfach, well known throughout the world, were laid during the last months of the war by Wilhelm Süss, and he took care that Seifert and Threlfall were among the first to work at this institute. This is where they were when the war ended.

Soon Seifert tried to return to Heidelberg. First he came for short visits and finally in November 1945 for good. So he was here when, at the beginning of 1946, the Faculty of Science of the university reopened. The American Military Government had closed the university in the spring of 1945 and fired many professors in the course of denazification. At the beginning of 1946 only 4 full professors in the whole Faculty of Science were in office. Seifert was one of them, and in February 1946 he was the only one whom the American military authorities would accept as dean. So he did this job for 4 months although he did not like it at all.

In the time to come Seifert and Threlfall (who was again at Frankfurt/Main) were very much interested in continuing their cooperation. Each tried to get the other to his place. What finally worked out was that Threlfall got an offer from Heidelberg and became the second full professor of mathematics in the winter term 1946/47. However, the expectations of further extended cooperation did not materialize. Threlfall spent part of the winter term 1946/47 in Switzerland; Seifert was invited to the Institute for Advanced Study at Princeton by Marston Morse and he went there for the winter term 1948/49; and on April 4, 1949 Threlfall died unexpectedly at the age of 60.

For the next 3 years Seifert was again the only full professor in mathematics at Heidelberg and had the main responsibility for the department, supported by the associate professor Hans Maaß and the lecturers Walter Habicht and Horst Schubert. Topology was taught and studied intensively. Seifert had a series of graduate students whom he supervised in the best possible way. Recent progress in Algebraic Topology which now mainly came from USA, France, England and the Soviet Union was systematically studied in the topology seminar which has met on Thursday afternoons ever since. International connections were re- and newly established. Some of the best mathematicians in the world came to give talks at Heidelberg.

In 1952 the department of mathematics started to grow, during some periods at a breathtaking pace. The second chair was filled with the algebraist F.K. Schmidt. Then many new professorships were created and finally new institutes were founded, so that at present several mathematical disciplines are active at Heidelberg. Topology still plays an important role.

On September 13, 1949 Herbert Seifert married Dr. Katharina Seifert, née Korn. If he had been the head of the mathematical community at Heidelberg for a long time, Mrs. Seifert was its soul. She took care that people got together not only for work. She organized parties in her house and in other places which have become legendary. She also had an important part in establishing and keeping relations with many mathematicians outside Heidelberg.

At the end of the summer term 1975 Herbert Seifert retired. After that he did not go to the university very often. Together with Mrs. Seifert he enjoyed his beautiful house on the hill near the old town of Heidelberg and above all his garden in which he did a lot of work himself. For many years a large group of friends, colleagues and former students regularly came to visit the Seiferts. It was only a short time before his death that he lost his strength. A rich life which had been important for many people ended.

Many honours were bestowed on Herbert Seifert. He was a member of the Heidelberger Akademie der Wissenschaften (Heidelberg Academy of Sciences and Letters), of the Akademie der Wissenschaften zu Göttingen and of the Accademia Mediterranea delle Scienze in Catania. He was one of very few honorary members of the Deutsche Mathematiker-Vereinigung (German Association of Mathematicians). He himself did not care much about external honours. He avoided any kind of public attention. In particular he did not allow public notice to be taken of any of his birthdays. On his 75-th birthday many people gathered to congratulate him but he insisted that there should be no speeches mentioning the birthday.

Seifert will be remembered as a great mathematician by the topologists and by many other mathematicians – by those who had the privilege of personal acquaintance he will also be remembered as an upright, sometimes stern, but always warm and lovable person.

This Page Intentionally Left Blank

Appendix: Some Dates

John Frank ADAMS 1930–1989
José ADEM 1921–1991
Pavel Sergeevich ALEKSANDROV 1896–1982
James Waddell ALEXANDER 1888–1971
Izea BERSTEIN 1926–1991
Enrico BETTI 1823–1892
R.H. BING 1914–1986
Karol BORSUK 1905–1982
Luitzen Egbertus Jan BROUWER 1881–1966
Georg CANTOR 1845–1918
Elie CARTAN 1869–1951
Eduard ČECH 1893–1960
Richard DEDEKIND 1831–1916
Max DEHN 1878–1952
Clifford Hugh DOWKER 1912–1982
James DUGUNDJI 1919–1985
Walther Franz Anton von DYCK 1856–1934
Eldon DYER 1929–1993
Charles EHRESMANN 1905–1979
Samuel EILENBERG 1913–1998
Leonhard EULER 1707–1783
Georg FEIGL 1890–1945
Jacques FELDBAU 1914–1945
Ralph Hartzler FOX 1913–1973
Hans FREUDENTHAL 1905–1990
Tudor GANEA 1922–1974
Carl Friedrich GAUSS 1777–1855
Hermann Guenther GRASSMANN 1809–1877
Victor K.A.M. GUGENHEIM 1923–1995
Felix HAUSDORFF 1868–1942
Poul HEEGARD 1871–1948
David HILBERT 1862–1943
Guy HIRSCH 1915–1993
Witold HUREWICZ 1904–1956
Heinz HOPF 1894–1971

Camille JORDAN 1838–1921
Egbertus R. van KAMPEN 1908–1942
Gustav Robert KIRCHHOFF 1824–1887
Christian Felix KLEIN 1849–1925
Hellmuth KNESER 1898–1973
Leopold KRONECKER 1823–1891
Andrei Nikolaevich KOLMOGOROV 1903–1987
Hermann KUNNETH 1892–1975
Kazimierz KURATOWSKI 1896–1980
Henri Leon LEBESGUE 1875–1941
Solomon LEFSCHETZ 1884–1972
Gottfried Wilhelm LEIBNIZ 1646–1716
Jean LERAY 1906–1998
Simon Antoine Jean L'HUILIER 1750–1840
Marius Sophus LIE 1842–1899
Johann Benedict LISTING 1808–1882
August Ferdinand MÖBIUS 1790–1868
Eliakim Hastings MOORE 1862–1932
Robert Lee MOORE 1882–1974
Harold Calvin Marston MORSE 1892–1977
Maxwell Herman Alexander NEWMAN 1897–1984
Jakob NIELSEN 1890–1959
Amalie Emmy NOETHER 1882–1935
Christos PAPAKYRIAKOPOULOS 1914–1976
Charles Emile PICARD 1856–1941
Jules Henri POINCARÉ 1858–1912
Lev Semyonovich PONTRYAGIN 1908–1988
Georges REEB 1920–1993
Kurt Werner Friedrich REIDEMEISTER 1893–1971
Georges de RHAM 1903–1990
Georg Friedrich Bernhard RIEMANN 1826–1866
Vladimir Abramovich ROKHLIN 1919–1984
Ludwig SCHLÄFLI 1814–1895
Arthur Moritz SCHOENFLIES 1853–1928
Herbert SEIFERT 1907–1996
Paul SMITH 1900–1980
Edwin Henry SPANIER 1921–1996
Norman Earl STEENROD 1910–1971
Carl Georg Christian von STAUDT 1798–1868
Peter Guthrie TAIT 1831–1901
William THRELFALL 1888–1949
Heinrich Franz Friedrich TIETZE 1880–1964
Albert William TUCKER 1916–1995
Pavel Samuilovich URYSON 1898–1924
Alexandre-Theophile VANDERMONDE 1735–1796
Oswald VEBLEN 1880–1960
John Henry Constantine WHITEHEAD 1904–1960

Hassler WHITNEY 1907–1989
Gordon Thomas WHYBURN 1896–1982
Raymond Louis WILDER 1896–1982
Wilhelm WIRTINGER 1865–1945
Alexander ZABRODSKY 1936–1986

This Page Intentionally Left Blank

Index