SI Appendix: Phenotypic switching in gene regulatory networks

Philipp Thomas, Nikola Popović, Ramon Grima

September 3, 2014

Contents

1 Biochemical formulation of gene regulation

We consider a gene regulatory network that is composed of a set of promoters with joint state $\boldsymbol{\mathcal{G}} = (\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_{N_G})$, and the products of gene expression $\boldsymbol{\mathcal{M}} = (\mathcal{M}_1, \mathcal{M}_2, \ldots, \mathcal{M}_{N_M})$ and $\boldsymbol{\mathcal{P}} = (\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_{N_P})$, which comprise the set of all mRNA and protein species, respectively. We further assume that the promoter species are localized in some small volume, while mRNA and protein species are contained in a larger volume Ω.

First, we consider N_G effective reactions that describe mRNA transcription,

$$\mathcal{G}_j \xrightarrow{k_{q,j}^{(1)}} \mathcal{G}_j + \hat{q}_{1j}^{(1)}\mathcal{M}_1 + \ldots + \hat{q}_{N_M j}^{(1)}\mathcal{M}_{N_M}, \tag{1}$$

where j is a reaction index running from 1 to N_G, with N_G the number of promoter states. We note that $\hat{q}_{ij}^{(1)} = 1$ if mRNA $\mathcal{M}_i$ is transcribed from a promoter in state $\mathcal{G}_j$ and zero otherwise. We also note that the promoters appear in equal proportions on the left-hand sides and the right-hand sides in the above equation, and that they are hence conserved in the transcription process. Second, we consider translation as the synthesis of protein $\boldsymbol{\mathcal{P}}$ from mRNA $\boldsymbol{\mathcal{M}}$:

$$\mathcal{M}_j \xrightarrow{k_{r,j}^{(1)}} \mathcal{M}_j + \hat{r}_{1j}^{(1)}\mathcal{P}_1 + \ldots + \hat{r}_{N_P j}^{(1)}\mathcal{P}_{N_P}; \tag{2}$$

here, j is a reaction index running from 1 to N_M, where N_M is the number of mRNA species. Moreover, $\hat{r}_{ij}^{(1)} = 1$ if $\mathcal{M}_j$ is an mRNA-encoding protein $\mathcal{P}_i$.

Next, we consider the set of reactions which alter mRNA and protein concentrations post-transcriptionally and post-translationally. In order to ease notation, we summarize mRNA and protein species in the vector of gene products $\boldsymbol{\mathcal{Z}} = (\mathcal{Z}_1, \ldots, \mathcal{Z}_{N_Z}) = (\mathcal{M}_1, \ldots, \mathcal{M}_{N_M}, \mathcal{P}_1, \ldots, \mathcal{P}_{N_P})$, where $N_Z = N_M + N_P$. Such processes typically represent the largest set of reactions, and may include: degradation of gene products; post-transcriptional regulation, such as, for instance, by protein-mRNA interactions; post-translational modification, such as phosphorylation; or other types of downstream pathways. The corresponding chemical equation can be written in the generic form

$$s_{1j}^{(1)}\mathcal{Z}_1 + \ldots + s_{N_Z j}^{(1)}\mathcal{Z}_{N_Z} \xrightarrow{k_{s,j}^{(1)}} \hat{s}_{1j}^{(1)}\mathcal{Z}_1 + \ldots + \hat{s}_{N_Z j}^{(1)}\mathcal{Z}_{N_Z}, \tag{3}$$

where j is a reaction index running from 1 to T_0, $k_{s,j}^{(1)}$ is the reaction rate of the j^{th} reaction, and $s_{ij}^{(1)}$ and $\hat{s}_{ij}^{(1)}$ are the respective stoichiometric coefficients.

Then, we include transcriptional feedback in which proteins act as transcription factors by binding to the promoter,

$$t_{1j}^{(2)}\mathcal{G}_1 + \ldots + t_{N_G j}^{(2)}\mathcal{G}_{N_G} + s_{1j}^{(2)}\mathcal{P}_1 + \ldots + s_{N_P j}^{(2)}\mathcal{P}_{N_P} \xrightarrow{k_{t,j}^{(2)}} \hat{t}_{1j}^{(2)}\mathcal{G}_1 + \ldots + \hat{t}_{N_G j}^{(2)}\mathcal{G}_{N_G} + \hat{s}_{1j}^{(2)}\mathcal{P}_1 + \ldots + \hat{s}_{N_P j}^{(2)}\mathcal{P}_{N_P}. \tag{4}$$

Here, the reaction index runs from 1 to T_1 and $t_{ij}^{(2)}$, $\hat{t}_{ij}^{(2)}$ and $s_{ij}^{(2)}$, $\hat{s}_{ij}^{(2)}$ are the stoichiometric coefficients of promoter and transcription factor in the binding and unbinding reactions, respectively. Finally, we include interactions that spontaneously change the promoter state,

$$u_{1j}^{(2)}\mathcal{G}_1 + \ldots + u_{N_G j}^{(2)}\mathcal{G}_{N_G} \xrightarrow{k_{u,j}^{(2)}} \hat{u}_{1j}^{(2)}\mathcal{G}_1 + \ldots + \hat{u}_{N_G j}^{(2)}\mathcal{G}_{N_G}, \tag{5}$$

such as are induced by chromatin remodeling or DNA looping in the presence of a transcription factor. Here, j runs from 1 to T_2, $k_{u,j}^{(2)}$ is the reaction rate of the j^{th} reaction mediating transcriptional feedback, and $u_{ij}^{(2)}$ and $\hat{u}_{ij}^{(2)}$ are the stoichiometric coefficients for the promoter states. We note that, thus far, the above decomposition has been completely general, in that it has not required any specific assumptions on the underlying kinetics. We also note that, although we are only considering actively transcribed mRNA here, our approach can be readily extended to include non-coding RNA and the resulting interactions.

2 Probabilistic formulation of gene regulatory networks

A probabilistic formulation of the gene regulatory networks outlined in the previous section requires the specification of the probabilities per unit time for the above reactions to occur. Assuming well mixed conditions, the resulting model can be formulated independently of the positions of the molecules involved. The state of the network at any given time can then be described by the instantaneous promoter state

$\boldsymbol{G} = (n_{G_1}, n_{G_2}, \ldots, n_{G_{N_G}})$, the vector of mRNA molecule numbers $\boldsymbol{n}_M = (n_{M_1}, n_{M_2}, \ldots, n_{M_{N_M}})$, and the vector of protein numbers $\boldsymbol{n}_P = (n_{P_1}, n_{P_2}, \ldots, n_{P_{N_P}})$. We further assume the RNA polymerase and ribosome concentrations to be very abundant, so that they may be taken constant, and need not be modeled explicitly.

The propensities of the transcription reaction in (1) then depend only on the rate of transcription and the number of promoters in a given state. Similarly, the translational propensities in (2) are dependent on the translation rates and the number of mRNA molecules:

$$\begin{aligned} \hat{f}_j^{(1)}(\boldsymbol{G}) &= \Omega k_{q,j}^{(1)} n_{G_j}, \\ \hat{f}_j^{(1)}(\boldsymbol{n}_M) &= k_{r,j}^{(1)} n_{M_j}. \end{aligned} \tag{6}$$

We note that the rate constant in the transcriptional propensity $\hat{f}_j^{(1)}(\boldsymbol{G})$ has been rescaled by a factor of Ω, compared to the conventional notation for elementary reactions. We also remark that transcriptional and translational activity implicitly depend on the RNA polymerase concentration in the nucleus and the ribosome concentration in the cytosol, which have been absorbed into the rate constants $k_{q,j}^{(1)}$. Making use of the combined vector of gene products $\boldsymbol{n}_Z = (n_{Z_1}, n_{Z_2}, \ldots, n_{Z_{N_Z}})$, as introduced in the previous section, the propensity of protein-RNA interactions and protein-protein interactions, Eq. (3), can be formulated according to the law of mass action,

$$\hat{f}_j^{(1)}(\boldsymbol{n}_Z) = \Omega k_{s,j}^{(1)} \prod_{z=1}^{N_Z} \Omega^{-s_{zj}^{(1)}} \left(n_{Z_z}\right)_{s_{zj}^{(1)}}; \tag{7}$$

here, $(n_{Z_z})_m$ is the falling factorial (or Pochhammer symbol), which is defined as $n_{Z_z}(n_{Z_z}-1)\ldots(n_{Z_z}-m+1)$. The propensities of protein-DNA interactions, Eq. (4), are then simply proportional to the occupation of the promoter states and the protein concentrations. Similarly, for spontaneous promoter transitions, recall Eq. (5), we find

$$\begin{aligned} \hat{f}_j^{(2)}(\boldsymbol{G}, \boldsymbol{n}_Z) &= k_{s,j}^{(2)} \prod_{g=1}^{N_Z} \left(n_{G_g}\right)_{u_{gj}^{(2)}} \prod_{z=1}^{N_Z} \Omega^{-s_{1j}^{(2)}} \left(\boldsymbol{n}_{Z_z}\right)_{s_{zj}^{(2)}}, \\ \hat{f}_j^{(2)}(\boldsymbol{G}) &= k_{t,j}^{(2)} \prod_{g=1}^{N_G} \left(n_{G_g}\right)_{t_{gj}^{(1)}}. \end{aligned} \tag{8}$$

The biochemical processes described by the reactions in (1) through (5) can be divided into two groups: (i) reactions that leave the promoter state unchanged, as given in (1) through (3); and (ii) reactions that alter the promoter configuration and hence affect transcriptional activity, which are given by (4) and (5). Group (i) comprises a total of $R_1 = T_0 + N_M + N_G$ reactions, while group (ii) consists of $R_2 = T_1 + T_2$ reactions. It is worth noting that the propensities of group (i) are proportional to the cellular volume Ω, while the propensities for the reactions in group (ii) are proportional to the occupation of the promoter state $\boldsymbol{G}$, whose components only take the values 0 and 1 in the case of a single promoter. This difference can be made explicit by denoting the instantaneous gene product concentration by

$$\boldsymbol{Z} = \frac{\boldsymbol{n}_Z}{\Omega}, \tag{9}$$

and by expressing the propensities as

$$\hat{f}_j^{(i)}(\boldsymbol{G}, \boldsymbol{n}_Z) = \begin{cases} \Omega f_j^{(1)}(\boldsymbol{G}, \boldsymbol{Z}) & i = 1, \\ f_j^{(2)}(\boldsymbol{G}, \boldsymbol{Z}) & i = 2. \end{cases} \tag{10}$$

We note that $f_j^{(2)}(\boldsymbol{G},\boldsymbol{Z})$ is now the probability per unit time per unit volume for a reaction to occur, and that it is thus a true reaction rate.

We are now able to formulate the Chemical Master Equation (CME) for the general family of gene regulatory networks under consideration, which describes the time-evolution equation for the probability $\Pi(\boldsymbol{G},\boldsymbol{Z},t)$ to find the system in a particular mesoscopic state $(\boldsymbol{G},\boldsymbol{Z})$ at time t:

$$\begin{aligned}\frac{d\Pi\left(\boldsymbol{G},\boldsymbol{Z},t\right)}{dt} &= \Omega\sum_{j=1}^{R_1}\left(\prod_{z=1}^{N_Z}E^{-S_{zj}^{(1)}}-1\right)f_j^{(1)}\left(\boldsymbol{G},\boldsymbol{Z}\right)\Pi\left(\boldsymbol{G},\boldsymbol{Z},t\right)\\ &+\sum_{j=1}^{R_2}\left(\prod_{g=1}^{N_G}E^{-R_{gj}}\prod_{z=1}^{N_Z}E^{-S_{zj}^{(2)}}-1\right)f_j^{(2)}\left(\boldsymbol{G},\boldsymbol{Z}\right)\Pi\left(\boldsymbol{G},\boldsymbol{Z},t\right),\end{aligned} \tag{11}$$

where $\underline{\mathrm{R}}$ is an $N_G\times R_2$ block matrix with elements $[\underline{\mathrm{R}}]_{ij}=[\hat{t}_{ij}-t_{ij},\hat{u}_{ij}-u_{ij}]$, $\underline{\mathrm{S}}^{(2)}=\hat{s}_{ij}^{(2)}-s_{ij}^{(2)}$ is an $N_Z\times R_2$ matrix, and $\underline{\mathrm{S}}^{(1)}$ is an $N_Z\times R_1$ matrix given by

$$\underline{\mathrm{S}}^{(1)}=\left[\begin{array}{cc|c}\hat{\underline{\mathrm{q}}} & 0 & \hat{\underline{\mathrm{s}}}^{(1)}-\underline{\mathrm{s}}^{(1)}\\ 0 & \hat{\underline{\mathrm{r}}} & \end{array}\right], \tag{12}$$

with $[\hat{\underline{\mathrm{q}}}]_{ij}=\hat{q}_{ij}$, $[\hat{\underline{\mathrm{r}}}]_{ij}=\hat{r}_{ij}$, $[\hat{\underline{\mathrm{s}}}^{(1)}]_{ij}=\hat{s}_{ij}^{(1)}$ and $[\underline{\mathrm{s}}^{(1)}]_{ij}=s_{ij}^{(1)}$. The step operator $E_i^{-S_{ij}}$ is defined by its action on a function of the molecular populations as $E_i^{-S_{ij}}g(n_1,...,n_i,...,n_N)=g(n_1,...,n_i-S_{ij},...,n_N)$ [1].

2.1 Conventional Linear Noise Approximation of the Chemical Master Equation

The CME, Eq. (11), can only be solved in particular simple cases [2, 3, 4, 5]. A common way of circumventing this intractability is the use of a linear noise approximation (LNA), which can be derived by truncating van Kampen's system size expansion after the leading order term [1, 6]:

$$\frac{\boldsymbol{n}_G}{\Omega}=[\boldsymbol{G}]+\Omega^{-\frac{1}{2}}\boldsymbol{\epsilon}_G\quad\text{and}\quad\frac{\boldsymbol{n}_Z}{\Omega}=[\boldsymbol{Z}]+\Omega^{-\frac{1}{2}}\boldsymbol{\epsilon}_Z, \tag{13}$$

where the $O(\Omega^0)$-term is given by the solution of the macroscopic rate equations and the term proportional to $\Omega^{-\frac{1}{2}}$ denotes a contribution due to stochastic fluctuations. The resulting rate equations read

$$\begin{aligned}\frac{d[\boldsymbol{G}]}{dt} &= \underline{\mathrm{R}}\,\tilde{\boldsymbol{f}}^{(2)}([\boldsymbol{G}],[\boldsymbol{Z}]),\\ \frac{d[\boldsymbol{Z}]}{dt} &= \underline{\mathrm{S}}^{(1)}\boldsymbol{f}^{(1)}([\boldsymbol{G}],[\boldsymbol{Z}])+\underline{\mathrm{S}}^{(2)}\tilde{\boldsymbol{f}}^{(2)}([\boldsymbol{G}],[\boldsymbol{Z}]),\end{aligned} \tag{14}$$

where $\tilde{\boldsymbol{f}}^{(2)}([\boldsymbol{G}],[\boldsymbol{Z}])=\Omega^{-1}\boldsymbol{f}^{(2)}(\langle\boldsymbol{G}\rangle,[\boldsymbol{Z}])$ have been rescaled to represent macroscopic rate functions. The covariance matrix $\underline{\Sigma}$ now satisfies the matrix equation

$$\frac{d\,\underline{\Sigma}}{dt}=\underline{\mathrm{J}}\,\underline{\Sigma}+\underline{\Sigma}\,\underline{\mathrm{J}}^T+\frac{1}{\Omega}\big(\underline{\mathrm{D}}^{(1)}+\underline{\mathrm{D}}^{(2)}\big); \tag{15}$$

here, $\underline{\mathrm{J}}$ is the Jacobian of the rate equations in (14), while $\underline{\mathrm{D}}^{(1)}=\underline{\mathrm{S}}^{(1)}\mathrm{diag}(\boldsymbol{f}^{(1)})\,\underline{\mathrm{S}}^{(1)T}$ and

$$\underline{\mathrm{D}}^{(2)}=\left(\begin{array}{cc}\underline{\mathrm{R}}\,\mathrm{diag}(\boldsymbol{f}^{(2)})\,\underline{\mathrm{R}}^T & \underline{\mathrm{R}}\,\mathrm{diag}(\boldsymbol{f}^{(2)})\,\underline{\mathrm{S}}^{(2)T}\\ \underline{\mathrm{S}}^{(2)}\mathrm{diag}(\boldsymbol{f}^{(2)})\,\underline{\mathrm{R}}^T & \underline{\mathrm{S}}^{(2)}\mathrm{diag}(\boldsymbol{f}^{(2)})\,\underline{\mathrm{S}}^{(2)T}\end{array}\right) \tag{16}$$

denote noise matrices that correspond to the reactions in groups (i) and (ii), respectively, as defined above. (Similar partitions of the noise matrix have been analyzed in [7].) Eq. (15) can be solved algebraically at

steady state, or numerically in time course using the freely available software iNA [8]. Moreover, given the solution of Eqs. (14) and (15), the corresponding probability distributions can also be inferred via the system size expansion. The marginal distribution of gene products according to the LNA is thus given by a multivariate Gaussian distribution

$$\Pi(\boldsymbol{Z},t) = (2\pi)^{-\frac{N_Z}{2}} \det(\underline{\Sigma}_Z(t))^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} \left(\boldsymbol{Z} - [\boldsymbol{Z}](t)\right)^T \underline{\Sigma}_Z^{-1}(t) \left(\boldsymbol{Z} - [\boldsymbol{Z}](t)\right) \right\}, \tag{17}$$

where $\underline{\Sigma}_Z$ is the reduced covariance of gene products only. It is well known that Eq. (17) is centered about the solution of the deterministic rate equations and, hence, that it is clearly unimodal. It is however questionable if the ansatz in (13) is justified when considering the expression from a single promoter. This concern is in part due to the fact that, in the single-promoter case, fluctuations are treated as small perturbations of the deterministic dynamics.

In particular, a system size expansion requires that the reaction propensities must satisfy the following scaling,

$$\hat{\boldsymbol{f}}^{(1)}(n_G, n_Z) = \Omega \boldsymbol{f}^{(1)}\left(\frac{n_G}{\Omega}, \boldsymbol{Z}\right) + O(\Omega^0) \quad \text{and} \quad \hat{\boldsymbol{f}}^{(2)}(n_G, n_Z) = \Omega \tilde{\boldsymbol{f}}^{(2)}\left(\frac{n_G}{\Omega}, \boldsymbol{Z}\right) + O(\Omega^0), \tag{18}$$

in order for a deterministic limit to exist. Comparing the above expressions with the propensities given in the preceding section, we see that (18) is indeed satisfied by the reactions in group (i), but not by those in group (ii), since the propensities are proportional to the number of promoters in a particular state – which is typically a small integer at low gene dosage – in the latter case. In the following we analyze, albeit using a different technique, the respective regimes where the approximation in (17) is appropriate, and where it fails to predict qualitatively the observed distribution of proteins. In the past, only little attention has been paid to the accuracy of such distributions [9], or of the first two moments for Eq. (17) [10], in the context of gene regulation.

3 Slow promoter fluctuations

We now devise a method that allows us to obtain insight into the full probability distributions induced by expression of a single gene. Specifically, we analyze how the separation between promoter and product timescales gives rise to qualitatively different distributions. We first note that the reactions in group (ii) are the only ones that may change the promoter state. Hence, it follows that, if the reactions in group (i) are fast while those in group (ii) are slow, the promoter is a slow species, while the process of gene expression is fast. In order to account quantitatively for the resulting scale separation, we define the parameter μ as the ratio between the propensities of the reactions in groups (i) and (ii):

$$\mu = \frac{\Omega \min_j f_j^{(1)}}{\max_j f_j^{(2)}}, \tag{19}$$

where the factor of Ω is required for μ to be dimensionless, since $\boldsymbol{f}^{(1)}$ is a rate function, while $\boldsymbol{f}^{(2)}$ is a propensity. Rescaling $\boldsymbol{f}_*^{(1)} = \mu^{-1}\boldsymbol{f}^{(1)}$ and dropping the asterisk in the CME, Eq. (11), we find

$$\begin{aligned} \frac{d\Pi(\boldsymbol{G},\boldsymbol{Z},t)}{dt} = {} & \mu\Omega \sum_{j=1}^{R_1} \left(\prod_{z=1}^{N_Z} E^{-S_{zj}^{(1)}} - 1 \right) f_j^{(1)}(\boldsymbol{G},\boldsymbol{Z}) \, \Pi(\boldsymbol{G},\boldsymbol{Z},t) \\ & + \sum_{j=1}^{R_2} \left(\prod_{g=1}^{N_G} E^{-R_{gj}} \prod_{z=1}^{N_Z} E^{-S_{zj}^{(2)}} - 1 \right) f_j^{(2)}(\boldsymbol{G},\boldsymbol{Z}) \, \Pi(\boldsymbol{G},\boldsymbol{Z},t) \,. \end{aligned} \tag{20}$$

This procedure allows the fast dynamics of the products of gene expression, summarized in the vector $\boldsymbol{Z}$, to be described conditional on the state of promoter $\boldsymbol{G}$. The slow dynamics then simply evolves with the instantaneous protein concentrations replaced by their average values conditioned on the state of the promoter. We note that the first term on the right-hand side in Eq. (20) gives the evolution of expression products in group (i) reactions, while the second term describes the corresponding evolution – and, in particular, the transition between promoter states – in group (ii). We also note that this second term is conditionally independent of its products only in the absence of transcriptional feedback, i.e., if $\underline{\mathrm{S}}^{(2)} = 0$.

Using Bayes' theorem, we can write $\Pi(\boldsymbol{G}, \boldsymbol{Z}, t) = \Pi(\boldsymbol{Z}\,|\boldsymbol{G}, t)\Pi(\boldsymbol{G}, t)$, where

$$\Pi(\boldsymbol{G}, t) = \sum_{\boldsymbol{Z}} \Pi(\boldsymbol{G}, \boldsymbol{Z}, t) \tag{21}$$

is the marginal distribution of promoter states and the summation in Eq. (21) is taken over all possible gene product concentrations $\boldsymbol{Z}$. Substituting the above into Eq. (20), we obtain

$$\begin{aligned}\frac{\partial \Pi(\boldsymbol{Z}|\boldsymbol{G}, t)}{\partial t}\Pi(\boldsymbol{G}, t) + \Pi(\boldsymbol{Z}|\boldsymbol{G}, t)\frac{\partial \Pi(\boldsymbol{G}, t)}{\partial t} \\ = \mu\Pi(\boldsymbol{G}, t)\mathcal{L}_0\Pi(\boldsymbol{Z}|\boldsymbol{G}, t) + \mathcal{L}_1\Pi(\boldsymbol{Z}|\boldsymbol{G}, t)\Pi(\boldsymbol{G}, t),\end{aligned} \tag{22}$$

where

$$\mathcal{L}_0 = \Omega \sum_{j=1}^{R_1} \left(\prod_{z=1}^{N_Z} E^{-S_{zj}^{(1)}} - 1 \right) f_j^{(1)}(\boldsymbol{G}, \boldsymbol{Z}), \tag{23a}$$

$$\mathcal{L}_1 = \sum_{j=1}^{R_2} \left(\prod_{g=1}^{N_G} E^{-R_{gj}} \prod_{z=1}^{N_Z} E^{-S_{zj}^{(2)}} - 1 \right) f_j^{(2)}(\boldsymbol{G}, \boldsymbol{Z}). \tag{23b}$$

In the limit as $\mu \to \infty$ in Eq. (22), we have

$$\mathcal{L}_0\, \pi(\boldsymbol{Z}|\boldsymbol{G}) = 0, \tag{24}$$

where $\pi(\boldsymbol{Z}|\boldsymbol{G}) = \lim_{\mu\to\infty} \Pi(\boldsymbol{Z}|\boldsymbol{G}, t)$. We note that, since $\mathcal{L}_0$ is time-independent and a genuine transition matrix, Eq. (24) implies a stationary probability distribution $\pi(\boldsymbol{Z}|\boldsymbol{G})$ which can at least in principle be obtained by solving Eq. (24). Marginalizing (22) using Eq. (21), we find

$$\frac{\partial \Pi(\boldsymbol{G}, t)}{\partial t} = \left(\sum_{\boldsymbol{Z}} \mathcal{L}_1\, \pi(\boldsymbol{Z}|\boldsymbol{G}) \right) \Pi(\boldsymbol{G}, t) + O(\mu^{-1}), \tag{25}$$

where the term in brackets is a conditional average of the slow dynamics over the protein concentrations. Hence, we conclude that slow gene transitions occur with rates at which the protein numbers are replaced by their expectation conditioned on the current promoter state. Similarly, the distribution of gene expression products is a weighted sum of the probability that a product is found given a particular promoter state, times the probability of the promoter being in that state:

$$\Pi(\boldsymbol{Z}, t) = \sum_{\boldsymbol{G}} \pi(\boldsymbol{Z}\,|\boldsymbol{G})\Pi(\boldsymbol{G}, t) + O(\mu^{-1}). \tag{26}$$

In simple terms, that result can be formulated as the dynamics of gene products being enslaved by the dynamics of the promoter. In particular, the distribution of gene expression is then a superposition of the stationary distributions $\pi(\boldsymbol{Z}\,|\boldsymbol{G})$, with promoter held constant, weighted with the probabilities of the promoter being in a particular state $\Pi(\boldsymbol{G}, t)$ (as a function of time).

3.1 Binary promoter switching: exact solution

We now illustrate the theory developed above in the example of a promoter with two internal states [11]:

$$\mathcal{G} \underset{k_{\rm on}}{\overset{k_{\rm off}}{\rightleftharpoons}} \mathcal{G}^*, \tag{27a}$$

$$\mathcal{G} \xrightarrow{k_0^{(\rm on)}} \mathcal{G} + \mathcal{M}, \quad \mathcal{G}^* \xrightarrow{k_0^{(\rm off)}} \mathcal{G}^* + \mathcal{M},$$

$$\mathcal{M} \xrightarrow{k_1} \mathcal{M} + \mathcal{P}, \quad \mathcal{M} \xrightarrow{k_2} \varnothing, \quad \mathcal{P} \xrightarrow{k_3} \varnothing. \tag{27b}$$

We note that $k_0^{(\rm off)}$ represents the basal rate of transcription here, according to the convention in Eq. (6); moreover, we denote the promoter state by $\boldsymbol{G} = (n_G, n_{G^*})$ and the vector of gene product concentrations by $\boldsymbol{Z} = (M, P)$. According to the separation into groups (i) and (ii) proposed above, Eq. (27a) clearly represents the only pair of reactions affecting $\boldsymbol{G}$. Hence, we have $f_1^{(2)} = k_{\rm on}$ and $f_2^{(2)} = k_{\rm off}$, and $f_1^{(1)} = k_0 n_G$, $f_2^{(1)} = k_0^* n_{G^*}$, $f_3^{(1)} = k_2 \frac{n_M}{\Omega}$, and $f_4^{(1)} = k_1 \frac{n_P}{\Omega}$, where $\frac{n_M}{\Omega}$ and $\frac{n_P}{\Omega}$ are the instantaneous concentrations of mRNA and protein, respectively. Writing

$$p_{m,n}^{(\rm on)} = \pi\left(M = \tfrac{m}{\Omega}, P = \tfrac{n}{\Omega} \,\middle|\, n_G = 1, n_{G^*} = 0\right) \quad \text{and} \quad p_{m,n}^{(\rm off)} = \pi\left(M = \tfrac{m}{\Omega}, P = \tfrac{n}{\Omega} \,\middle|\, n_G = 0, n_{G^*} = 1\right), \tag{28}$$

we obtain two decoupled equations for the conditional probabilities of the on-state and the off-state at steady state, which are identical modulo a factor of $k_0^{(\rm on)}$ or $k_0^{(\rm off)}$:

$$\begin{aligned} &k_0^{(\rm on,off)} \Omega \left(p_{m-1,n}^{(\rm on,off)} - p_{m,n}^{(\rm on,off)} \right) + k_2 \left((m+1) p_{m+1,n}^{(\rm on,off)} - m p_{m,n}^{(\rm on)} \right) \\ &+ k_1 \left(p_{m,n-1}^{(\rm on,off)} - p_{m,n}^{(\rm on,off)} \right) + k_3 \left((n+1) p_{m,n+1}^{(\rm on,off)} - n p_{m,n}^{(\rm on,off)} \right) = 0. \end{aligned} \tag{29}$$

The above equations have been solved exactly by Bokes et al. [4], who found that the marginal distribution of mRNA is Poissonian, while the marginal protein distribution $p_n = \sum_m p_{m,n}$ can be obtained from the recursion relation

$$\begin{aligned} p_n^{(\rm on,off)} &= \Omega k_0^{(\rm on,off)} \sum_{i=0}^{n-1} \frac{\beta}{n} \frac{\beta^{n-1-i}}{(1+\lambda)_{(n-1-i)}} M\left(n-i, n-i+\lambda, -\beta\right) p_i^{(\rm on,off)}, \\ p_0^{(\rm on,off)} &= \exp\left(-\Omega k_0^{(\rm on,off)} \beta \int_0^1 ds\, M\left(1, 1+\lambda, -\beta(s-1)\right) \right), \end{aligned} \tag{30}$$

where $\beta = \frac{k_1}{k_2}$, $\lambda = \frac{k_2}{k_3}$, $(1+\lambda)_{(n-1-i)}$ denotes the falling factorial or Pochhammer symbol, as before, and $M(a,b,z)$ is Kummer's function. Noting that the on-off transitions of promoter are independent of protein levels, we can deduce the time evolution of the two promoter states from

$$\frac{\partial \Pi^{(\rm on)}}{\partial t} = k_{\rm on}\left(1 - \Pi^{(\rm on)}\right) - k_{\rm off} \Pi^{(\rm on)}; \tag{31}$$

here, $\Pi^{(\rm on)} = \Pi(n_G = 1, n_{G^*} = 0)$, and we have made use of the conservation of gene copy numbers, with $\Pi^{(\rm on)} = 1 - \Pi^{(\rm off)}$. Eq. (31) has the steady state solution $\Pi^{(\rm on)} = \frac{k_{\rm on}}{k_{\rm on}+k_{\rm off}}$, which implies $\Pi^{(\rm off)} = \frac{k_{\rm off}}{k_{\rm on}+k_{\rm off}}$. Using Eq. (26), we then conclude that the distribution of mRNA is a weighted sum of Poissonians

$$\Pi(n_M = m) = \frac{k_{\rm on}}{k_{\rm on}+k_{\rm off}} \frac{\mathrm{e}^{-\frac{k_0^{(\rm on)}}{k_2}}}{m!} \left(\frac{k_0^{(\rm on)}}{k_2} \right)^m + \frac{k_{\rm off}}{k_{\rm on}+k_{\rm off}} \frac{\mathrm{e}^{-\frac{k_0^{(\rm off)}}{k_2}}}{m!} \left(\frac{k_0^{(\rm off)}}{k_2} \right)^m. \tag{32}$$

Similarly, the protein distribution is given by

$$\Pi(n_P = n) = \frac{k_{\rm on}}{k_{\rm on}+k_{\rm off}} p_n^{(\rm on)} + \frac{k_{\rm off}}{k_{\rm on}+k_{\rm off}} p_n^{(\rm off)}, \tag{33}$$

with $p_n^{(\text{on,off})}$ as in Eq. (30). We note that the explicit form of Eq. (30) does not permit to draw easy conclusions on the overall distribution functions. In the following, we hence develop approximative techniques which will allow us to gain insight into these distributions in a more straightforward manner.

3.2 General solution using the conditional LNA

For any reasonably sized reaction network, the conditional distribution of gene products typically cannot be obtained in closed form. In particular, this is the case when bimolecular reactions such as protein-mRNA or protein-protein interactions are considered. It is, however, possible to describe the network dynamics by performing a linear noise approximation for the expression products, while retaining the particulate character of promoter fluctuations.

3.2.1 Conditional gene product distribution

We now seek to solve Eq. (24) via the LNA, which will yield a closed-form solution for the conditional distribution of gene products. Explicitly, we make the ansatz

$$\boldsymbol{Z}|\,\boldsymbol{G} = [\boldsymbol{Z}|\boldsymbol{G}] + \Omega^{-\frac{1}{2}}\boldsymbol{\epsilon}_{Z|\boldsymbol{G}}, \tag{34}$$

which separates the instantaneous protein concentrations conditioned on a particular promoter state into a deterministic component and the conditional fluctuations about it. Expanding the step operator

$$\prod_{z=1}^{N_Z} E^{-S_{zj}^{(1)}} - 1 = -\Omega^{-\frac{1}{2}}\sum_z \nabla_z S_{zj}^{(1)} + \frac{1}{2}\sum_{z,q}\nabla_z S_{zj}^{(1)} S_{qj}^{(1)}\nabla_z\nabla_q + O\big(\Omega^{-\frac{3}{2}}\big), \tag{35}$$

where $\nabla_z = \frac{\partial}{\partial(\boldsymbol{\epsilon}_{Z|\boldsymbol{G}})_z}$ and, equally, the vector of rate functions $f_j^{(1)}$ of the fast reactions, we obtain

$$f_j^{(1)}(\boldsymbol{G},\boldsymbol{Z}|\boldsymbol{G}) = f_j^{(1)}(\boldsymbol{G},[\boldsymbol{Z}|\boldsymbol{G}]) + \Omega^{-\frac{1}{2}}\sum_{z=0}^{N_Z}(\boldsymbol{\epsilon}_{Z|\boldsymbol{G}})_z\,\frac{\partial f_j^{(1)}(\boldsymbol{G},[\boldsymbol{Z}|\boldsymbol{G}])}{\partial[Z_z|\boldsymbol{G}]} + O(\Omega^{-1}). \tag{36}$$

Using the fact that the conditional distribution of protein numbers and the distribution of $\boldsymbol{\epsilon}$ are related by the inverse Jacobian determinant

$$\pi\left(\boldsymbol{Z}|\boldsymbol{G}\right) = \pi\left(\boldsymbol{\epsilon}_{Z|\boldsymbol{G}}\right)\left|\frac{\partial(\boldsymbol{Z}|\boldsymbol{G})}{\partial(\boldsymbol{\epsilon}_{Z|\boldsymbol{G}})}\right|^{-1}, \tag{37}$$

in combination with Eqs. (23a), (35) and (36) in Eq. (24), we obtain

$$0 = \mathcal{L}_0\,\pi\left(\boldsymbol{\epsilon}_{Z|\boldsymbol{G}}\right) = \left(-\Omega^{\frac{1}{2}}\nabla\cdot\underline{\mathrm{S}}^{(1)}\boldsymbol{f}^{(1)}(\boldsymbol{G},[\boldsymbol{Z}|\boldsymbol{G}]) - \nabla^T\underline{\mathrm{J}}_{\boldsymbol{G}}\,\boldsymbol{\epsilon}_{Z|\boldsymbol{G}} + \frac{1}{2}\nabla^T\underline{\mathrm{D}}_{\boldsymbol{G}}\nabla\right)\pi\left(\boldsymbol{\epsilon}_{Z|\boldsymbol{G}}\right) + O\big(\Omega^{-\frac{1}{2}}\big), \tag{38}$$

where $[\underline{\mathrm{J}}_{\boldsymbol{G}}]_{ij} = \sum_\alpha S_{i\alpha}^{(1)}\partial_{[Z_i|\boldsymbol{G}]}f_\alpha^{(1)}(\boldsymbol{G},[\boldsymbol{Z}|\boldsymbol{G}])$ and $[\underline{\mathrm{D}}_{\boldsymbol{G}}]_{ij} = \sum_\alpha S_{i\alpha}^{(1)}S_{j\alpha}^{(1)}f_\alpha^{(1)}(\boldsymbol{G},[\boldsymbol{Z}|\boldsymbol{G}])$. The above equation can be solved in a straightforward manner in terms of successive powers of $\Omega^{\frac{1}{2}}$. To leading order ($\Omega^{\frac{1}{2}}$), we have

$$0 = \underline{\mathrm{S}}^{(1)}\boldsymbol{f}^{(1)}(\boldsymbol{G},[\boldsymbol{Z}|\boldsymbol{G}]), \tag{39}$$

the solution of which is simply the steady state of the rate equations for the group (i) reactions, with the promoter state held constant. The next order Ω^0 is given by a Fokker-Planck Equation with linear coefficients, which determines the conditional fluctuations

$$0 = \left(-\nabla^T\underline{\mathrm{J}}_{\boldsymbol{G}}\,\boldsymbol{\epsilon}_{Z|\boldsymbol{G}} + \frac{1}{2}\nabla^T\underline{\mathrm{D}}_{\boldsymbol{G}}\nabla\right)\pi\left(\boldsymbol{\epsilon}_{Z|\boldsymbol{G}}\right). \tag{40}$$

We note that the coefficients in Eq. (40) depend parametrically on the solution of Eq. (39). We also note that the solution of Eq. (40) is a multivariate Gaussian distribution [1]. Transforming back to the original variables $\boldsymbol{Z}$, we find

$$\pi(\boldsymbol{Z}|\boldsymbol{G}) = (2\pi)^{-\frac{N_Z}{2}} \det(\Sigma_{\boldsymbol{Z}|\boldsymbol{G}})^{-\frac{1}{2}} \mathrm{e}^{-\frac{1}{2}(\boldsymbol{Z}-[\boldsymbol{Z}|\boldsymbol{G}])^T \Sigma^{-1}_{\boldsymbol{Z}|\boldsymbol{G}} (\boldsymbol{Z}-[\boldsymbol{Z}|\boldsymbol{G}])}. \tag{41}$$

The covariance matrix $\Sigma_{\boldsymbol{Z}|\boldsymbol{G}} = \Omega^{-1}\langle \boldsymbol{\epsilon}_{Z|\boldsymbol{G}}\, \boldsymbol{\epsilon}^T_{Z|\boldsymbol{G}}\rangle$ satisfies the equation

$$\underline{\mathrm{J}}_{\boldsymbol{G}} \Sigma_{\boldsymbol{Z}|\boldsymbol{G}} + \Sigma_{\boldsymbol{Z}|\boldsymbol{G}} \underline{\mathrm{J}}^T_{\boldsymbol{G}} + \frac{1}{\Omega} \underline{\mathrm{D}}_{\boldsymbol{G}} = 0, \tag{42}$$

which can be derived from Eq. (40). Moreover, we note that the coefficient matrices $\underline{\mathrm{J}}_{\boldsymbol{G}}$ and $\underline{\mathrm{D}}_{\boldsymbol{G}}$ depend on the conditional expectations $[\boldsymbol{Z}|\boldsymbol{G}]$, as is already implied in the definition after Eq. (38). The result is simply the LNA for the conditional distribution of gene products under steady state conditions.

3.2.2 Distribution of promoter states

We now seek to find an approximation for the distribution of the promoter states. Similarly to our derivation of Eq. (35) in the previous section, we expand the step operator and the reaction propensities via

$$\prod_{z=1}^{N_Z} E^{-S^{(2)}_{zj}} = 1 - \Omega^{-\frac{1}{2}} \sum_{z=1}^{N_Z} \nabla_z S^{(1)}_{zj} + O(\Omega^{-1}) \tag{43}$$

and

$$f^{(2)}_j(\boldsymbol{G}, \boldsymbol{Z}|\boldsymbol{G}) = f^{(2)}_j(\boldsymbol{G}, [\boldsymbol{Z}|\boldsymbol{G}]) + \Omega^{-\frac{1}{2}} \sum_{z=0}^{N_Z} (\boldsymbol{\epsilon}_{Z|\boldsymbol{G}})_z \frac{\partial f^{(2)}_j(\boldsymbol{G}, [\boldsymbol{Z}|\boldsymbol{G}])}{\partial [Z_z|\boldsymbol{G}]} + O(\Omega^{-1}), \tag{44}$$

respectively. Performing the conditional average over $\pi(\boldsymbol{Z}|\boldsymbol{G})$, as given by Eq. (41), we find

$$\begin{aligned} \frac{\partial \Pi(\boldsymbol{G}, t)}{\partial t} &= \left(\sum_{\boldsymbol{Z}} \mathcal{L}_1\, \pi(\boldsymbol{Z}|\boldsymbol{G}) \right) \Pi(\boldsymbol{G}, t) = \left(\int d(\boldsymbol{\epsilon}_{Z|\boldsymbol{G}})\ \mathcal{L}_1\, \pi(\boldsymbol{\epsilon}_{Z|\boldsymbol{G}}) \right) \Pi(\boldsymbol{G}, t) \\ &= \sum_{j=1}^{R_2} \left(\prod_{g=1}^{N_G} E^{-R_{gj}} - 1 \right) f^{(2)}_j\left(\boldsymbol{G}, [\boldsymbol{Z}|\boldsymbol{G}]\right) \Pi(\boldsymbol{G}, t) + O(\Omega^{-1}), \end{aligned} \tag{45}$$

where we have used partial integration and the fact that $\langle \boldsymbol{\epsilon}_{Z|\boldsymbol{G}} \rangle_\pi = 0$. The result is a CME for the slow promoter state whose propensities depend on the average product concentrations conditioned on the promoter state. We note that, unlike in the previously considered binary gene, the current example includes the effects of feedback, since the propensities depend explicitly on the downstream network.

Solution of the resulting CME, Eq. (45), is typically tractable, as (i) it involves only effective "unimolecular" reactions, (ii) its state space is finite, and (iii) the number of interacting genes is small, e.g., tens to a hundred in *E. coli* [12], and genes are present only in one or two copies numbers per cell. In order to demonstrate how an explicit solution to Eq. (45) can be obtained, let us enumerate all of the N_G possible promoter states as follows,

$$\boldsymbol{s}_i = (n_{G_1} = 0, .., n_{G_i} = 1, .., n_{G_{N_G}} = 0), \tag{46}$$

such that the probability of being in either state can be written as the vector

$$\boldsymbol{\Pi}(t) = (\Pi_1(t), \Pi_2(t), ..., \Pi_{N_G}(t)). \tag{47}$$

The above definition implies $\sum_{i=0}^{N_G} \Pi_i = 1$. Eq. (45) is then represented by the Markov chain

$$\frac{d\boldsymbol{\Pi}}{dt} = \underline{\mathrm{W}}\,([\boldsymbol{Z}|\boldsymbol{G}])\boldsymbol{\Pi}(t). \tag{48}$$

The transition matrix $\underline{\mathrm{W}}$ is a function of the conditional expectations $[\boldsymbol{Z}|\boldsymbol{G}]$ and assumes a particularly simple form in the case of a single promoter with N_G internal states, $[\underline{\mathrm{W}}]_{ij}([\boldsymbol{Z}|\boldsymbol{G}]) = \sum_k R_{ik} f_k^{(2)}(\boldsymbol{s}_j, [\boldsymbol{Z}|\boldsymbol{s}_j])$. We note that, in this case, the above equations are simply the rate equations for the reactions in group (ii).

3.2.3 Distribution of expression products

Using Eq. (41), we find that the marginal distribution of gene expression products is a mixed Gaussian distribution,

$$\Pi(\boldsymbol{Z}, t) = \sum_{\boldsymbol{G}} \pi(\boldsymbol{Z}|\boldsymbol{G})\Pi(\boldsymbol{G}, t), \tag{49}$$

since $\pi(\boldsymbol{Z}|\boldsymbol{G})$, as given in Eq. (41), is Gaussian. We note that the time-dependence of the resulting reduced distribution is only due to slow promoter fluctuations, as well as that is no particular restriction on the general form of the distribution. However, in practice, we are interested in finite and small mixtures.

3.3 Modes of distribution

It is obvious that the distribution in (49) is a multivariate Gaussian and, hence, unimodal if the promoter state is constant. The simplest situation – which is still non-trivial – arises in a configuration with two promoter states:

$$\Pi(\boldsymbol{Z}, t) = \Pi_0(t)\pi_0(\boldsymbol{Z}) + \big(1 - \Pi_0(t)\big)\pi_1(\boldsymbol{Z}). \tag{50}$$

Let us denote the conditional means and variances of π_i by $\boldsymbol{\mu}_i$ and $\underline{\Sigma}_i$, respectively, where $i = 0, 1$. The general theory of the modes of a distribution is very difficult to obtain analytically. In the case of proportional covariance matrices, i.e., for $\underline{\Sigma}_1 = \sigma^2 \underline{\Sigma}_0$, it can be shown that the distribution can have at most two modes [13]. (We remark that graphical techniques for the general case are also devised there.) Using the fact that the corresponding marginal distributions are Gaussian, as well, that result can immediately be applied to the marginal univariate distribution for a single gene product $n_{P,i}$:

$$\Pi(Z_i, t) = \Pi_0(t)\pi_0(Z_i) + \big(1 - \Pi_0(t)\big)\pi_1(Z_i). \tag{51}$$

Hence, the marginal distributions of a two-state promoter can have at most two modes that are independent of the underlying downstream network. We note, however, that this statement does not extend to joint distributions; in fact, a gene network with a binary promoter and N_Z products, Eq. (50), can have as many as $N_Z + 1$ modes [14]. Specifically, it follows that a marginal bivariate density of a binary mixture can have three modes, which is remarkable, as these modes are not evident from the reduced probability distributions.

3.4 Binary promoter switching: approximate solution

Next, we apply our approximation – which is based on the conditional LNA – to the example of binary gene expression in Eq. (27). It is now clear that the conditional expectation values of the fast mRNA and protein concentrations can simply be obtained by solving the rate equations for the standard model of gene expression

$$\begin{aligned} &\mathcal{G} \xrightarrow{k_0^{(\mathrm{on,off})}} \mathcal{G} + \mathcal{M}, \quad \mathcal{M} \xrightarrow{k_1} \mathcal{M} + \mathcal{P}, \\ &\mathcal{M} \xrightarrow{k_2} \varnothing, \quad \mathcal{P} \xrightarrow{k_3} \varnothing. \end{aligned} \tag{52}$$

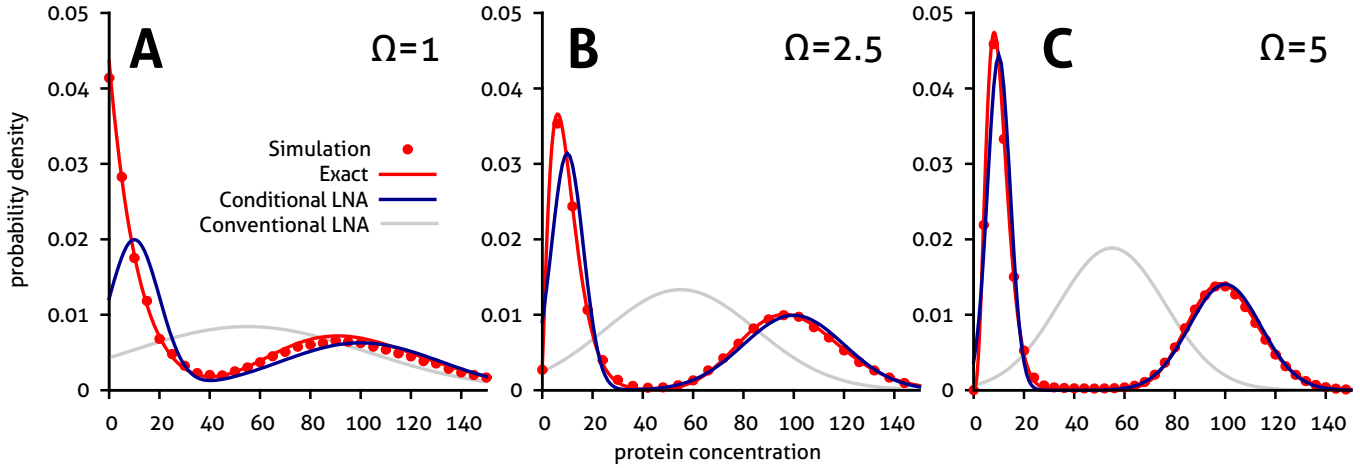


Figure S1: **Accuracy of the conditional LNA.** We compared the distributions obtained via the conditional LNA (Eq. (56), solid blue) against the exact distributions derived in the case of slow promoter fluctuations (Eq. (33), solid red) and stochastic simulation (dotted red). (A) For small cell volumes ($\Omega = 1$), we find that both the exact distribution and the one obtained from simulation have the off-state peaked at zero, whereas the conditional LNA distribution peaks away from zero. Nevertheless, the number of modes as well as the on-state contribution are qualitatively well captured by the conditional LNA. (B, C) The conditional LNA becomes increasingly more accurate as the cell volume Ω is increased, reflecting the limit in which it has been obtained. For comparison, we also show the conventional LNA (solid gray), which predicts the correct concentration and variance on average in this case; however, it cannot account for the bimodality of the distribution functions, as its solution is invariably Gaussian. Parameter values are $k_{\text{on}} = 0.01$ $k_{\text{off}} = 0.01$, $k_0^{(\text{on})} = 10$, $k_0^{(\text{off})} = 1$, $k_1 = 100$, $k_2 = 10$, and $k_3 = 1$.

We note that the above reaction diagram represents two models of gene expression with different translation rates $k_0^{(\text{on})}$ and $k_0^{(\text{off})}$. The conditional averages of the mRNA concentration in the two promoter states are then given by

$$[Z_M|(1,0)] = \frac{k_0^{(\text{on})}}{k_2} \quad \text{and} \quad [Z_M|(0,1)] = \frac{k_0^{(\text{off})}}{k_2}, \tag{53}$$

where $(1,0)$ and $(0,1)$ denote the state $\boldsymbol{G} = (n_G, n_{G*})$ in the on-configuration and the off-configuration, respectively. The conditional expectation values for the protein concentration are then related to the ones for mRNA by

$$[Z_P|\boldsymbol{G}] = \frac{k_1}{k_3}[Z_M|\boldsymbol{G}]. \tag{54}$$

Similarly, we can obtain the variances for each of the conditional models above; the result is

$$\underline{\Sigma}_{\boldsymbol{Z}|\boldsymbol{G}} = \frac{[Z_M|\boldsymbol{G}]}{\Omega} \begin{pmatrix} 1 & \frac{k_1}{k_2+k_3} \\ \frac{k_1}{k_2+k_3} & \frac{k_1}{k_3}\frac{k_1+k_2+k_3}{k_2+k_3} \end{pmatrix}. \tag{55}$$

Hence, it follows that the joint distribution of $\boldsymbol{Z} = (Z_M, Z_P)$ is given by

$$\begin{aligned} \Pi(\boldsymbol{Z}) = &\frac{k_{\text{on}}}{k_{\text{on}}+k_{\text{off}}}(2\pi)^{-1}\det(\underline{\Sigma}_{\boldsymbol{Z}|(1,0)})^{-\frac{1}{2}}\mathrm{e}^{-\frac{1}{2}(\boldsymbol{Z}-[\boldsymbol{Z}|(1,0)])^T\underline{\Sigma}^{-1}_{\boldsymbol{Z}|(1,0)}(\boldsymbol{Z}-[\boldsymbol{Z}|(1,0)])} \\ &+ \frac{k_{\text{off}}}{k_{\text{on}}+k_{\text{off}}}(2\pi)^{-1}\det(\underline{\Sigma}_{\boldsymbol{Z}|(0,1)})^{-\frac{1}{2}}\mathrm{e}^{-\frac{1}{2}(\boldsymbol{Z}-[\boldsymbol{Z}|(0,1)])^T\underline{\Sigma}^{-1}_{\boldsymbol{Z}|(0,1)}(\boldsymbol{Z}-[\boldsymbol{Z}|(0,1)])}. \end{aligned} \tag{56}$$

From Eq. (55), it is obvious that the covariance matrices for the on-state and the off-state are proportional, which implies that the distribution can have at most two modes.

In Fig. S1, we compare the distributions obtained from the conditional LNA against the exact solution (see Section 3.1) and against stochastic simulation. We find that, for low volumes Ω (corresponding to very low numbers of protein molecules), the conditional LNA only accounts for the shape of the on-state contribution, but not for the one of the off-state (Fig. S1 A); nevertheless, it qualitatively predicts the bimodal character of the distribution and its relative weights. For moderate Ω-values (Figs. S1 B,C), the overall agreement is excellent, which hence justifies the use of the underlying approximation.

3.5 Moments of distribution

Thus far, we have been concerned with "full" probability distributions. In the following, we sketch the derivation of the moments of the distribution in Eq. (49), which are often of interest in practice. It is straightforward to infer the mean value, which is defined as the sum of conditional expectations, weighted with the probabilities of being in a particular promoter state:

$$\langle \boldsymbol{Z}(t)\rangle = \sum_{\boldsymbol{G}} \langle \boldsymbol{Z}|\boldsymbol{G}\rangle_\pi \Pi(\boldsymbol{G},t). \tag{57}$$

However, the above relation does not carry over to the covariances, which are given by

$$\underline{\Sigma}\,(t) = \sum_{\boldsymbol{G}} \Pi(\boldsymbol{G},t)\left[\underline{\Sigma}_{\boldsymbol{Z}|\boldsymbol{G}} + ([\boldsymbol{Z}|\boldsymbol{G}] - \langle \boldsymbol{Z}(t)\rangle)([\boldsymbol{Z}|\boldsymbol{G}] - \langle \boldsymbol{Z}(t)\rangle)^T\right], \tag{58}$$

where the first term is due to fast gene product fluctuations, and is of order Ω^{-1}, while the second term is due to slow promoter fluctuations, and of the order Ω^0. These scalings are conspicuous, as they imply the presence of two contributions to the protein fluctuations, one from conditional protein fluctuations whose coefficient of variation is expected to scale as $\langle \boldsymbol{Z}\rangle^{-\frac{1}{2}}$ for small protein numbers and a second contribution which, by the above equation, scales as $\langle \boldsymbol{Z}\rangle^0$ due to promoter noise. The latter hence resembles contributions that are expected from extrinsic fluctuation sources, as are common to single-cell studies [15].

Making use of the definition of conditional moments as

$$\begin{aligned} \mu_i(t) &= [Z_i|\boldsymbol{G}] - \langle Z_i(t)\rangle, \\ \mu_{i,j,\ldots,z}^{\alpha_i,\alpha_j,\ldots,\alpha_z} &= \langle (Z_i|\boldsymbol{G} - \langle Z_i|\boldsymbol{G}\rangle)^{\alpha_i}(Z_j|\boldsymbol{G} - \langle Z_j|\boldsymbol{G}\rangle)^{\alpha_j}\ldots(Z_z|\boldsymbol{G} - \langle Z_z|\boldsymbol{G}\rangle)^{\alpha_z}\rangle, \end{aligned} \tag{59}$$

all higher-order moments can be expressed in terms of these via

$$\begin{aligned} &\langle (Z_i(t) - \langle Z_i(t)\rangle)^{\alpha_i}(Z_j(t) - \langle Z_j(t)\rangle)^{\alpha_j}\ldots(Z_z(t) - \langle Z_z(t)\rangle)^{\alpha_z}\rangle \\ &\quad = \sum_{\boldsymbol{G}} \Pi(\boldsymbol{G},t) \sum_{\beta_i=0}^{\alpha_i}\sum_{\beta_j=0}^{\alpha_j}\ldots\sum_{\beta_z=0}^{\alpha_z} \binom{\alpha_i}{\beta_i}\binom{\alpha_j}{\beta_j}\ldots\binom{\alpha_z}{\beta_z} \mu_i^{\alpha_i-\beta_i}\mu_j^{\alpha_j-\beta_j}\ldots\mu_z^{\alpha_z-\beta_z}\ \mu_{i,j,\ldots,z}^{\beta_i,\beta_j,\ldots,\beta_z}, \end{aligned} \tag{60}$$

for all times. (The above expression is obtained from $\langle (Z_i(t) - \langle Z_i(t)\rangle)^{\alpha_i} = \langle (Z_i(t) - [Z_i|\boldsymbol{G}] + [Z_i|\boldsymbol{G}] - \langle Z_i(t)\rangle)^{\alpha_i}$ and an application of the binomial theorem.)

3.6 Time correlation functions and power spectra

We now evaluate two-time correlation functions under *stationary conditions* and obtain explicit expressions for the resulting power spectra. Using the ansatz in (34) and the fact that $\langle \boldsymbol{\epsilon}_{Z|\boldsymbol{G}}\rangle_\pi = 0$, we find

$$\left\langle \left(\boldsymbol{Z}(t) - \langle \boldsymbol{Z}\rangle\right)\left(\boldsymbol{Z}(t') - \langle \boldsymbol{Z}\rangle\right)^T \right\rangle = \underline{\Lambda}\,(t,t') + \Omega^{-1}\langle \boldsymbol{\epsilon}_{Z|\boldsymbol{G}}(t)\,\boldsymbol{\epsilon}_{Z|\boldsymbol{G}'}(t')^T\rangle, \tag{61}$$

where the first term

$$\underline{\Lambda}\,(t,t') = \sum_{\boldsymbol{G},\boldsymbol{G}'} \langle([\boldsymbol{Z}|\boldsymbol{G}] - \langle\boldsymbol{Z}\rangle)\left([\boldsymbol{Z}|\boldsymbol{G}'] - \langle\boldsymbol{Z}\rangle\right)^T\rangle \Pi(\boldsymbol{G},t|\boldsymbol{G}',t')\Pi(\boldsymbol{G}') \tag{62}$$

denotes the autocorrelation due to slow (promoter) fluctuations, while the second term is the autocorrelation function of fast fluctuations:

$$\langle \boldsymbol{\epsilon}_{Z|\boldsymbol{G}}(t)\,\boldsymbol{\epsilon}^T_{Z|\boldsymbol{G}'}(t')\rangle = \sum_{\boldsymbol{G},\boldsymbol{G}'} \langle \boldsymbol{\epsilon}_{Z|\boldsymbol{G}}(t)\,\boldsymbol{\epsilon}^T_{Z|\boldsymbol{G}'}(t')\rangle_\pi \Pi(\boldsymbol{G},t|\boldsymbol{G}',t')\Pi(\boldsymbol{G}'). \tag{63}$$

We note that, under stationary conditions, Eqs. (61) through (63) are functions of the time difference $t-t'$ only. Eq. (63) can be simplified significantly by noting that the first term decays on the fast timescale μ, while the second one decays on a much slower scale. Hence, we can approximate $\Pi(\boldsymbol{G},t|\boldsymbol{G}',t')$ by $\delta_{\boldsymbol{G},\boldsymbol{G}'}$ such that the above expression becomes

$$\langle \boldsymbol{\epsilon}_{Z|\boldsymbol{G}}(t)\,\boldsymbol{\epsilon}^T_{Z|\boldsymbol{G}'}(t')\rangle \approx \sum_{\boldsymbol{G}} \langle \boldsymbol{\epsilon}_{Z|\boldsymbol{G}}(t)\boldsymbol{\epsilon}^T_{Z|\boldsymbol{G}}(t')\rangle_\pi \Pi(\boldsymbol{G}). \tag{64}$$

Specifically, this approximation neglects the impact of promoter switching on the gene product dynamics; it can thus be evaluated using a standard LNA [16], albeit applied to the conditional fluctuations:

$$\langle \boldsymbol{\epsilon}_{Z|\boldsymbol{G}}(t)\boldsymbol{\epsilon}^T_{Z|\boldsymbol{G}}(t')\rangle_\pi = \frac{1}{\underline{\mathrm{J}}_{\boldsymbol{G}} - i\omega}\,\underline{\mathrm{D}}_{\boldsymbol{G}}\,\frac{1}{\underline{\mathrm{J}}^T_{\boldsymbol{G}} + i\omega}, \tag{65}$$

where $\underline{\mathrm{J}}_{\boldsymbol{G}}$ and $\underline{\mathrm{D}}_{\boldsymbol{G}}$ have been defined after Eq. (38). Since, however, no such argument holds for $\underline{\Lambda}\,(t,t')$, an explicit expression for the transition probability $\Pi(\boldsymbol{G},t|\boldsymbol{G}',t')$ needs to be obtained, as follows.

Solving Eq. (48) with initial condition δ_{ij} for $t > t'$ and the corresponding adjoint equation for $t < t'$, we find

$$\Pi(\boldsymbol{G},t|\boldsymbol{G}',t') = H(t-t')\mathrm{e}^{\underline{\mathrm{W}}\,(t-t')} + H(t'-t)\mathrm{e}^{\underline{\mathrm{W}}^T(t'-t)}, \tag{66}$$

where $H(\tau)$ is the Heaviside function. It then follows that, under stationarity, the two-time probability distribution is given by

$$\Pi(\boldsymbol{G},t|\boldsymbol{G}',t')\Pi(\boldsymbol{G}') \equiv \underline{\mathrm{G}}\,(t-t') = H(t-t')\mathrm{e}^{\underline{\mathrm{W}}\,(t-t')}\mathrm{diag}(\boldsymbol{\Pi}) + H(t'-t)\mathrm{diag}(\boldsymbol{\Pi})\mathrm{e}^{\underline{\mathrm{W}}^T(t'-t)}; \tag{67}$$

also,

$$\underline{\Lambda}\,(t,t') = ([\boldsymbol{Z}|\boldsymbol{G}] - \langle\boldsymbol{Z}\rangle)\;\underline{\mathrm{G}}\,(t-t')\left([\boldsymbol{Z}|\boldsymbol{G}'] - \langle\boldsymbol{Z}\rangle\right)^T. \tag{68}$$

Using the identity

$$\underline{\mathrm{F}}\,(\omega) = \int_0^\infty d\tau\,\mathrm{e}^{(\underline{\mathrm{W}} - i\omega)\tau} = \frac{1}{i\omega - \underline{\mathrm{W}}}, \tag{69}$$

we find the Fourier transform of the function $\underline{\mathrm{G}}\,(\tau)$ as

$$\underline{\mathrm{G}}\,(\omega) = \int_{-\infty}^\infty d\tau\,\mathrm{e}^{-i\omega\tau}\,\underline{\mathrm{G}}\,(\tau) = \underline{\mathrm{F}}\,(\omega)\mathrm{diag}(\boldsymbol{\Pi}) + \mathrm{diag}(\boldsymbol{\Pi})\,\underline{\mathrm{F}}^\dagger(\omega). \tag{70}$$

We note that the above expression is singular for $\omega \to 0$ because of the zero eigenvalue of the matrix $\underline{\mathrm{W}}$, which is due to conservation of probability; however, one can easily convince oneself that the resulting expression in (70) is not. It now follows that the Fourier transform of Eq. (61) is given by

$$\underline{\mathrm{S}}\,(\omega) = ([\boldsymbol{Z}|\boldsymbol{G}] - \langle\boldsymbol{Z}\rangle)\;\underline{\mathrm{G}}\,(\omega)\,([\boldsymbol{Z}|\boldsymbol{G}] - \langle\boldsymbol{Z}\rangle)^T + \frac{1}{\Omega}\sum_{\boldsymbol{G}} \Pi(\boldsymbol{G})\frac{1}{\underline{\mathrm{J}}_{\boldsymbol{G}} - i\omega}\,\underline{\mathrm{D}}_{\boldsymbol{G}}\,\frac{1}{\underline{\mathrm{J}}^T_{\boldsymbol{G}} + i\omega}, \tag{71}$$

where the term multiplying Ω^{-1} agrees with the usual result, found via LNA, for the conditional processes [16] averaged over all promoter states. Finally, we note that integration of Eq. (71) yields the stationary covariance

$$\underline{\Sigma} = \int \frac{d\omega}{2\pi} \underline{\mathrm{S}}(\omega) = \sum_{\boldsymbol{G}} \Pi(\boldsymbol{G}) \left[\underline{\Sigma}_{\boldsymbol{Z}|\boldsymbol{G}} + ([\boldsymbol{Z}|\boldsymbol{G}] - \langle \boldsymbol{Z} \rangle)([\boldsymbol{Z}|\boldsymbol{G}] - \langle \boldsymbol{Z} \rangle)^T \right], \tag{72}$$

which is in agreement with Eq. (58), as derived from the distribution of expression products in (49).

3.7 Multiple identical gene copies

The scenario of multiple identical gene copies can be analyzed in a straightforward manner on the basis of the case of just one promoter. Let us consider a single promoter that can be in one of N_G internal states. Then, the probability of being in either of these states can be written as the vector

$$\boldsymbol{\Pi}(t) = (\Pi_1(t), \Pi_2(t), \ldots, \Pi_{N_G}(t)), \tag{73}$$

where $\Pi_i(t) = \Pi(n_{G_1} = 0, \ldots, n_{G_i} = 1, .., n_{G_{N_G}} = 0, t)$ and $\sum_{i=0}^{N_G} \Pi_i = 1$ because of promoter conservation. We note that Π_i can be obtained from Eq. (48), and that it thus depends on conditional product concentrations through the transition rates.

The case of N_c identical promoters now follows from the fact that the probability of finding n promoters in state i is proportional to $\frac{\Pi_i^n}{n!}$, since these promoters are independent. The joint probability for the population in each state is hence multinomial in that case,

$$\Pi(n_{G_1}, \ldots, n_{G_{N_G}}, t) = N_c! \frac{\Pi_1^{n_{G_1}}}{n_{G_1}!} \cdots \frac{\Pi_{N_G-1}^{n_{N_G-1}}}{n_{G_{N_G-1}}!} \frac{\Pi_{N_G}^{n_{N_G}}}{n_{G_{N_G}}!} \tag{74}$$

with mean values $\langle \frac{n_{G_i}}{N_c} \rangle = \Pi_i$ that equal the average numbers of promoters in state i and covariance

$$\underline{\Sigma}_{\boldsymbol{G}} = \left\langle \frac{n_{G_i}}{N_c} \frac{n_{G_j}}{N_c} \right\rangle - \left\langle \frac{n_{G_i}}{N_c} \right\rangle \left\langle \frac{n_{G_j}}{N_c} \right\rangle = \begin{cases} \frac{\Pi_i(1-\Pi_i)}{N_c} & i = j, \\ -\frac{\Pi_i \Pi_j}{N_c} & i \neq j. \end{cases} \tag{75}$$

For $N_c \to \infty$, a Central Limit Theorem applies, which implies that the distribution $\Pi(\boldsymbol{G}, t)$ is asymptotically normal. It then follows that the joint distribution $\Pi(\boldsymbol{Z}, \boldsymbol{G}, t) = \pi(\boldsymbol{Z}|\boldsymbol{G})\Pi(\boldsymbol{G}, t)$ is a multivariate Gaussian in this limit and, hence, strictly unimodal, independently of the underlying network; moreover, the mean values correspond to solutions of the associated deterministic rate equations. The same conclusion then applies to the marginal distribution of gene products, in agreement with results obtained via a standard LNA.

4 Fast promoter fluctuations

For completeness, we now derive an analogous approximation for the probability distributions of gene products in the case of fast promoter fluctuations. To that end, we rescale $\boldsymbol{f}_*^{(2)} = \mu \boldsymbol{f}^{(2)}$ in Eq. (11); dropping the asterisk, we then obtain

$$\frac{d\Pi(\boldsymbol{G}, \boldsymbol{Z}, t)}{dt} = \mathcal{L}_0 \Pi(\boldsymbol{G}, \boldsymbol{Z}, t) + \frac{1}{\mu} \mathcal{L}_1 \Pi(\boldsymbol{G}, \boldsymbol{Z}, t), \tag{76}$$

where $\mathcal{L}_0$ and $\mathcal{L}_1$ are defined as in Eqs. (23a) and (23b), respectively. Hence, in the limit of $\mu \to 0$, the first term corresponding to the reactions in group (i) is to be treated as slow, while the second one,

which corresponds to group (ii) reactions, is to be considered fast. Using Bayes' theorem, we may write $\Pi(\boldsymbol{G},\boldsymbol{Z},t)=\Pi(\boldsymbol{G}|\boldsymbol{Z},t)\Pi(\boldsymbol{Z},t)$ to find

$$\frac{\partial\Pi(\boldsymbol{G}|\boldsymbol{Z},t)}{\partial t}\Pi(\boldsymbol{Z},t)+\Pi(\boldsymbol{G}|\boldsymbol{Z},t)\frac{\partial\Pi(\boldsymbol{Z},t)}{\partial t} \\ =\Pi(\boldsymbol{Z},t)\mathcal{L}_0\Pi(\boldsymbol{G}|\boldsymbol{Z},t)+\frac{1}{\mu}\mathcal{L}_1\Pi(\boldsymbol{G}|\boldsymbol{Z},t)\Pi(\boldsymbol{Z},t). \tag{77}$$

For $\mu\to 0$, we thus have $\Pi(\boldsymbol{G}|\boldsymbol{Z},t)\to\pi(\boldsymbol{G}|\boldsymbol{Z})$, and the reduced equations become

$$0=\mathcal{L}_1\pi(\boldsymbol{G}|\boldsymbol{Z}), \tag{78a}$$

$$\frac{\partial\Pi(\boldsymbol{Z},t)}{\partial t}=\left(\sum_{\boldsymbol{G}}\mathcal{L}_0\pi(\boldsymbol{G}|\boldsymbol{Z})\right)\Pi(\boldsymbol{Z},t), \tag{78b}$$

analogously to the case of slow promoter fluctuations.

4.1 Distribution of promoter states

The promoter distribution now follows from Eq. (78a). Let us enumerate all of the N_G possible promoter states as follows:

$$\boldsymbol{s}_i|\boldsymbol{Z}\equiv(n_{G_1}=0,..,n_{G_i}=1,..,n_{G_{N_G}}=0)|\boldsymbol{Z} \tag{79}$$

for fixed $\boldsymbol{Z}$, such that the probability of being in either state can be written as the vector

$$\boldsymbol{\Pi}=\left(\Pi(\boldsymbol{s}_1|\boldsymbol{Z}),\Pi(\boldsymbol{s}_2|\boldsymbol{Z}),\ldots,\Pi(\boldsymbol{s}_{N_G}|\boldsymbol{Z})\right), \tag{80}$$

which depends only on the value of $\boldsymbol{Z}$, but not on time. The above vector satisfies $\sum_{i=0}^{N_G}\Pi(\boldsymbol{s}_i|\boldsymbol{Z})=1$ because of promoter conservation; Eq. (78a) then becomes the Markov chain

$$0=\underline{\mathrm{W}}\,(\boldsymbol{Z})\boldsymbol{\Pi}. \tag{81}$$

Assuming that all promoters are present only in a single copy, the matrix $\underline{\mathrm{W}}$ is defined as $[\underline{\mathrm{W}}]_{ij}(\boldsymbol{Z})=\sum_k R_{ik}f_k^{(2)}(\boldsymbol{s}_j|\boldsymbol{Z},\boldsymbol{Z})$; hence, the above equation is formally equivalent to a set of macroscopic rate equations for the promoter state.

4.2 Distribution of expression products

Since the promoter state $\boldsymbol{G}$ enters $\mathcal{L}_0$ only linearly via the transcription propensity, Eq. (6), the latter is given by

$$\sum_{\boldsymbol{G}}\hat{f}_j^{(1)}(\boldsymbol{G},\boldsymbol{Z})\pi(\boldsymbol{G}|\boldsymbol{Z})=\Omega f_j^{(1)}(\langle\boldsymbol{G}|\boldsymbol{Z}\rangle,\boldsymbol{Z}). \tag{82}$$

It is obvious that $\langle\boldsymbol{G}|\boldsymbol{Z}\rangle$ can be calculated using the conventional statistical framework developed by Shea and Ackers [17, 18]. Making use of the ansatz

$$\boldsymbol{Z}=[\boldsymbol{Z}]+\Omega^{-\frac{1}{2}}\boldsymbol{\epsilon}_Z \tag{83}$$

for the gene product concentrations, we can expand Eq. (78b) as

$$\left(\frac{\partial}{\partial t}-\Omega^{\frac{1}{2}}\frac{d[\boldsymbol{Z}]}{dt}\cdot\nabla\right)\Pi(\boldsymbol{\epsilon}_Z,t) \\ =\left(-\Omega^{\frac{1}{2}}\nabla\cdot\underline{\mathrm{S}}^{(1)}\boldsymbol{f}^{(1)}(\langle\boldsymbol{G}|\boldsymbol{Z}\rangle,[\boldsymbol{Z}])-\nabla^T\underline{\mathrm{J}}_{\boldsymbol{Z}}\boldsymbol{\epsilon}_Z+\nabla^T\underline{\mathrm{D}}_{\boldsymbol{Z}}\nabla+O\big(\Omega^{-\frac{1}{2}}\big)\right)\Pi(\boldsymbol{\epsilon}_Z,t), \tag{84}$$

where $[\underline{\mathrm{J}}_{\boldsymbol{Z}}]_{ij} = \sum_\alpha S^{(1)}_{i\alpha}\partial_{[Z_i]} f^{(1)}_\alpha(\langle \boldsymbol{G}|\boldsymbol{Z}\rangle, [\boldsymbol{Z}])$ and $[\underline{\mathrm{D}}_{\boldsymbol{Z}}]_{ij} = \sum_\alpha S^{(1)}_{i\alpha} S^{(1)}_{j\alpha} f^{(1)}_\alpha(\langle \boldsymbol{G}|\boldsymbol{Z}\rangle, [\boldsymbol{Z}])$. Equating terms of orders $\Omega^{\frac{1}{2}}$ and Ω^0, respectively, we find

$$\frac{d[\boldsymbol{Z}]}{dt} = \underline{\mathrm{S}}^{(1)} \boldsymbol{f}^{(1)}(\langle \boldsymbol{G}|\boldsymbol{Z}\rangle, [\boldsymbol{Z}]), \tag{85a}$$

$$\frac{\partial \Pi(\boldsymbol{\epsilon}_Z, t)}{\partial t} = \left(-\nabla^T \underline{\mathrm{J}}_{\boldsymbol{Z}} \boldsymbol{\epsilon}_Z + \nabla^T \underline{\mathrm{D}}_{\boldsymbol{Z}} \nabla\right) \Pi(\boldsymbol{\epsilon}_Z, t) \tag{85b}$$

The solution to the last equation is a multivariate Gaussian with mean $[\boldsymbol{Z}]$ and variance $\underline{\Sigma}_{\boldsymbol{Z}}(t)$, as determined by the equation

$$\frac{\partial \underline{\Sigma}_{\boldsymbol{Z}}}{\partial t} = \underline{\mathrm{J}}_{\boldsymbol{Z}} \underline{\Sigma}_{\boldsymbol{Z}} + \underline{\Sigma}_{\boldsymbol{Z}} \underline{\mathrm{J}}^T_{\boldsymbol{Z}} + \frac{1}{\Omega} \underline{\mathrm{D}}_{\boldsymbol{Z}}. \tag{86}$$

Transforming back to the concentration variable $\boldsymbol{Z}$, we find the probability distribution

$$\Pi(\boldsymbol{Z}, t) = (2\pi)^{-\frac{N_Z}{2}} \det(\Sigma_{\boldsymbol{Z}}(t))^{-\frac{1}{2}} \mathrm{e}^{-\frac{1}{2}(\boldsymbol{Z}-[\boldsymbol{Z}(t)])^T \Sigma^{-1}_{\boldsymbol{Z}}(t)(\boldsymbol{Z}-[\boldsymbol{Z}(t)])}. \tag{87}$$

Hence, the gene product distribution is clearly unimodal, in contrast to the case of slow promoter fluctuations discussed in the previous section.

5 Applications

5.1 Global control of gene expression promotes multimodality

We consider a switch of two mutually repressing genes which also includes translational regulation, and which is described by the set of reactions

$$\mathcal{P}_1 + \mathcal{G}_2 \underset{a_{-1}}{\overset{a_1}{\rightleftharpoons}} \mathcal{P}_1\mathcal{G}_2, \quad \mathcal{P}_2 + \mathcal{G}_1 \underset{b_{-1}}{\overset{b_1}{\rightleftharpoons}} \mathcal{P}_2\mathcal{G}_1, \tag{88a}$$

$$\mathcal{G}_1 \xrightarrow{a_2} \mathcal{G}_1 + \mathcal{M}_1, \quad \mathcal{G}_1\mathcal{P}_2 \xrightarrow{a_3} \mathcal{G}_1\mathcal{P}_2 + \mathcal{M}_1, \quad \mathcal{M}_1 \xrightarrow{a_4} \mathcal{M}_1 + \mathcal{P}_1, \quad \mathcal{M}_1 \xrightarrow{a_5} \varnothing, \quad \mathcal{P}_1 \xrightarrow{a_6} \varnothing,$$

$$\mathcal{G}_2 \xrightarrow{b_2} \mathcal{G}_2 + \mathcal{M}_2, \quad \mathcal{G}_2\mathcal{P}_1 \xrightarrow{b_3} \mathcal{G}_2\mathcal{P}_1 + \mathcal{M}_2, \quad \mathcal{M}_2 \xrightarrow{b_4} \mathcal{M}_2 + \mathcal{P}_2, \quad \mathcal{M}_2 \xrightarrow{b_5} \varnothing, \quad \mathcal{P}_2 \xrightarrow{b_6} \varnothing, \tag{88b}$$

$$\mathcal{P}_1 + \mathcal{M}_2 \underset{b_7}{\overset{b_8}{\rightleftharpoons}} \mathcal{M}_2\mathcal{P}_1 \xrightarrow{b_9} \varnothing. \tag{88c}$$

Here, the reactions in (88a) account for transcriptional regulation without cooperativity, while those in (88b) represent transcription, translation, and consequent degradation. We note that we consider the particular case of an additional post-transcriptional regulation mechanism, as given by (88c), in which protein species $\mathcal{P}_1$ inhibits the mRNA $\mathcal{M}_2$ of protein species $\mathcal{P}_2$. If the reactions in (88a) are slow, so is the joint promoter state $\boldsymbol{G} = (n_{G_1}, n_{G_2})$. The vector of gene product concentrations $\boldsymbol{Z} = (M_1, M_2, P_1, P_2, M_2P_1)$ can hence be characterized by its distribution conditional on the promoter state, the distribution of which is evaluated explicitly in the following.

Conditional means

Promoter conservation implies the relations $n_{G_1P_2} = 1 - n_{G_1}$ and $n_{G_2P_1} = 1 - n_{G_2}$; hence, the total promoter state can be expressed as $\boldsymbol{G} = (n_{G_1}, n_{G_2})$, which can be in one of four states, since both genes can be either on or off. Letting $\alpha(n_{G_1}) = a_2 n_{G_1} + a_3(1 - n_{G_1})$ and $\beta(n_{G_2}) = b_2 n_{G_2} + b_3(1 - n_{G_2})$ denote the transcription rates of either promoter, where the transcription rates $a_{2,3}$ and $b_{2,3}$ are defined according to the convention

in Eq. (6), we can then write

$$\begin{aligned}
\frac{d[M_1|\boldsymbol{G}]}{dt} &= \alpha(\boldsymbol{G}) - a_5[M_1|\boldsymbol{G}],\\
\frac{d[M_2|\boldsymbol{G}]}{dt} &= \beta(\boldsymbol{G}) - b_5[M_2|\boldsymbol{G}] + b_7[M_2P_1|\boldsymbol{G}] - b_8[M_2|\boldsymbol{G}][P_1|\boldsymbol{G}],\\
\frac{d[P_1|\boldsymbol{G}]}{dt} &= a_4[M_1|\boldsymbol{G}] - a_6[P_1|\boldsymbol{G}] + b_7[M_2P_1|\boldsymbol{G}] - b_8[M_2|\boldsymbol{G}][P_1|\boldsymbol{G}],\\
\frac{d[P_2|\boldsymbol{G}]}{dt} &= b_4[M_2|\boldsymbol{G}] - b_6[P_2|\boldsymbol{G}],\\
\frac{d[M_2P_1|\boldsymbol{G}]}{dt} &= -b_7[M_2P_1|\boldsymbol{G}] + b_8[M_2|\boldsymbol{G}][P_1|\boldsymbol{G}] - b_9[M_2P_1|\boldsymbol{G}].
\end{aligned} \tag{89}$$

In the absence of post-translational regulation, $b_8 = 0$, the above equations are readily solved at steady state by setting the time derivatives to zero, which yields

$$\begin{aligned}
[M_2P_1|\boldsymbol{G}] &= 0,\\
[M_1|\boldsymbol{G}] = [M_1|n_{G_1}] &= \frac{\alpha(n_{G_1})}{a_5}, \quad [P_1|\boldsymbol{G}] = [P_1|n_{G_1}] = \frac{\alpha(n_{G_1})a_4}{a_5a_6},\\
[M_2|\boldsymbol{G}] = [M_2|n_{G_2}] &= \frac{\beta(n_{G_2})}{b_5}, \quad [P_2|\boldsymbol{G}] = [P_2|n_{G_2}] = \frac{\beta(n_{G_2})b_4}{b_5b_6}.
\end{aligned} \tag{90}$$

The conditional expectation values for the concentrations of P_1 and P_2 thus only depend on their respective promoter states G_1 and G_2. Analogously, the same statement can be shown to hold for the variances, which implies that P_1 and P_2 are independent conditional on $\boldsymbol{G}$. Hence, Eq. (90) admits only two sets of solutions. Moreover, and as shown below, P_1 and P_2 are also independent unconditionally.

Considering the case of post-transcriptional regulation, with $b_8 \neq 0$, we find

$$\begin{aligned}
[M_2P_1|\boldsymbol{G}] &= \frac{\beta(n_{G_2}) - b_5[M_2|\boldsymbol{G}]}{b_9},\\
[M_1|\boldsymbol{G}] &= \frac{\alpha(n_{G_1})}{a_5}, \quad [M_2|\boldsymbol{G}] = \frac{b_9(a_5\beta(n_{G_2}) - a_4\alpha(n_{G_1})) - a_5a_6b_5K_I + R(\boldsymbol{G})}{2a_5b_5b_9},\\
[P_1|\boldsymbol{G}] &= \frac{a_4\alpha(n_{G_1}) - a_5\beta(n_{G_2}) + a_5b_5[M_2|\boldsymbol{G}]}{a_5a_6}, \quad [P_2|\boldsymbol{G}] = \frac{b_4[M_2|\boldsymbol{G}]}{b_6},
\end{aligned} \tag{91}$$

where we have introduced the mRNA inhibition constant

$$K_I = \frac{b_7 + b_9}{b_8} \tag{92}$$

and the abbreviation

$$R(\boldsymbol{G}) = \sqrt{2\alpha(n_{G_1})a_4a_5b_9(a_6b_5K_I - \beta(n_{G_2})b_9) + a_5^2(b_9\beta(n_{G_2}) + a_6b_5K_I)^2 + a_4^2b_9^2\alpha(n_{G_1})^2}. \tag{93}$$

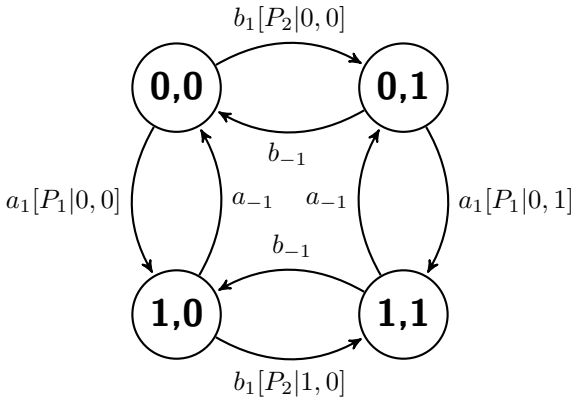
We observe that, since $[P_1|\boldsymbol{G}]$ and $[P_2|\boldsymbol{G}]$ depend on both α and β in this case, P_1 and P_2 are not conditionally independent. In particular, $[P_1|\boldsymbol{G}]$ and $[P_2|\boldsymbol{G}]$ differ for each of the four joint promoter states.

Distribution of promoter states

Under steady state conditions, the promoter distribution is now obtained from

$$0 = \begin{pmatrix}
-b_1[P_2|0,0] - a_1[P_1|0,0] & a_{-1} & 0 & b_{-1}\\
a_1[P_1|0,0] & -b_1[P_2|1,0] - a_{-1} & b_{-1} & 0\\
0 & b_1[P_2|1,0] & -a_{-1} - b_{-1} & a_1[P_1|0,1]\\
b_1[P_2|0,0] & 0 & a_{-1} & -b_{-1} - a_1[P_1|0,1]
\end{pmatrix}\boldsymbol{\Pi}, \tag{94}$$

where $\mathbf{\Pi} = (\Pi(0,0), \Pi(1,0), \Pi(0,1), \Pi(1,1))$, and where the above transition matrix has been constructed using the explicit expression for $\underline{\mathrm{W}}$ for the single-promoter case given after Eq. (48). A graphical representation of the transition matrix linking the promoter states helps understand the effect of the two regulatory regimes:



Here, horizontal transitions represent the binding of P_1, while vertical ones correspond to binding of P_2 due to transcriptional regulation. In the absence of post-transcriptional regulation, we have $[P_2|0,0] = [P_2|1,0] \equiv [P_2|0]$, $[P_1|0,0] = [P_1|0,1] \equiv [P_1|0]$; hence, promoter binding does not correlate the promoters in that regime. Defining the abbreviations $K_0 = \frac{a_{-1}}{b_{-1}}$, $K_1 = \frac{a_1}{a_{-1}}$, and $K_2 = \frac{b_1}{b_{-1}}$, the corresponding solution for $\mathbf{\Pi}$ can be obtained from the conditional expectations in Eq. (90):

$$\begin{aligned}
\Pi(0,0) &= \left(\frac{1}{1+K_1[P_1|0]}\right)\left(\frac{1}{1+K_2[P_2|0]}\right), \\
\Pi(1,0) &= \left(\frac{K_1[P_1|0]}{1+K_1[P_1|0]}\right)\left(\frac{1}{1+K_2[P_2|0]}\right), \\
\Pi(0,1) &= \left(\frac{1}{1+K_1[P_1|0]}\right)\left(\frac{K_2[P_2|0]}{1+[P_2|0]K_2}\right), \\
\Pi(1,1) &= \left(\frac{K_1[P_1|0]}{1+K_1[P_1|0]}\right)\left(\frac{K_2[P_2|0]}{1+[P_2|0]K_2}\right).
\end{aligned} \tag{95}$$

Despite the presence of transcriptional regulation, the above result implies that both promoters and, hence, their respective proteins are independent, i.e., that $\Pi(n_{G_1}, n_{G_2}) = \Pi(n_{G_1})\Pi(n_{G_2})$ as well as that $\Pi(P_1, P_2) = \Pi(P_1)\Pi(P_2)$.

When post-transcriptional regulation is present, i.e., for $b_8 \neq 0$, the promoter distribution is obtained from Eq. (94) using the conditional expectations (91) and taking into account that $[P_1|0,0] \neq [P_1|0,1]$ and $[P_2|0,0] \neq [P_2|1,0]$:

$$\begin{aligned}
\Pi(0,0) &= \frac{1 + K_0 + K_2[P_2|1,0] + K_1K_0[P_1|0,1]}{U}, \\
\Pi(1,0) &= \frac{(K_1K_2[P_2|0,0][P_1|0,1] + K_1[P_1|0,0]) + K_0(K_1^2[P_1|0,0][P_1|0,1] + K_1[P_1|0,0])}{U}, \\
\Pi(0,1) &= \frac{K_2[P_2|0,0]([P_2|1,0]K_2 + K_0 + 1) + K_2K_1K_0[P_1|0,0][P_2|1,0]}{U}, \\
\Pi(1,1) &= \frac{K_1K_2([P_2|1,0]((K_2[P_2|0,0][P_1|0,1] + [P_1|0,0]) + K_1K_0[P_1|0,0][P_1|0,1]) + K_0[P_2|0,0][P_1|0,1])}{U},
\end{aligned} \tag{96}$$

where

$$U = K_0(K_1[P_1|0,1]+1)([P_2|0,0]K_2+[P_2|1,0]K_1K_2[P_1|0,0]+K_1[P_1|0,0]+1) \\ +([P_2|1,0]K_2+1)\{[P_2|0,0](K_1K_2[P_1|0,1]+K_2)+K_1[P_1|0,0]+1\}. \tag{97}$$

Eq. (96) shows that both promoters are effectively correlated as a result of the regulation. A particularly interesting aspect of joint regulation is that of detailed balance, which concerns the question of whether the two promoters are in mutual equilibrium; in other words, the flux from the clockwise and counter-clockwise transitions in the transition graph must be equal, leading to the condition

$$\frac{[P_1|0,0]}{[P_1|0,1]} = \frac{[P_2|0,0]}{[P_2|1,0]}. \tag{98}$$

We note that, in this case, the regulation of the two-promoter dynamics mimics that of an allosteric complex. The above condition can for instance be satisfied for the particular choice of rate constants $a_2 = b_2$, $a_3 = b_3$, $a_4 = b_4$, $a_5 = b_5$, and $a_6 = b_6$, as in Table S2, which implies that the transcription and translation rates of both proteins are identical; however, it does not necessarily hold in general. We note that, although the lack of detailed balance stemming from the interplay of transcriptional and post-transcriptional regulation has been demonstrated for a specific two-promoter network here, our argument can be extended in a straightforward manner to any gene regulatory network that involves slow promoters.

Conditional gene product distribution

It remains to find the solution of the Lyapunov equation, Eq. (42). The corresponding Jacobian can be constructed either from its definition, as given after Eq. (38), or directly from Eq. (89); the result is

$$\underline{\mathrm{J}}_{\boldsymbol{G}} = \begin{pmatrix} -a_5 & 0 & 0 & 0 & 0 \\ 0 & -b_5 - b_8[P_1|\boldsymbol{G}] & -b_8[M_2|\boldsymbol{G}] & 0 & b_7 \\ a_4 & -b_8[P_1|\boldsymbol{G}] & -a_6 - b_8[M_2|\boldsymbol{G}] & 0 & b_7 \\ 0 & b_4 & 0 & -b_6 & 0 \\ 0 & b_8[P_1|\boldsymbol{G}] & b_8[M_2|\boldsymbol{G}] & 0 & -b_7 - b_9 \end{pmatrix}. \tag{99}$$

Using the definition of the matrix $\underline{\mathrm{D}}_{\boldsymbol{G}}$ given after Eq. (38), we identify the following non-vanishing matrix elements

$$\begin{aligned} D_{1,1} &= \alpha(n_{G_1}) + a_5[M_1|\boldsymbol{G}], \\ D_{2,2} &= \beta(n_{G_2}) + b_5[M_2|\boldsymbol{G}] + b_7[M_2P_1|\boldsymbol{G}] + b_8[M_2|\boldsymbol{G}][P_1|\boldsymbol{G}], \\ D_{3,3} &= a_4[M_1|\boldsymbol{G}] + a_6[P_1|\boldsymbol{G}] + b_7[M_2P_1|\boldsymbol{G}] + b_8[M_2|\boldsymbol{G}][P_1|\boldsymbol{G}], \\ D_{4,4} &= b_4[M_2|\boldsymbol{G}] + b_6[P_2|\boldsymbol{G}], \\ D_{5,5} &= b_7[M_2P_1|\boldsymbol{G}] + b_9[M_2P_1|\boldsymbol{G}] + b_8[M_2|\boldsymbol{G}][P_1|\boldsymbol{G}], \\ D_{2,3} &= D_{3,2} = -D_{2,5} = -D_{5,2} = -D_{3,5} = -D_{5,3} = b_7[M_2P_1|\boldsymbol{G}] + b_8[M_2|\boldsymbol{G}][P_1|\boldsymbol{G}], \end{aligned} \tag{100}$$

which can then be used in Eq. (15) to obtain $\underline{\Sigma}_{\boldsymbol{Z}|\boldsymbol{G}}$. For each promoter, the latter yields a unique solution for $\underline{\Sigma}_{\boldsymbol{Z}|(0,0)}$, $\underline{\Sigma}_{\boldsymbol{Z}|(1,0)}$, $\underline{\Sigma}_{\boldsymbol{Z}|(0,1)}$, and $\underline{\Sigma}_{\boldsymbol{Z}|(1,1)}$ whose explicit form is quite elaborate, and is hence omitted here. More straightforwardly, the solution is obtained numerically using Eqs. (91) and (96), from which the steady state gene product distribution can be constructed as

$$\Pi(\boldsymbol{Z}) = \sum_{n_{G_1}=0}^{1}\sum_{n_{G_2}=0}^{1} \Pi(n_{G_1},n_{G_2})(2\pi)^{-\frac{N_Z}{2}} \det(\Sigma_{\boldsymbol{Z}|(n_{G_1},n_{G_2})})^{-\frac{1}{2}} \mathrm{e}^{-\frac{1}{2}(\boldsymbol{Z}-[\boldsymbol{Z}|n_{G_2},n_{G_2}])^T\Sigma^{-1}_{\boldsymbol{Z}|(n_{G_1},n_{G_2})}(\boldsymbol{Z}-[\boldsymbol{Z}|n_{G_1},n_{G_2}])}; \tag{101}$$

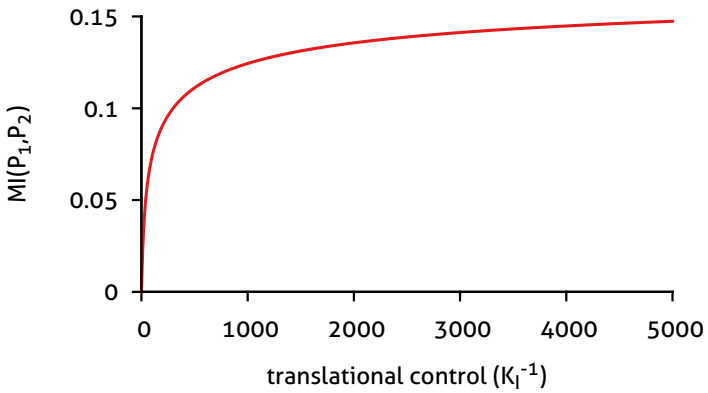


Figure S2: **Mutual information shared between two proteins interacting via translational inhibition**, shown as a function of the inverse mRNA inhibition constant K_I (strength of post-translational control). Here, K_I has been varied via the bimolecular association constant b_8, using Eq. (103), while the remaining parameters are as given in Table S2, with the exception of $a_1 = 0 = b_1$. The measure does not exceed 0.15bit of information over the considered range of translational control.

similarly, one can obtain the reduced distribution functions $\Pi(P_1, P_2)$, $\Pi(P_1)$, and $\Pi(P_2)$. The mutual information shared between the two protein species is defined by [19]

$$\mathrm{MI}(P_1, P_2) = \int_{-\infty}^{\infty} dP_1 \int_{-\infty}^{\infty} dP_2\, \Pi(P_1, P_2) \log_2 \left(\frac{\Pi(P_1, P_2)}{\Pi(P_1)\Pi(P_2)} \right). \tag{102}$$

Since the proteins are independent in the absence of post-transcriptional regulation, we have $\Pi(P_1, P_2) = \Pi(P_1)\Pi(P_2)$ and, hence, $\mathrm{MI}(P_1, P_2) = 0$. In the absence of transcriptional regulation, the promoter distribution is $\Pi(n_{G_1}, n_{G_2}) = \delta_{1,1}$; the joint distribution is thus a Gaussian, for which the mutual information reduces to the well known expression

$$\mathrm{MI}(P_1, P_2) = -\frac{1}{2} \log_2 \left(1 - \frac{[\Sigma_{\boldsymbol{Z}|(1,1)}]^2_{3,4}}{[\Sigma_{\boldsymbol{Z}|(1,1)}]_{3,3}[\Sigma_{\boldsymbol{Z}|(1,1)}]_{4,4}} \right), \tag{103}$$

where the second term in the brackets is simply the correlation coefficient of the conventional LNA. In Fig. S2, we show the dependence of the mutual information on the mRNA inhibition constant K_I which measures the strength of post-translational control. In the general case which involves both transcriptional and post-transcriptional regulation, the mutual information has to be evaluated numerically.

5.2 Birhythmicity in the expression of a genetic oscillator

For the genetic oscillator described in the main text, we formulated a simplified model that is based on complex elementary reactions of Michaelis-Menten type, similarly to the negative feedback oscillator

employed in [20]. The reactions and their associated propensities are given by

$$\begin{aligned}
\mathcal{G} &\xrightarrow{k_{\rm off}} \mathcal{G}^* &:\ & k_{\rm off} n_G, \\
\mathcal{G}^* &\xrightarrow{k_{\rm on}} \mathcal{G} &:\ & k_{\rm on} n_{G^*}, \\
\mathcal{G} &\xrightarrow{k_0^{\rm on}} \mathcal{G} + \mathcal{P} &:\ & k_0^{\rm on} \Omega n_G, \\
\mathcal{G}^* &\xrightarrow{k_0^{\rm off}} \mathcal{G}^* + \mathcal{P} &:\ & k_0^{\rm off} \Omega n_{G^*}, \\
\mathcal{P} &\xrightarrow{k_1} \varnothing &:\ & k_1 n_P, \\
\mathcal{R}_P + \mathcal{P} &\xrightarrow{k_2} \mathcal{R}_P &:\ & \tfrac{k_2}{\Omega} n_P n_{R_P}, \\
\mathcal{Y} &\xrightarrow{k_3} \mathcal{Y}_P &:\ & \tfrac{k_3 n_P n_Y}{n_Y + \Omega K_{M,1}}, \\
\mathcal{Y}_P &\xrightarrow{k_4} \mathcal{Y} &:\ & \Omega \tfrac{k_4 n_{Y_P}}{n_{Y_P} + \Omega K_{M,2}}, \\
\mathcal{R} &\xrightarrow{k_5} \mathcal{R}_P &:\ & \Omega \tfrac{k_5 n_{Y_P} n_R}{n_R + \Omega K_{M,3}}, \\
\mathcal{R}_P &\xrightarrow{k_6} \mathcal{R} &:\ & \Omega \tfrac{k_6 n_{R_P}}{n_{R_P} + \Omega K_{M,3}}.
\end{aligned} \tag{104}$$

We note that the first two reactions are the only ones that affect the promoter state, and that they are henceforth assumed to be slow. Since the above reaction scheme obeys the conservation laws $n_Y + n_{Y_P} = [Y_T]\Omega$, $n_R + n_{R_P} = [R_T]\Omega$, and $n_G + n_{G^*} = N_c$, where N_c is the gene dosage, the reduced vector of gene product concentrations can be written as $\boldsymbol{Z} = (P, Y_P, R_P)$, while the full promoter state is given by $\boldsymbol{G} = (n_G, n_G^*)$.

We now calculate the power spectrum

$$\underline{\mathrm{S}}\,(\omega) = (\,[\boldsymbol{Z}|\boldsymbol{G}] - \langle \boldsymbol{Z} \rangle\,)\ \underline{\mathrm{G}}\,(\omega)\,(\,[\boldsymbol{Z}|\boldsymbol{G}] - \langle \boldsymbol{Z} \rangle\,)^T + \frac{1}{\Omega} \sum_{\boldsymbol{G}} \Pi(\boldsymbol{G}) \frac{1}{\underline{\mathrm{J}}_{\boldsymbol{G}} - i\omega}\, \underline{\mathrm{D}}_{\boldsymbol{G}}\, \frac{1}{\underline{\mathrm{J}}_{\boldsymbol{G}}^T + i\omega}, \tag{105}$$

as found by Eq. (71). The above expression depends on (i) the promoter distribution, (ii) the conditional means, and (iii) the conditional power spectra, which correspond to the last term in Eq. (105). The calculation is performed explicitly for the cases of one and two promoters in the following.

Promoter distribution

For a single promoter, the steady state promoter distribution is given by

$$\Pi_1 \equiv \Pi(1,0) = \frac{k_{\rm on}}{k_{\rm on} + k_{\rm off}} \quad \text{and} \quad \Pi_0 \equiv \Pi(0,1) = \frac{k_{\rm off}}{k_{\rm on} + k_{\rm off}}. \tag{106}$$

Using the above expression, the probability of N_c identical promoters can be obtained from Eq. (74),

$$\Pi(n_G, n_{G^*}) = \left(\begin{array}{c} N_c \\ n_G \end{array} \right) \Pi_1^{n_G} \Pi_0^{N_c - n_G} \delta_{n_{G^*}, N_c - n_G}, \tag{107}$$

which is the binomial distribution.

Conditional means

By virtue of Eq. (39), the conditional means of gene product concentrations obey

$$\begin{aligned}
\frac{d[P|\boldsymbol{G}]}{dt} &= k_0^{\rm on} n_G + k_0^{\rm off} n_{G^*} - k_1 [P|\boldsymbol{G}] - k_2 [P|\boldsymbol{G}][R|\boldsymbol{G}], \\
\frac{d[Y_P|\boldsymbol{G}]}{dt} &= \frac{k_3([Y_T] - [Y_P|\boldsymbol{G}])}{([Y_T] - [Y_P|\boldsymbol{G}]) + K_{M,1}} - \frac{k_4 [Y_P|\boldsymbol{G}]}{[Y_P|\boldsymbol{G}] + K_{M,2}}, \\
\frac{d[R_P|\boldsymbol{G}]}{dt} &= \frac{k_5([R_T] - [R_P|\boldsymbol{G}])}{([R_T] - [R_P|\boldsymbol{G}]) + K_{M,3}} - \frac{k_6 [R_P|\boldsymbol{G}]}{[R_P|\boldsymbol{G}] + K_{M,3}}.
\end{aligned} \tag{108}$$

The numerical values for one and two identical promoters, based on the parameter values in Table S3, are given in Table S5.

Power spectra of oscillations

Power spectra can be obtained by taking into account the explicit form of the relevant transition matrices. For a single promoter, the latter reads simply

$$\underline{\mathrm{W}} = \begin{pmatrix} -k_{\mathrm{off}} & k_{\mathrm{on}} \\ k_{\mathrm{off}} & -k_{\mathrm{on}} \end{pmatrix}, \tag{109}$$

from which we find

$$\underline{\mathrm{G}}\,(\omega) = \frac{2k_{\mathrm{off}}\Pi_1}{(k_{\mathrm{off}}+k_{\mathrm{on}})^2+\omega^2} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}; \tag{110}$$

hence, the first term in Eq. (105) becomes

$$([Z_i|\boldsymbol{G}] - \langle Z_i\rangle)\ \underline{\mathrm{G}}\,(\omega)\,([Z_i|\boldsymbol{G}] - \langle Z_i\rangle)^T = \frac{2k_{\mathrm{off}}\Pi_1}{(k_{\mathrm{off}}+k_{\mathrm{on}})^2+\omega^2}([Z_i|(1,0)] - [Z_i|(0,1)])^2. \tag{111}$$

For two identical promoters, the transition matrix between the states $(2,0)$, $(0,2)$, and $(1,1)$ is given by

$$\underline{\mathrm{W}} = \begin{pmatrix} -2k_{\mathrm{off}} & 0 & k_{\mathrm{on}} \\ 0 & -2k_{\mathrm{on}} & k_{\mathrm{off}} \\ 2k_{\mathrm{off}} & 2k_{\mathrm{on}} & -k_{\mathrm{off}}-k_{\mathrm{on}} \end{pmatrix}. \tag{112}$$

A similar calculation as above now implies

$$\begin{aligned} ([Z_i|\boldsymbol{G}] - \langle Z_i\rangle)\,\underline{\mathrm{G}}\,(\omega)\,([Z_i|\boldsymbol{G}] - \langle Z_i\rangle)^T &= D^{-1}k_{\mathrm{off}}(4k_{\mathrm{off}}^2+\omega^2)([Z_i|(0,2)] - [Z_i|(1,1)])^2 \\ &+ D^{-1}k_{\mathrm{on}}\left(4k_{\mathrm{on}}^2+\omega^2\right)([Z_i|(2,0)] - [Z_i|(1,1)])^2 \\ &+ D^{-1}([Z_i|(2,0)] - [Z_i|(0,2)])k_{\mathrm{off}}^2k_{\mathrm{on}}([Z_i|(2,0)] - 5[Z_i|(0,2)] + 4[Z_i|(1,1)]) \\ &+ D^{-1}([Z_i|(2,0)] - [Z_i|(0,2)])k_{\mathrm{off}}k_{\mathrm{on}}^2(5[Z_i|(2,0)] - [Z_i|(0,2)] - 4[Z_i|(1,1)]), \end{aligned} \tag{113}$$

with

$$D = \frac{\omega^4(k_{\mathrm{off}}+k_{\mathrm{on}})^2 + 5\omega^2(k_{\mathrm{off}}+k_{\mathrm{on}})^4 + 4(k_{\mathrm{off}}+k_{\mathrm{on}})^6}{4k_{\mathrm{off}}k_{\mathrm{on}}}. \tag{114}$$

In order to evaluate the power spectra of the conditional fast variables, as given by the second term in Eq. (105), we employ the representation of the corresponding Jacobian

$$\begin{aligned} \underline{\mathrm{J}}_{\boldsymbol{G}} &= \begin{pmatrix} -\alpha_1 & 0 & \beta_3 \\ \beta_1 & -\alpha_2 & 0 \\ 0 & \beta_2 & -\alpha_3 \end{pmatrix} \\ &= \begin{pmatrix} -k_1 - k_2[R_P|\boldsymbol{G}] & 0 & -k_2[P|\boldsymbol{G}] \\ \frac{k_3([Y_T]-[Y_P|\boldsymbol{G}])}{K_{M,3}-[Y_P|\boldsymbol{G}]+[Y_T]} & -\frac{k_3K_{M,3}[P|\boldsymbol{G}]}{(K_{M,3}-[Y_P|\boldsymbol{G}]+[Y_T])^2} - \frac{k_4K_{M,4}}{(K_{M,4}+[Y_P|\boldsymbol{G}])^2} & 0 \\ 0 & \frac{k_5([R_T]-[R_P|\boldsymbol{G}])}{K_{M,5}-[R_P|\boldsymbol{G}]+[R_T]} & -\frac{k_5K_{M,5}[Y_P|\boldsymbol{G}]}{(K_{M,5}-[R_P|\boldsymbol{G}]+[R_T])^2} - \frac{k_6K_{M,6}}{(K_{M,6}+[R_P|\boldsymbol{G}])^2} \end{pmatrix} \end{aligned} \tag{115}$$

in terms of its eigenvalues, $\underline{\mathrm{J}}_{\boldsymbol{G}}\,\underline{\mathrm{U}}\,(\boldsymbol{G}) = \mathrm{diag}(\lambda_1(\boldsymbol{G}),\lambda_2(\boldsymbol{G}),\lambda_3(\boldsymbol{G}))\,\underline{\mathrm{U}}\,(\boldsymbol{G})$ for each state $\boldsymbol{G}$. Using a result that can be found in [21], we obtain

$$\sum_{\boldsymbol{G}} \Pi(\boldsymbol{G}) \frac{1}{\underline{\mathrm{J}}_{\boldsymbol{G}} - i\omega}\,\underline{\mathrm{D}}_{\boldsymbol{G}}\,\frac{1}{\underline{\mathrm{J}}_{\boldsymbol{G}}^T + i\omega} = \sum_{\boldsymbol{G}} \Pi(\boldsymbol{G}) \sum_{ij} U_{si}(\boldsymbol{G}) \frac{1}{\lambda_i(\boldsymbol{G}) - i\omega}\,\widetilde{\underline{\mathrm{D}}}_{\boldsymbol{G}}\,\frac{1}{\lambda_j^*(\boldsymbol{G}) + i\omega} U_{js}^\dagger(\boldsymbol{G}) \tag{116}$$

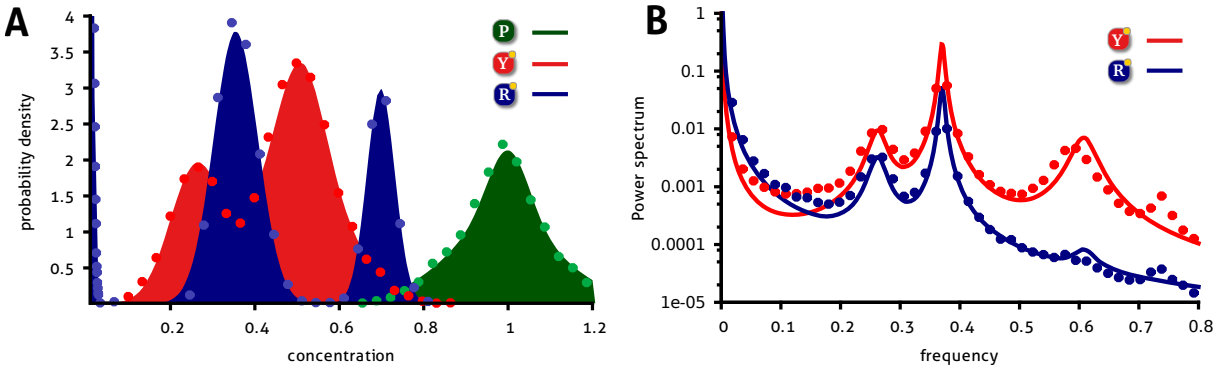


Figure S3: **Increasing the gene dosage of the genetic oscillator.** We doubled the copy number in the genetic oscillator shown in Fig. 4A of the main text. For comparison with Figs. 4C and 4D there, we have also accounted for dosage compensation by halving the transcription rate. In (A), we observe that increased dosage leads to a new state in the oscillator output, while the modality of the downstream components remains robust as a consequence of the ultrasensitive response. That new, undecided state corresponds to one of the promoters occupying the high state, with the remaining one being in the low state. For better comparison, the concentration of P (green) has been divided by a factor of two. Surprisingly, in (B) we also observe a new period of oscillation in the power spectra of both kinases, which is predicted by the conditional LNA (solid) and confirmed by stochastic simulation (dotted). We note that the new mode in distribution and its characteristic frequency are not associated with those of any of the individual promoter states, but that they correspond to the average promoter state; hence, they represent an emergent property of the gene regulatory network. We also note that the conditional LNA captures all features of the oscillatory dynamics, apart from higher-order harmonics which our theory does not describe [21].

for the second term in Eq. (105), where $\widetilde{\underline{\mathrm{D}}}_{\boldsymbol{G}} = \underline{\mathrm{U}}^{-1}\,\underline{\mathrm{D}}\,\underline{\mathrm{U}}^{-\dagger}$. The matrix of eigenvectors $\underline{\mathrm{U}}$ can be written in terms of the eigenvalues $\boldsymbol{\lambda}(\boldsymbol{G})$, using the characteristic equation of $\underline{\mathrm{J}}_{\boldsymbol{G}}$:

$$U_{1j} = \beta_3, \quad U_{2j} = (\alpha_3 + \lambda_j)(\alpha_1 + \lambda_j), \quad \text{and} \quad U_{3j} = (\alpha_1 + \lambda_j), \tag{117}$$

where $j = 1, 2, 3$ and where the eigenvalues for the examples considered here are given in Table S5. The noise matrix $\underline{\mathrm{D}}_{\boldsymbol{G}}$ has the following non-vanishing elements,

$$\begin{aligned} D_{11} &= k_0^{\mathrm{on}} n_G + k_0^{\mathrm{off}} n_{G^*} + k_1[P|\boldsymbol{G}] + k_2[P|\boldsymbol{G}][R_P|\boldsymbol{G}], \\ D_{22} &= \frac{k_4[Y_P|\boldsymbol{G}]}{K_{M,4} + [Y_P|\boldsymbol{G}]} + \frac{k_3[P|\boldsymbol{G}]([Y_T] - [Y_P|\boldsymbol{G}])}{K_{M,3} - [Y_P|\boldsymbol{G}] + [Y_T]}, \\ D_{33} &= \frac{k_6[R_P|\boldsymbol{G}]}{K_{M,6} + [R_P|\boldsymbol{G}]} + \frac{k_5([R_T] - [R_P|\boldsymbol{G}])[Y_P|\boldsymbol{G}]}{K_{M,5} - [R_P|\boldsymbol{G}] + [R_T]}. \end{aligned} \tag{118}$$

In particular, the form of Eq. (116) implies that the power spectrum has peaks at eigenvalues with small real parts; the imaginary parts of these eigenvalues yield the frequencies of the resulting oscillations [21]. Using the parameter values given in Table S3, we deduce that, in the case of a single promoter, the frequencies are obtained as 0.26 and 0.61, see Table S5; in the case of two identical promoters, on the other hand, there appears an additional frequency at 0.37. These predictions agree well with both the approximated and the simulated power spectra, as shown in Fig. 4 of the main text and in Fig. S3.

Summarizing, the first term in Eq. (105) is given by (111) for a single promoter and by (113) for two identical promoters. Using Eqs. (118) and (117) together with the eigenvalues given in Table S5 in (116), the second term in Eq. (105) can be evaluated.

5.3 Phenotype induction: transient bimodality and hysteresis

We consider a hypothetical experiment in which a transcriptional activator $\mathcal{T}$, induced at rate $s(t)$, increases transcriptional activity by binding to a target promoter. The corresponding reaction diagram can be written as

$$\varnothing \xrightarrow{s(t)} \mathcal{T}, \tag{119a}$$

$$\mathcal{G} + \mathcal{T} \underset{k_{\text{off}}}{\overset{k_{\text{on}}}{\rightleftharpoons}} \mathcal{GT}, \tag{119b}$$

$$\mathcal{G} \xrightarrow{k_0^{(\text{off})}} \mathcal{G} + \mathcal{M}, \tag{119c}$$

$$\mathcal{GT} \xrightarrow{k_0^{(\text{on})}} \mathcal{GT} + \mathcal{M},$$

$$\mathcal{M} \xrightarrow{k_1} \mathcal{M} + \mathcal{P}, \tag{119d}$$

$$\mathcal{M} \xrightarrow{k_2} \varnothing, \quad \mathcal{P} \xrightarrow{k_3} \varnothing, \quad \mathcal{T} \xrightarrow{k_4} \varnothing. \tag{119e}$$

Here, the stimulus $s(t)$ in Eq. (119a) is assumed to be varying slowly, while the degradation of $\mathcal{T}$ is fast. The promoter state is then given by $\boldsymbol{G} = (n_G, n_{GT})$, for which we will denote the on-state and the off-state by (on) $= (0, 1)$ and (off) $= (1, 0)$, respectively. If we postulate fast degradation and slow binding of $\mathcal{T}$, as described by the reactions in (119b), we have

$$[T|\text{off}] = \frac{s(t)}{k_4}. \tag{120}$$

Averaging the promoter state distribution over the slow promoter association dynamics, we find

$$\frac{d\Pi^{(\text{on})}}{dt} = -k_{\text{off}}\Pi^{(\text{on})} + \frac{k_{\text{on}}}{k_4}s(t)(1 - \Pi^{(\text{on})}); \tag{121}$$

which implies that $s(t)$ drives promoter activation. The above equation has the general solution

$$\Pi^{(\text{on})}(t) = \Pi^{(\text{on})}(t_0)\mathrm{e}^{-\beta(t_0,t)} + \frac{k_{\text{on}}}{k_4}\int_{t_0}^{t} dt'\,\mathrm{e}^{-\beta(t',t)}s(t'), \tag{122}$$

with

$$\beta(t', t) = k_{\text{off}}(t - t') + \frac{k_{\text{on}}}{k_4}\int_{t'}^{t} d\tau\, s(\tau). \tag{123}$$

While it is often easier to solve Eq. (121) numerically, we carry out the solution procedure explicitly for a simple stimulus which can be realized experimentally, and which is given by a ramp of finite duration α:

$$s(t) = \begin{cases} s_0\frac{t}{\alpha} & 0 \le t \le \alpha, \\ s_0 & t > \alpha, \end{cases} \tag{124}$$

In particular, we are interested in scenarios where the promoter is induced up to some $t > \alpha$, at which point the time-reverse experiment is performed by decreasing the ramp from the induced state to zero. We denote the solutions to these two scenarios by $\Pi_f^{(\text{on})}(t)$ and $\Pi_r^{(\text{on})}(t)$, respectively.

In order to understand the implications of our experiment, we focus on three dynamical regimes: (i) the slow-induction regime of a slowly varying ramp ($\alpha \gg \tau_G \gg \mu^{-1}$, with τ_G being the promoter timescale); (ii) the fast-induction regime of a fast ramp resulting in a step-function ($\tau_G \gg \alpha \gg \mu^{-1}$); and, finally, (iii) the intermediate regime, where $\tau_G \simeq \alpha \gg \tau_Z$.

Slow induction

Regime (i) is realized by setting the left-hand side of Eq. (122) equal to zero, and by solving for $\Pi^{(\mathrm{on})}$. For $\alpha \gg \tau_G$, the probability of induction is proportional to the stimulus

$$\Pi_f^{(\mathrm{on})}(t) = \Pi_r^{(\mathrm{on})}(t) = \begin{cases} 0 & t < 0 \\ A\frac{t}{\alpha} & 0 \leq t \leq \alpha \\ A & t > \alpha \end{cases}, \tag{125}$$

to lowest order, where $A = \frac{[T]_0}{[T]_0 + K_{\mathrm{eq}}}$ denotes the steady state probability after induction, $[T]_0 = \frac{s_0}{k_4}$ is the corresponding TF concentration, and $K_{\mathrm{eq}} = \frac{k_{\mathrm{off}}}{k_{\mathrm{on}}}$ denotes the equilibrium constant of TF binding. Hence, for slow induction, the forward and reverse induction experiments yield the same result. We also note that, for $s_0 \gg \frac{k_{\mathrm{off}} k_4}{k_{\mathrm{on}}}$, the final state is fully induced, in the sense that $A \simeq 1$ then.

Fast induction

Regime (ii) follows by taking $\alpha \to 0$ in Eq. (124): direct integration of the resulting piecewise constant signal $s(t)$ yields

$$\Pi_f^{(\mathrm{on})}(t) = \begin{cases} 0 & t < 0 \\ A\left(1 - \mathrm{e}^{-\left(k_{\mathrm{off}} + s_0 \frac{k_{\mathrm{on}}}{k_4}\right)t}\right) & t > 0 \end{cases}, \tag{126a}$$

$$\Pi_r^{(\mathrm{on})}(t) = \begin{cases} A\mathrm{e}^{k_{\mathrm{off}} t} & t < 0 \\ A & t > 0 \end{cases}. \tag{126b}$$

We note that, in contrast to the regime of slow induction, the time-dependence is exponential, with different rates for the forward and reverse experiments.

Intermediate regime

We now show that hysteresis persists if the promoter and induction timescales are of the same order; however, as we will see, the solution for $\Pi_{f,r}^{(\mathrm{on})}$ can only be performed semi-analytically in that regime. Defining $\lambda = \frac{k_{\mathrm{on}}}{k_4}\frac{s_0}{\alpha}$ and writing $\mathrm{erfi}(x) = \frac{2}{\sqrt{\pi}}\int_0^x dt\, \mathrm{e}^{t^2}$ for the imaginary error function, we can express the solution to the forward experiment as

$$\Pi_f^{(\mathrm{on})}(t) = \begin{cases} 0 & t < 0 \\ 1 - \mathrm{e}^{-k_{\mathrm{off}} t - \lambda\frac{t^2}{2}} + k_{\mathrm{off}}\sqrt{\frac{\pi}{2\lambda}}\mathrm{e}^{-\frac{(k_{\mathrm{off}}+\lambda t)^2}{2\lambda}}\left(\mathrm{erfi}\left(\frac{k_{\mathrm{off}}}{\sqrt{2\lambda}}\right) - \mathrm{erfi}\left(\frac{k_{\mathrm{off}}+\lambda t}{\sqrt{2\lambda}}\right)\right) & 0 \leq t \leq \alpha \\ \Pi^{(\mathrm{on})}(\alpha)\mathrm{e}^{-\left(k_{\mathrm{off}} + s_0\frac{k_{\mathrm{on}}}{k_4}\right)(t-\alpha)} + A\left(1 - \mathrm{e}^{-\left(k_{\mathrm{off}} + s_0\frac{k_{\mathrm{on}}}{k_4}\right)(t-\alpha)}\right) & t > \alpha \end{cases}; \tag{127}$$

here, $\Pi^{(\mathrm{on})}$ is given as in Eq. (122).

When considering the reverse experiment, one starts from the induced state and slowly decreases the induction rate, thus inverting time. The result is

$$\Pi_r^{(\mathrm{on})}(t) = \begin{cases} \Pi^{(\mathrm{on})}(0)\mathrm{e}^{k_{\mathrm{off}} t} & t < 0 \\ A\mathrm{e}^{k_{\mathrm{off}}(t-\alpha) - \lambda(t^2-\alpha^2)} + 1 - \mathrm{e}^{k_{\mathrm{off}}(t-\alpha) + \lambda\frac{t^2-\alpha^2}{2}} & \\ +k_{\mathrm{off}}\sqrt{\frac{\pi}{2\lambda}}\mathrm{e}^{\frac{(k_{\mathrm{off}}+\lambda t)^2}{2\lambda}}\left(\mathrm{erf}\left(\frac{k_{\mathrm{off}}+\lambda t}{\sqrt{2\lambda}}\right) - \mathrm{erf}\left(\frac{k_{\mathrm{off}}+\alpha\lambda}{\sqrt{2\lambda}}\right)\right) & 0 \leq t \leq \alpha \\ A & t > \alpha \end{cases}, \tag{128}$$

where $\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x dt\, \mathrm{e}^{-t^2}$ denotes the error function. Clearly, the two solutions coincide in the limits of $t \to \pm\infty$; however, for intermediate times, they are generally different, reflecting promoter memory.

Distribution of expression products

Thus far, we have been concerned with determining promoter state distributions. In all three regimes considered above, the gene product distribution is given by the Gaussian mixture

$$\Pi_{f,r}^{(\mathrm{on})}(\boldsymbol{Z}_P,t)=\Pi_{f,r}^{\mathrm{on}}(t)\pi(\boldsymbol{Z}_P|\mathrm{on})+\left(1-\Pi_{f,r}^{(\mathrm{on})}(t)\right)\pi(\boldsymbol{Z}_P|\mathrm{off}), \tag{129}$$

since we have assumed that the promoter timescales are much slower than those of gene products; here, the conditional distributions in the above expression are as given in Section 4. We note that the above expression is valid both for slow and for fast induction kinetics. Moreover, it is clear that, since the transcription rates in both states are significantly different, there will be some intermediate time interval in which the protein distribution enters a transient bimodal state. In regime (i), this observation is true for forward and reverse induction experiments at equal induction rates. However, in regime (ii), increasing and decreasing the induction rates does not yield identical protein distributions; rather, one observes a dependence on the induction history and, hence, a hysteresis phenomenon. In particular, it follows from the case of fast induction, Eq. (126), that the timescale ratio for induction and promoter decay in the reverse experiment is given by

$$\frac{\tau_r}{\tau_f}=1+\frac{[T]_0}{K_{\mathrm{eq}}}, \tag{130}$$

which is a measure of the hysteresis observable in the response of the gene network. In particular, the memory of the system is greatly enhanced for either large perturbations $[T]_0$ or transcription factors with small K_{eq}. Finally, we remark that regime (iii), for which the timescales of promoter and induction are of the same order, has been considered in the main text using the parameter values in Table S4.

5.4 Positive cooperative feedback: comparison against deterministic bistability

Populations of isogenic cells exhibiting distinct subpopulations are commonly modeled by deterministic rate equations whose steady state exhibits multiple steady states. Multistability often requires positive genetic feedback interlinked with post-transcriptional kinetics in order to achieve an increase in effective cooperativity, leading to an ultrasensitive response as, for example, in mycobacterial persistence [22].

Here, we study a simple example of bistability in positive autoregulation where cooperativity is achieved via post-translational formation of oligomeric complexes. We consider a single gene $\mathcal{G}$ encoding a protein $\mathcal{P}$ at basal levels and its dimeric and tetrameric forms $\mathcal{D}$ and $\mathcal{Q}$, respectively. For simplicity, we do not consider mRNA transcription explicitly here. The tetrameric complex promotes the expression of $\mathcal{P}$ via the DNA-protein complex $\mathcal{GQ}$; hence, induction is cooperative. The set of elementary reactions comprising this model are given by

$$\begin{aligned}
&\mathcal{G}+\mathcal{Q}\underset{k_{\mathrm{off}}}{\overset{k_{\mathrm{on}}}{\rightleftharpoons}}\mathcal{GQ},\\
&\mathcal{G}\xrightarrow{\delta}\mathcal{G}+\mathcal{P},\quad \mathcal{GQ}\xrightarrow{k_T+\delta}\mathcal{GQ}+\mathcal{P},\quad \mathcal{P}\xrightarrow{k_{\mathrm{dp}}}\varnothing,\\
&2\mathcal{P}\underset{k_{mD}}{\overset{k_D}{\rightleftharpoons}}\mathcal{D},\quad 2\mathcal{D}\underset{k_{mQ}}{\overset{k_Q}{\rightleftharpoons}}\mathcal{Q},
\end{aligned} \tag{131}$$

where we have assumed the oligomeric complexes to be stable so that their decay can be neglected. In the following, we assume the association and dissociation of the dimer $\mathcal{Q}$ and the promoter $\mathcal{G}$ to be slow compared to the remaining reactions. Next, we compare the solution of the deterministic rate equation model corresponding to Eq. (131) against our conditional LNA of the CME, and against the results of stochastic simulation.

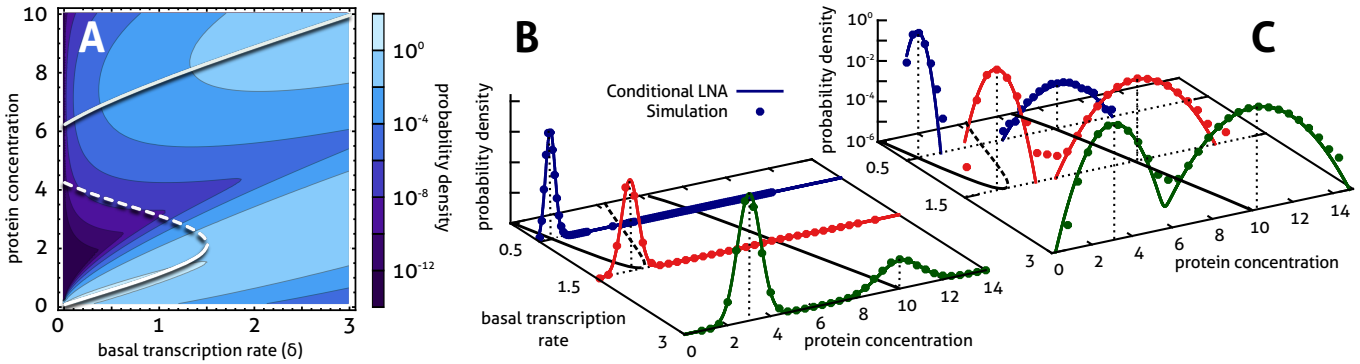


Figure S4: **Comparison of deterministic bistability against stochastic bimodality.** (A) Solutions of the deterministic rate equation (solid white) are compared to the contours of the probability density that is obtained from the conditional LNA as a function of the basal transcription rate δ. (B) Conditional LNA solution for three parameters in the deterministically bistable ($\delta = 0.5$, blue), marginally bistable ($\delta = 1.5$, red), and deterministically monostable ($\delta = 3$, green) regimes. The solution of the deterministic equation is also shown for comparison (solid black). We observe bimodality to be most evident outside of the region of deterministic bistability ($\delta > 1.5$), which is well confirmed by stochastic simulation (dotted). (C) The same distributions as in (B) are presented on a logarithmic scale, demonstrating that, within the region of deterministic bistability, the distributions obtained from the conditional LNA and from stochastic simulation are indeed bimodal. In particular, the deterministic theory correctly predicts the modality of the distributions (solid horizontal lines) in the deterministically bistable region, whereas it fails in the monostable region. Parameter values are $k_{\rm on} = 10^{-3}$, $k_{\rm off} = 2 \times 10^{-4}$ $k_{\rm dp} = 1$, $k_T = 7$, $k_D = 1$. $k_Q = 0.1$, $k_{mD} = 10$, $k_{mQ} = 1$, and $\Omega = 10$.

Deterministic bistability

We define the apparent dissociation constant

$$K_d = \frac{k_{\rm off}}{k_{\rm on}} \frac{k_{mQ}}{k_Q} \left(\frac{k_{mD}}{k_D} \right)^2 ; \tag{132}$$

then, the deterministic rate equations for the reaction network (131) lead to the following condition

$$0 = \delta + \frac{k_T [P]^4}{K_d + [P]^4} - k_{\rm dp}[P] \tag{133}$$

for the steady state protein concentration $[P]$, where δ and k_T are the transcription rates according to our convention in Eq. (6). Since the above equation cannot be solved in closed form, in Fig. S4A we illustrate stable (solid white) and unstable (dashed white) numerical solutions.

Conditional LNA solution

By contrast, we now show that, in the case of slow promoter dynamics, the conditional LNA does indeed yield a closed-form solution for the full probability distribution, which is derived in the following. Denoting the set of slow promoter states by $\Pi_{\rm off} = \Pi(n_G = 1, n_{GQ} = 0)$ and $\Pi_{\rm on} = \Pi(n_G = 0, n_{GQ} = 1)$, and making use of promoter conservation, i.e., of $\Pi_{\rm on} = (1 - \Pi_{\rm off})$, we find with Eq. (48) that the promoter distribution is governed by

$$\frac{\partial \Pi_{\rm on}}{\partial t} = -k_{\rm off} \Pi_{\rm on} + k_{\rm on} [Q|1,0](1 - \Pi_{\rm on}), \tag{134}$$

where $[Q|1,0]$ is the conditional expectation of tetramer concentration in the off-state. The conditional expectations obtained from the conditional LNA, Eq. (39), satisfy the simple relation $0 = k_T n_{GQ} + \delta - k_{\rm dp}[P|n_G, n_{GQ}]$, which yields

$$[P|n_G, n_{GQ}] = \frac{k_T n_{GQ} + \delta}{k_{\rm dp}}, \tag{135}$$

together with the relations

$$[Q|n_G, n_{GQ}] = \left(\frac{k_{mQ}}{k_Q}\right)[D|n_G, n_{GQ}]^2 = \left(\frac{k_{mQ}}{k_Q}\right)\left(\frac{k_{mD}}{k_D}\right)^2 [P|n_G, n_{GQ}]^4, \tag{136}$$

for the conditional expectations for the concentrations of dimeric and tetrameric complexes. Using the above expressions together with Eq. (134), we conclude that the probability of finding the promoter in the complexed state is given by

$$\Pi_{\rm on} = \frac{[P|1,0]^4}{K_d + [P|1,0]^4}. \tag{137}$$

Hence, the distribution of protein concentrations is determined by the Gaussian mixture

$$\begin{aligned}\Pi(P) =& \frac{K_d}{K_d + [P|1,0]^4}\frac{1}{\sqrt{2\pi\Sigma_{P|(1,0)}}}\mathrm{e}^{-(P-[P|1,0])^2/(2\Sigma_{P|(1,0)})} \\ &+ \frac{[P|1,0]^4}{K_d + [P|1,0]^4}\frac{1}{\sqrt{2\pi\Sigma_{P|(0,1)}}}\mathrm{e}^{-(P-[P|0,1])^2/(2\Sigma_{P|(0,1)})},\end{aligned} \tag{138}$$

where the conditional protein fluctuations have variance

$$\Sigma_{P|(n_G,n_{GQ})} = \frac{1}{\Omega}[P|n_G, n_{GQ}], \tag{139}$$

as is obtained by solving Eq. (42).

The resulting probability distributions are compared to the deterministic solutions of Eq. (133) in Fig. S4A, in dependence of the basal transcription rate δ. We observe that the modes predicted by the conditional LNA are in qualitative agreement with the solution of the deterministic equation; however, the latter cannot predict its relative weights. Fig. S4A suggests that bimodality is most evident outside of the region of deterministic bistability. We have verified this claim here by stochastic simulation of the full set of elementary reactions in (131) in three regimes in which the solutions of the deterministic Eq. (133) are bistable (blue), marginally bistable (red), and monostable (green), respectively, as shown in Figs. S4B and C. The deterministic theory qualitatively predicts the modal values of the distributions (Fig. S4C, black dashed) within the deterministically bistable region, but misses the stochastic bimodality in the monostable region. By contrast, the conditional LNA accurately predicts the full distributions over the entire parameter space.

References

[1] N. G. Van Kampen. *Stochastic Processes in Physics and Chemistry*. North Holland (1992).

[2] J. Peccoud and B. Ycart. Markovian modeling of gene-product synthesis. *Theor Popul Biol*, **48**, 222 (1995).

[3] V. Shahrezaei and P. S. Swain. Analytical distributions for stochastic gene expression. *Proc Natl Acad Sci*, **105**, 17256 (2008).

[4] P. Bokes, et al. Exact and approximate distributions of protein and mRNA levels in the low-copy regime of gene expression. *J Math Biol*, **64**, 829 (2012).

[5] R. Grima, D. Schmidt, and T. Newman. Steady-state fluctuations of a genetic feedback loop: An exact solution. *J Chem Phys*, **137**, 035104 (2012).

[6] J. Elf and M. Ehrenberg. Fast evaluation of fluctuations in biochemical networks with the linear noise approximation. *Genome Res*, **13**, 2475 (2003).

[7] M. Komorowski, J. Miekisz, and M. P. Stumpf. Decomposing noise in biochemical signaling systems highlights the role of protein degradation. *Biophys J*, **104**, 1783 (2013).

[8] P. Thomas, H. Matuschek, and R. Grima. Intrinsic noise analyzer: A software package for the exploration of stochastic biochemical kinetics using the system size expansion. *PLoS One*, **7**, e38518 (2012).

[9] F. Hayot and C. Jayaprakash. The linear noise approximation for molecular fluctuations within cells. *Phys Biol*, **1**, 205 (2004).

[10] P. Thomas, H. Matuschek, and R. Grima. How reliable is the linear noise approximation of gene regulatory networks? *BMC Genomics*, **14**, S5 (2013).

[11] M. Kærn, et al. Stochasticity in gene expression: from theories to phenotypes. *Nat Rev Genet*, **6**, 451 (2005).

[12] M. Babu, et al. Genetic interaction maps in Escherichia coli reveal functional crosstalk among cell envelope biogenesis pathways. *PLoS Genet*, **7**, e1002377 (2011).

[13] S. Ray and B. G. Lindsay. The topography of multivariate normal mixtures. *Ann Stat*, **33**, 2042 (2005).

[14] S. Ray and D. Ren. On the upper bound of the number of modes of a multivariate normal mixture. *J Multivar Anal*, **108**, 41 (2012).

[15] M. B. Elowitz, et al. Stochastic gene expression in a single cell. *Science*, **297**, 1183 (2002).

[16] C. W. Gardiner et al. *Handbook of Stochastic Methods*, volume 3. Springer Berlin (1985).

[17] M. A. Shea and G. K. Ackers. The O_R control system of bacteriophage lambda: A physical-chemical model for gene regulation. *J Mol Biol*, **181**, 211 (1985).

[18] L. Bintu, et al. Transcriptional regulation by the numbers: models. *Curr Opin Genet Dev*, **15**, 116 (2005).

[19] T. M. Cover and J. A. Thomas. *Elements of information theory*. John Wiley & Sons (2012).

[20] J. J. Tyson, K. C. Chen, and B. Novak. Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Curr Opin Cell Biol*, **15**, 221 (2003).

[21] P. Thomas, et al. Signatures of nonlinearity in single cell noise-induced oscillations. *J Theor Biol*, **335**, 222 (2013).

[22] A. Tiwari, et al. The interplay of multiple feedback loops with post-translational kinetics results in bistability of mycobacterial stress response. *Phys Biol*, **7**, 036005 (2010).

parameter	value	parameter	value
k_{on}	100 (fast), 0.1 (slow)	k_{off}	100 (fast), 0.1 (slow)
$k_0^{(\text{on})}$	40	$k_0^{(\text{off})}$	20
k_1	10	k_2	10
k_3	10	Ω	5

Table S1: Parameter values for the model of binary promoter switching, Eq. (27), used in generating Fig. 2 of the main text. The rates denoted by $k_{\text{on,off}}$ distinguish the two limiting regimes of fast and slow promoter transitions shown in Fig. 2.

parameter	value	parameter	value
a_1, b_1	0.028	a_{-1}, b_{-1}	0.01
a_2	1.25	b_2	1.25
a_3	0.75	b_3	0.75
a_4	10	b_4	10
a_5	10	b_5	10
a_6	1	b_6	1
b_7	1	b_8	57
b_9	1	Ω	300

Table S2: Parameter values for which the mutual information shared between the two protein species is optimal in the mutually repressing gene mechanism, Eq. (88). We note that the strength of translational control can be measured by the inverse mRNA inhibition constant $K_I = \frac{b_7+b_9}{b_8}$, while transcriptional control is proportional to the inverse DNA-protein dissociation constant $K = \frac{a_{-1}}{a_1} = \frac{b_{-1}}{b_1}$. The latter two have been varied in Fig. 3 in the main text via the rate constants b_9, a_1, and b_1, respectively.

parameter	value	parameter	value
$[Y_T]$	1	$[R_T]$	1
k_{off}	0.001	k_{on}	0.001
k_0^{on}	$\frac{140}{N_c}$	k_0^{off}	$\frac{2.2}{N_c}$
k_1	0.1	k_2	100
k_3	1	k_4	2
k_5	1	k_6	0.5
$K_{M,1}, K_{M,2}$	0.01	$K_{M,3}, K_{M,4}$	0.01
Ω	6000		

Table S3: Parameter values for the genetic oscillator used in generating Figs. 4 (main text) and S3. We note that we have accounted for dosage compensation here by rescaling the transcription rates with gene dosage N_c.

parameter	value	parameter	value
Ω	5	α	10
k_{on}	0.1	k_{off}	0.1
k_0^{on}	100	k_0^{on}	200
k_1	10	k_2	10
k_3	10	k_4	10

Table S4: Parameter values for the theoretical induction experiment used in generating Fig. 5 in the main text. The analytical prediction of the protein distribution is constructed using Eq. (129) together with (127) and (128) for forward and reverse induction, respectively.

quantity	$N_c = 1$		$N_c = 2$		
$\boldsymbol{G}$	(1,0)	(0,1)	(2,0)	(0,2)	(1,1)
$[\boldsymbol{Z}\|\boldsymbol{G}]$	$(2.00, 0.51, 0.67)$	$(1.95, 0.26, 0.01)$	as (1,0)	as (0,1)	$(2.00, 0.35, 0.49)$
$\boldsymbol{\lambda}\|\boldsymbol{G}$	$(-11.14,\ -0.014 \pm i0.26)$	$(-2.12, -0.020 \pm i0.61)$	as (1,0)	as (0,1)	$(-5.68, -0.0040 \pm i0.37)$

Table S5: Conditional means and conditional eigenvalues of the gene product concentration vector $\boldsymbol{Z} = (P, Y_P, R_P)$ and the Jacobian $\underline{\mathrm{J}}_{\boldsymbol{G}}$, respectively. We note that the imaginary parts of the eigenvalues correspond to the frequencies observed in the power spectra for the genetic oscillator. The underlying parameter values are given in Table S3.