2012

Relating network rigidity, time scale hierarchies, and expression noise in gene networks

PHYSICAL REVIEW E, COLLEGE PK, v. 85, n. 4, supl. 1, Part 1, pp. 393-399, APR 26, 2012

Relating network rigidity, time scale hierarchies, and expression noise in gene networks

Ovidiu Radulescu
Laboratoire de Dynamique des Interactions Membranaires Normales et Pathologiques, CNRS-UMR 5235, CC107, Université Montpellier II, Place Eugène Bataillon, 34095 Montpellier Cedex 5, France

Guilherme C. P. Innocentini and José Eduardo M. Hornos
Instituto de Física de São Carlos, Universidade de São Paulo, Caixa Postal 369, 13560-São Carlos, São Paulo, Brazil
(Received 4 April 2011; revised manuscript received 26 January 2012; published 26 April 2012)

Fluctuation-dissipation theorems can be used to predict characteristics of noise from characteristics of the macroscopic response of a system. In the case of gene networks, feedback control determines the "network rigidity," defined as resistance to slow external changes. We propose an effective Fokker-Planck equation that relates gene expression noise to topology and to time scales of the gene network. We distinguish between two situations referred to as normal and inverted time hierarchies. The noise can be buffered by network feedback in the first situation, whereas it can be topology independent in the latter.

DOI: 10.1103/PhysRevE.85.041919 PACS number(s): 87.10.Mn, 05.10.Gg, 05.70.Ln, 87.18.Tt

I. INTRODUCTION

Gene expression exhibits a high degree of stochasticity when studied at the level of individual cells. Expression "noise" is a critical, biologically relevant property of genetic circuits in both microbial and eukaryotic cells [1]. Many theoretical and experimental works underline the importance of network architecture and feedback for shaping gene expression noise [2–4]. It is well established that negative feedback increases network rigidity, which is defined as sensitivity of gene expression levels to slow external changes [5]. Several studies [2,3] suggest by extrapolation that noise generated by gene networks is buffered by negative feedback. However, this idea is challenged by recent theoretical [6,7] and experimental [8] work showing that negative feedback is not always noise reducing. In this paper, we relate fluctuations to time scales of the gene expression mechanisms and show that the independence of expression noise on network rigidity follows from an inversion between two such time scales.

In nonequilibrium thermodynamics, fluctuation theories are generally obtained from stochastic equations of motion [9]. The dynamics of gene networks can be described by Markov jump processes [10]. The master equation for such processes reads

$$\frac{\partial p}{\partial t}(X,t) = \sum_{i\in\mathscr{R}}[V_i(X-\gamma_i;\mu)p(X-\gamma_i,t) - V_i(X;\mu)p(X,t)]. \quad (1)$$

Here, the components of the state vector $X \in \mathbb{N}^n$ represent numbers of molecules of various species. The jump vectors $\gamma_i \in \mathbb{Z}^n$ are stoichiometric vectors of biochemical reactions and V_i are reaction rates. By μ, we denote a set of kinetic parameters and external conditions. Since Eq. (1) can be exactly solved only in some simple cases [11], general approximate solutions are essential.

II. TIME SCALES AND EFFECTIVE FOKKER-PLANCK EQUATION

Provided that all species are produced in large numbers, the ω expansion [4] or, equivalently, the central limit theorem [12] leads to diffusion (Fokker-Planck) approximations of the Markov jump process. In multiscale gene networks, this condition is not satisfied because only some of the species (typically proteins) are produced in large numbers. We build on results from [10] to obtain a diffusion approximation in this case as well. We call the large numbers species continuous because their trajectories are continuous in the concentration space, and denote them X_C. Other species are present only in a few copies per cell (the most prominent example is the DNA molecule). We denote these species X_D and call them discrete. The interactions among discrete and continuous species are suitably described by a partition of the reactions in four sets, $\mathscr{R} = \mathscr{R}_D \cup \mathscr{R}_{DC} \cup \mathscr{R}_{CD} \cup \mathscr{R}_C$. The reactions $\mathscr{R}_D$ act on X_D and have rates depending on X_D only. The reactions $\mathscr{R}_C$ act on X_C and have rates depending on X_C only. The reactions $\mathscr{R}_{DC}$ and $\mathscr{R}_{CD}$ act on X_C and X_D, respectively. Their rates depend on both X_D and X_C. The sets $\mathscr{R}_{DC}$ and $\mathscr{R}_{CD}$ can have a nonempty intersection.

The coupling between discrete and continuous variables produces *switched diffusions* [10], i.e., diffusion processes whose parameters (drift and diffusion coefficients) are commanded by discrete variables and change discontinuously at discrete times. Between successive changes of the discrete variables, the continuous species undergo simple diffusion in concentration space (Fig. 1). For slow switching, fluctuations of continuous species that result from switched diffusions can be far from Gaussian [10], whereas for fast switching the switched diffusion can be approximated by a simple diffusion. We provide an effective Fokker-Planck equation that includes the effect of switching and provides a reasonable approximation to switched diffusions in fast and intermediate fast switching regimes.

The definition of various switching regimes follows from the comparison of three time scales of the multiscale stochastic gene networks. The *discreteness time* τ_D represents the average waiting time between two successive reactions acting on continuous species,

$$\tau_D = \left(\sum_{i\in\mathcal{R}_C\cup\mathcal{R}_{DC}}^{r} V_i\right)^{-1}. \quad (2)$$

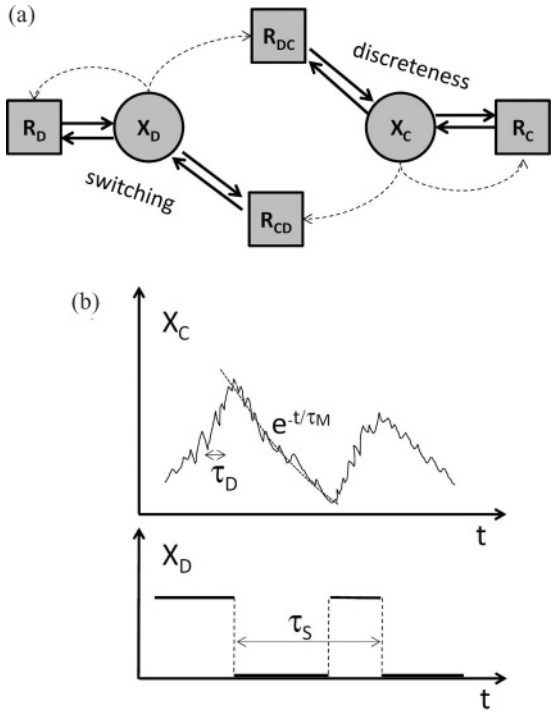

FIG. 1. (a) The partition of species and of the reactions; dotted lines mean that reaction rates depend on the corresponding species. (b) Typical trajectories of continuous and discrete variables: switched diffusions.

For each state of the discrete variable $X_D \in \{1,2,\ldots,M\}$, we define $\tau(k)$ as the average lifetime of the state $X_D = k$, i.e., the average time during which the system remains in this state before transiting to any other, different state. We define the *switching time* τ_S as the sum of these lifetimes,

$$\tau_S = \sum_{k=1}^{M} \tau(k). \tag{3}$$

The switching time sums the time scales of the reactions $\mathcal{R}_D$ and $\mathcal{R}_{CD}$. Finally, let us consider that the averaged (coarse-grained) drift dynamics of the continuous variable has an attractor. The *macroscopic time* τ_M is the relaxation time of the averaged system toward this attractor. For simplicity, we consider that there is only one such time scale, or in the case of many time scales, choose the fastest one.

A simple example of degradation reaction $X \xrightarrow{k} \emptyset$ shows that a single reaction introduces two time scales. The discreteness time $\tau_D = (kX)^{-1}$ is always much smaller than the macroscopic time $\tau_M = k^{-1}$. Without discrete species, the *normal two-time hierarchy* condition $\tau_D \ll \tau_M$ allows obtention of hydrodynamic equations for the macroscopic, continuous variables by coarse graining [13]. Discrete species introduce switching. Contrary to τ_D, the switching time τ_S can be independent from τ_M because it depends on reactions acting on discrete species, whereas τ_M depends on reactions acting on continuous species. This property of τ_S can lead to *inverted time hierarchies* when $\tau_D \ll \tau_S \ll \tau_M$ or $\tau_D \ll \tau_M \ll \tau_S$.

Let us introduce two scale parameters $\eta = \tau_D/\tau_M$ and $\epsilon = \tau_S/\tau_M$ and rescale the rates and species numbers $x_c = \eta X_c$, $v_i(X_D) = \epsilon V_i(X_D,x_c)$, $i \in \mathcal{R}_{CD} \cup \mathcal{R}_D$, $v_i(X_D,X_C) = \eta V_i(X_D,x_c)$, and $i \in \mathcal{R}_{DC} \cup \mathcal{R}_C$. Next, we obtain a diffusion approximation for x_c. As mentioned, η is always small and represents the inverse of a characteristic copy number of continuous species. We consider that ϵ is also small, meaning that $\tau_M \gg \tau_S$. This is enough to cover the crossover between discreteness and switching effects that we study in this paper, including the inverted hierarchy $\tau_D \ll \tau_S$. The case $\tau_M \ll \tau_S$ leading to non-Gaussian processes is covered by techniques in [10].

There are two contributions to the diffusion of continuous variables x_c. The discreteness term acts at constant X_D and results from the rapid succession of reactions $\mathcal{R}_C \cup \mathcal{R}_{DC}$. The discreteness term can be easily obtained by ω expansion [4,10,12]. The switching term results from changes of X_D by reactions $\mathcal{R}_D \cup \mathcal{R}_{CD}$. The system remains an average time $\tau(X_D;x_c)$ in the state X_D, then jumps to another state. Between jumps, the discreteness drift pushes x_c a distance in concentration space proportional to $\tau(X_D;x_c)$, in a direction that changes randomly each time the state changes. The corresponding diffusion term cannot be obtained by ω expansion and needs special treatment. In the Appendix, we obtain the following effective Fokker-Planck equation for x_c, by averaging the drift and by summing the discreteness and switching contributions to diffusion:

$$\frac{\partial p}{\partial t}(x_c,t) = -\nabla_{x_c}[\bar{b}(x_c)p(x_c,t)] + \frac{1}{2}\nabla^2_{x_c}\left\{\left[\sigma_d^2(x_c) + \sigma_s^2(x_c)\right]p(x_c,t)\right\}, \tag{4}$$

where

$$\begin{aligned}
\left(\sigma_d^2\right)_{kl}(x_c) &= \eta \sum_{i\in\mathcal{R}_C\cup\mathcal{R}_{DC}} \left(\gamma_i^C\right)_k \left(\gamma_i^C\right)_l \bar{v}_i(x_c),\\
\left(\sigma_s^2\right)_{kl}(x_c) &= \epsilon\tau_M \sum_{X_D} \left\{\sum_{i\in\mathcal{R}_{DC}} (\gamma_i)_k [v_i(X_D,x_c) - \bar{v}_i(x_c)]\right\}\\
&\quad \times \left\{\sum_{i\in\mathcal{R}_{DC}} (\gamma_i)_l [v_i(X_D,x_c) - \bar{v}_i(x_c)]\right\}\\
&\quad \times \rho(X_D;x_c)\hat{\tau}(X_D;x_c),\\
\bar{b}(x_c) &= \sum_{i\in\mathcal{R}_C\cup\mathcal{R}_{DC}} \gamma_i \bar{v}_i(x_c), \bar{v}_i = \sum_{X_D} v_i \rho(X_D),
\end{aligned} \tag{5}$$

where $\rho(X_D;x_c)$ and $\hat{\tau}(X_D;x_c) = \tau(X_D;x_c)/\sum_{X_D}\tau(X_D;x_c)$ are the steady-state probability and normalized lifetime of the state X_D, respectively (these can be computed from the reactions $\mathcal{R}_{CD}$ and $\mathcal{R}_D$). The discreteness term σ_d^2 accounts for the finite size of the jumps induced by reactions acting directly on x_c. It is well known that σ_d^2 vanishes like η in the thermodynamic limit [4,10,12]. The switching term σ_s^2 accounts for fluctuations transmitted from discrete variables to continuous variables, covering situations described as the bursting or the telegraph scenario [1]. The ω expansion [4] overlooks or underestimates this term because this expansion is valid only when all the species are in large numbers.

Let us consider that the averaged drift has a point attractor x_c^* [stable solution of $\bar{b}(x_c;\mu) = 0$]. Then, for small ϵ,η and for initial data close to x_c^*, we can linearize $\bar{b}$ and obtain an

Ornstein-Uhlenbeck approximation:

$$\frac{\partial p}{\partial t}(x_c,t) = -\nabla_{x_c}[J(x_c^*)(x_c - x_c^*)p(x_c,t)] + \frac{1}{2}\nabla_{x_c}^2[\sigma^2(x_c^*)p(x_c,t)], \quad (6)$$

where J is the Jacobian matrix of elements $J_{ij}(x_c^*) = \partial_{x_{c,j}}\bar{b}_i(x_c^*)$ and $\sigma^2 = \sigma_s^2 + \sigma_d^2$.

The covariance matrix Λ [$\Lambda_{i,j} = \mathrm{cov}(x_i^c, x_j^c)$] giving the steady-state fluctuations of x_c results from (6) by standard methods [14] as the unique symmetric solution of the equation

$$\Lambda(-J^T) + (-J)\Lambda = \sigma_d^2 + \sigma_s^2. \quad (7)$$

Several authors [4,15] interpreted (7) as a fluctuation-dissipation theorem. Indeed, (7) relates steady-state fluctuations to the Jacobian J. J is a measure of network rigidity and allows one to compute the sensitivity of the steady state to external forces (changes of the parameters μ) [5]. However, (7) contains also the diffusion matrix σ^2, sometimes interpreted as nonequilibrium "temperature" [9,15]. The nonequilibrium temperature can have unusual properties. It can depend on which species are observed [15] or change with the steady state [9]. In our Markovian approach, the discreteness and the switching terms have contrasting behavior with respect to changes of the steady state and of J. This means that a crossover of the two terms, resulting from an inversion in the time scale hierarchy, changes the dependence of fluctuations on rigidity. Using (5) and (7), one can find, for a particular model, the dependence of discreteness and switching terms on the network rigidity, and analyze the consequences of the time scale inversion. Next we study a simple two-state self-repressed promoter. We use this exactly solvable model to test our general approach.

III. SELF-REPRESSED BACTERIUM PROMOTER

This is the simplest, though nontrivial, model of a gene network with negative feedback for which we have a full analytic solution [11]. The discrete variable has two states $X_D = 0,1$ corresponding to the repressed and unrepressed promoter, respectively. The continuous variable is the number of proteins X_C and can take any positive integer value n. When $X_D = 1$, proteins are produced with rate g and degraded with rate kn. When $X_D = 0$, production falls to χg, $\chi < 1$. The transition between active and inactive states is commanded by the protein itself with a rate hn; the opposite transition is spontaneous with a constant unbinding rate f. The master equation is

$$\begin{aligned}\frac{\partial p_n^1}{\partial t} &= g\left[p_{n-1}^1 - p_n^1\right] + k\left[(n+1)p_{n+1}^1 - np_n^1\right] \\ &\quad - nhp_n^1 + fp_n^0, \\ \frac{\partial p_n^0}{\partial t} &= \chi g\left[p_{n-1}^0 - p_n^0\right] + k\left[(n+1)p_{n+1}^0 - np_n^0\right] \\ &\quad + nhp_n^1 - fp_n^0. \end{aligned} \quad (8)$$

Equations (8) can be reshaped in the form (1). The five reactions of the model can be partitioned as follows. $\mathscr{R}_{CD}$ and $\mathscr{R}_D$ include reactions of rates $V_1 = X_D hn$ and $V_2 = (1 - X_D)f$ and jump vectors $\gamma_1 = (-1,0)$ and $\gamma_2 = (1,0)$, respectively. $\mathscr{R}_{DC}$ includes reactions of rates $V_3 = X_D g$ and $V_4 = (1 - X_D)\chi g$ and jump vectors $\gamma_3 = (0,1)$ and $\gamma_4 = (0,1)$. $\mathscr{R}_C$ contains only the degradation reaction $V_5 = kn$, $\gamma_5 = (0,-1)$. We rescale time and protein numbers $\tau = t/\tau_M = tk$, $x = n/N_0$, $N_0 = g/k$. The rescaled model depends on four parameters: $n_r = f/(hN_0)$, χ, $\eta = 1/N_0$, and $\epsilon = k/f(1 + n_r/x)$. The first two are feedback parameters quantifying the affinity of the repressor to DNA, and the residual transcription activity of the promoter, respectively. n_r has been used in other studies relating noise and feedback (it is the inverse of the parameter α from [7]). The remaining parameters are time scale ratios. In these variables, (4) reads

$$\frac{\partial p}{\partial \tau}(x,t) = -\nabla_x[b(x)p(x,t)] + \frac{1}{2}\nabla_x^2\left\{\left[\sigma_s^2(x) + \sigma_d^2(x)\right]p(x,t)\right\}, \quad (9)$$

where

$$\begin{aligned} b(x) &= \rho(x) + \chi[1 - \rho(x)] - x \text{ is the averaged drift,} \\ \rho(x) &= \frac{n_r}{n_r + x} \text{ is the probability that } X_D = 1, \\ \sigma_d^2(x) &= \eta\{x + \rho(x) + [1 - \rho(x)]\chi\}, \\ \sigma_s^2(x) &= 2\epsilon\{\rho(x)[1 - \rho(x)](1 - \chi)\}^2. \end{aligned} \quad (10)$$

The drift $b(x)$ has a global attractor x^* which is the unique solution of the nonlinear equation $x = \rho(x) + \chi[1 - \rho(x)]$. We define the rigidity of the system as minus the Jacobian of the drift, $R = -db(x)/dx$ for $x = x^*$. Using (7) with $J = db(x)/dx$, we get the steady-state noise variance,

$$\mathrm{Var}(x) = \frac{\sigma_d^2(\eta, n_r, \chi, x^*) + \sigma_s^2(\epsilon, n_r, \chi, x^*)}{2R(n_r, \chi, x^*)}, \quad (11)$$

where the rigidity is

$$R(n_r, \chi, x) = 1 + \frac{n_r(1 - \chi)}{(n_r + x)^2}. \quad (12)$$

We check the validity of our approximations by direct comparison of the variance and the average obtained from relations (10) and (12) with the exact solution presented in [11]. As shown in Figs. 2(a) and 2(b), the agreement is excellent when $\epsilon, \eta < 1$, and less good for $\epsilon > 1$ when the diffusion approximation fails. Using (10) and (12), and straightforward algebra, we find the dependence of discreteness and switching terms on the rigidity,

$$\sigma_d^2 = 2\eta x, \quad \sigma_s^2 = 2\epsilon(R - 1)^2 x^2. \quad (13)$$

We note from (13) that the two "temperature" terms σ_d^2 and σ_s^2 have contrasting behavior. Further insight into this difference is obtained by investigating families of parameters. Two cases are particularly interesting.

Fixed average expression. We impose the average expression x. The free parameter is χ ($0 \leqslant \chi \leqslant x$) or, equivalently, n_r or R. This parametrization can be used to compare different promoters that have the same average expression, but different residual activities and repressor affinities. We decide to compare the promoters by the value of their rigidity. To this aim, we use (10) and (12) to express ρ, n_r, and χ as functions of R. Given x, a more "rigid" promoter will have a smaller residual activity and a more affine repressor. According to (13), at fixed

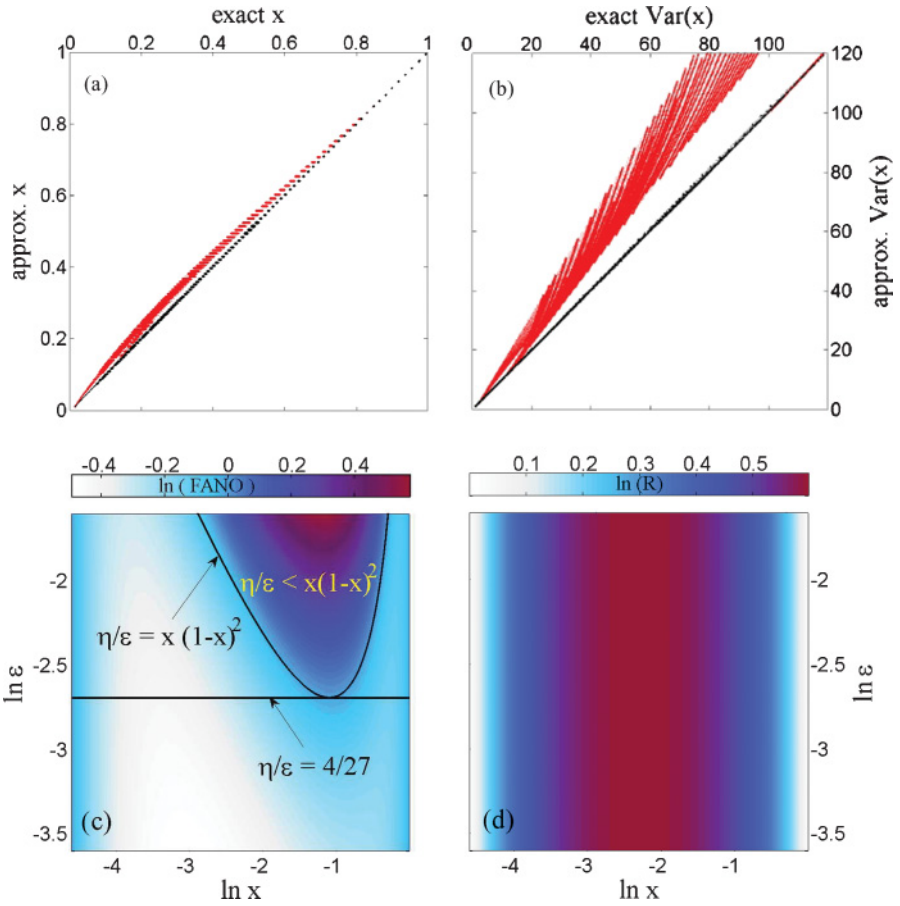

FIG. 2. (Color online) Comparison of exact and approximate (a) average and (b) variance of noise for $\chi = 0.001, 0.01, 0.1$, $\chi < x < 1$, $\eta = 0.01$, when $0.02 < \epsilon < 0.2$ (approximation applies, black dots) and when $1 < \epsilon < 2$ [approximation fails, red (grey) dots]. Comparison of (c) Fano factor and (d) rigidity dependence on x and ϵ, for $\chi = 0.01$, $0.02 < \epsilon < 0.2$. The Fano factor is maximum close to maximum rigidity in a region in parameter space where switching is dominant, $\eta/\epsilon < x(1-x)^2$.

x, the discreteness term is constant, whereas the switching term increases quadratically with the rigidity. The corresponding contributions to the variance follow from (11) and behave like $1/R$ and $(R-1)^2/R$, respectively. The discreteness and switching terms become equal for a critical rigidity R^* that satisfies $\eta/\epsilon = x(R^*-1)^2$. However, the values of rigidity are bounded. Using (12), we get $R = 1 + \frac{(x-\chi)(1-x)}{(1-\chi)x}$ from which $1 \leqslant R \leqslant 2-x$. This leads to an x-dependent condition for crossover, namely, $\eta/\epsilon < x(1-x)^2$. If this condition is not fulfilled, i.e., for rapid switching, then no crossover is possible. The discreteness term is dominant and the variance scales like $1/R$: more rigid promoters are less noisy. If the condition is fulfilled, inverted hierarchy behavior is possible. When switching terms become dominant, the variance scales like $(R-1)^2/R$: more rigid promoters are more noisy.

Fixed residual activity. Here we impose χ. The free parameter can be either x ($\chi \leqslant x \leqslant 1$) or n_r (n_r is strictly increasing with x). This parametrization allows one, for the same promoter, to study the effect of changes of the affinity of the repressor. In bacteria, such an adaptation process could result from stress or metabolic changes such as a switch in carbon source [8]. From (12) and (13), we get

$$\sigma_d^2 = 2\eta x, \quad \sigma_s^2 = 2\epsilon \left[\frac{(x-\chi)(1-x)}{(1-\chi)}\right]^2. \tag{14}$$

In this case, the variance is no longer appropriate to quantify noise amplitudes because the average is not constant. We can use instead the coefficient of variation, $\mathrm{CV} = \sqrt{\mathrm{Var}(x)}/x$, and the Fano factor, $F = N_0 \mathrm{Var}(x)/x$. From (14), having a crossover between discreteness and switching terms for some x boils down to having at least one positive root of the polynomial equation $\sigma_d^2(x;\eta) = \sigma_s^2(x;\chi,\epsilon)$. A necessary and sufficient condition for this reads $\eta/\epsilon < \psi^*(\chi)$, where $\psi^*(\chi)$ is a polynomial in χ satisfying $\psi^*(\chi) \approx 4/27$ for small χ. This condition places the inverted hierarchy behavior in a domain where ϵ is large. Indeed, Figs. 2(c) and 2(d) show that in this domain, the Fano factor has a maximum close to a place where also the rigidity is maximum. Far outside this domain, the Fano factor has small variation and CV scales roughly like $1/\sqrt{x}$. For bacterial promoters, $N_0 \sim 10^4$ and $k = 10^{-4}\ \mathrm{s}^{-1}$ for proteins, and $N_0 \sim 10^2$ and $k = 10^{-2}\ \mathrm{s}^{-1}$ for mRNA; normal time hierarchies correspond to switching times of less than 10 s.

IV. DISCUSSION

Theories ascertaining that network topology is the main factor that shapes noise in gene networks [2] can be misleading, even with amendments, such as considering that the extrinsic rather than the intrinsic noise is feedback dependent [3]. We show that in multiscale biochemical networks with both discrete and continuous variables, time scales of the molecular interactions can dictate the behavior of the molecular noise. In particular, negative feedback fails to buffer fluctuations of continuous components whose production depends on discrete slow processes.

We have obtained an effective Fokker-Planck equation that applies not only to situations when all chemical species in the network have large copy numbers, but also to situations when some of the species have small numbers and generate switching effects. For such effective diffusions, the Ornstein-Uhlenbeck result, also known as "fluctuation-dissipation," implies that fluctuation amplitudes and network rigidity are inversely proportional. However, the fluctuation amplitudes are also proportional to a "temperature" factor. The switching contribution to this temperature can increase with the network rigidity, so when it becomes dominant, as in the case of time hierarchy inversion, statically "rigid" networks can be very noisy.

Our findings offer mathematical substance to the counterintuitive result [6–8] that self-repression can be a noise maintaining mechanism, possibly important in bet-hedging adaptation strategies [16]. We estimate that switching times of ten seconds or larger should lead to mRNA and protein expression noise that is not buffered by negative feedback. Because these time scales are not rare in repressed bacterial or eukaryotic promoters, our results should be testable by accurate expression quantification experiments on large collections of promoters with different characteristics. A recent experimental study of a few carbon catabolic operons in Bacillus subtilis [8] showed that the noisiest repressed promoter is self-repressed. This result can be explained by the inverse proportionality of the switching time and of the transcriptional bursting amplitude with the concentration of active repressor, which, for self-repressed promoters, is low in the repressed state. More generally, our Eqs. (5) and (4) show the importance of long-lived discrete states for the switching contribution to noise. Examples of such long-lived states can be found not only in prokaryote transcription regulation, but also during formation of transcriptional complexes on the eukaryote *TATA* box [16]). Mechanisms that change the lifetime of such states can serve for effective noise tuning, with potential applications in synthetic biology.

In subsequent work, the validity of our formulas will be extended. Several improvements of the approximation are possible. First, a Stratonovich drift correction, like in [17], would be appropriate to take into account the "look into the future" property of diffusion approximations. Second, higher order approximations are needed to account for large fluctuations of the continuous variables. Moreover, (5) will be used to understand fluctuations in molecular systems with multiple discrete states, such as complex promoters, molecular motors, and genetic and protein transport networks.

ACKNOWLEDGMENT

This work was partially supported by the USP-COFECUB Grant No. 115/09.

APPENDIX

In this appendix, we obtain an effective Fokker-Planck equation (diffusion approximation) for the stochastic dynamics of the continuous species. Let $X_c(t) \in \mathbb{N}^n$ be the state of continuous species at time t. Let us define the rescaled process $x_c(t) = \eta X_c(t)$. $x_c(t)$ is a diffusion process in the high dimensional space of the concentrations, if and only if the following properties are satisfied (see, for instance, [18]):

(i) The increments $\Delta x_c(t;\Delta t) = x_c(t+\Delta t) - x_c(t)$ on nonoverlapping time intervals are independent and normally distributed for all t.

(ii) $\langle \Delta x_c(t;\Delta t)\rangle = b(x_c,t)\Delta t + o(\Delta t)$.

(iii) $\mathrm{Cov}[\Delta x_c(t;\Delta t)] = \sigma^2(x_c,t)\Delta t + o(\Delta t)$, where Cov is the variance-covariance matrix, σ^2 (a symmetric, positive matrix) is the diffusion matrix, and the vector field b is the drift.

The probability distribution $p(x_c,t)$ of such a diffusion process will satisfy the Fokker-Planck equation,

$$\frac{\partial p}{\partial t}(x_c,t) = -\nabla_{x_c}[b(x_c,t)p(x_c,t)] + \frac{1}{2}\nabla^2_{x_c}[\sigma^2(x_c,t)p(x_c,t)]. \tag{A1}$$

Here we will use (ii) and (iii) to find the drift and the diffusion parameters for an effective (coarse-grained) Fokker-Planck equation. We discuss the situation when the switching time τ_S and the discreteness time τ_D are much smaller than the macroscopic time τ_M (see the main text for the definition of these times), and coarse grain the process on times $\Delta t \gg \tau_S, \tau_D$. Δt is chosen sufficiently small (much smaller than τ_M) such that increments $\Delta x_c(t;\Delta t)$ are small.

Additionally, let us consider that given x_c, the stochastic dynamics of X_D is described by a discrete Markov process and that this process is ergodic (i.e., any two discrete states are connected by at least one path of nonzero probability transitions). Then, on the time scale $\tau_S \ll \Delta t$, X_D reaches the unique steady-state probability distribution $\rho(X_D;x_c)$. The probabilities $\rho(X_D;x_c)$ are found by solving the stationary master equation satisfied by X_D at fixed x_c (this is defined by the reactions $\mathcal{R}_D$ and $\mathcal{R}_{CD}$). Let us consider a trajectory of the process at fixed x_c and functions $f(X_D)$ depending on the discrete variable. We denote by $\Delta t_{X_D=k}$ the total time spent by the discrete variable in the state k. The Birkhoff-Khinchin formulation of the ergodic theorem (well known in statistical physics) says that for an ergodic system, the time average of an integrable function exists for almost all trajectories and is equal to the phase space average of the same function with respect to the unique equilibrium steady-state probability. The ergodic theorem applied to the Markov process X_D (at fixed x_c) reads

$$\sum_k f(k)\Delta t_{X_D=k} \Big/ \sum_k \Delta t_{X_D=k} = \sum_k f(k)\rho(k;x_c). \tag{A2}$$

Let us compute the increment $\Delta x_c(t;\Delta t)$ of x_c between t and $t+\Delta t$. In order to do so, we decompose the interval $[t,t+\Delta t]$ into a number $N(t,\Delta t)$ of successive, nonoverlapping subintervals $I_i = (t_i, t_i+\Delta t_i)$, $i \in [1,N(t,\Delta t)]$ along which the value of X_D is constant. To simplify notations, we denote the possible values of X_D by the integers $1,2,\ldots,M$.

Let M_{li} be the number of reactions of type $l \in \mathcal{R}_C \cup \mathcal{R}_{DC}$ that occur in the interval I_i. Up to small corrections of order $o(\Delta t)$ (because x_c is not strictly constant between t and $t+\Delta t$), M_{li} are independent Poisson variables of mean $\eta^{-1}v_l[X_D(t_i),x_c]\Delta t_i$, where $v_l = \eta V_l$ (see [12], for instance).

Then, we have

$$\Delta x_c(t;\Delta t) = \eta \sum_{i=1}^{N(t,\Delta t)} \sum_{l\in\mathcal{R}_C\cup\mathcal{R}_{DC}} M_{li}\gamma_l^C + o(\Delta t), \quad \text{(A3)}$$

where γ_l^C is the projection of the stoichiometric vector onto the continuous species. Straightforward algebra leads to the mean increment,

$$\langle \Delta x_c(t;\Delta t)\rangle = \sum_{k=1}^{M} b[x_c(t),k]\Delta t_{X_D=k} + o(\Delta t), \quad \text{(A4)}$$

where $b[x_c(t),k]$ is the drift of the continuous variables when the discrete variables are in the state k,

$$b(x_c,k) = \sum_{l\in\mathcal{R}_{DC}} \gamma_l^C v_l(x_c,k) + \sum_{l\in\mathcal{R}_C} \gamma_l^C v_l(x_c). \quad \text{(A5)}$$

Applying the ergodic theorem (A2), we find

$$\langle \Delta x_c(t;\Delta t)\rangle = \bar{b}[x_c(t)]\Delta t + o(\Delta t), \quad \text{(A6)}$$

where $\bar{b}(x_c) = \sum_{k=1}^{M} b(x_c,k)\rho(k;x_c)$.

Similarly, we compute the covariance matrix of the increment. By the law of total covariance, it contains two terms,

$$\mathrm{Cov}[\Delta x_c(t;\Delta t)] = \mathrm{Cov}[\Delta x_c(t;\Delta t)]^D + \mathrm{Cov}[\Delta x_c(t;\Delta t)]^S. \quad \text{(A7)}$$

The discreteness term corresponds to the Poissonian fluctuations of the number of reactions at fixed X_D, averaged over the values of X_D,

$$\mathrm{Cov}[\Delta x_c(t;\Delta t)]^D_{m,n} = \eta \sum_{k=1}^{M} \sum_{l\in\mathcal{R}_C\cup\mathcal{R}_{DC}} (\gamma_l)_m(\gamma_l)_n v_l[k,x_c(t)] \times \rho[k;x_c(t)]\Delta t + o(\Delta t). \quad \text{(A8)}$$

The switching term corresponds to random changes of the drift vector $b(x_c,X_D)$ induced by changes of X_D,

$$\mathrm{Cov}[\Delta x_c(t;\Delta t)]^S_{m,n} = \sum_{k=1}^{M} \{b[x_c(t),k] - \bar{b}[x_c(t)]\}_m \{b[x_c(t),k] - \bar{b}[x_c(t)]\}_n \Delta^2 t_{X_D=k} + o(\Delta t), \quad \text{(A9)}$$

where $\Delta^2 t_k = \sum_{X_D(t_i)=k} (\Delta t_i)^2$.

Applying again the ergodic theorem (A2), we find

$$\Delta^2 t_{X_D=k} = \Delta t \rho(k)\tau(k), \quad \text{(A10)}$$

where $\tau(k) = \langle \Delta t_i\rangle_{X_D=k}$ is the average lifetime of the state $X_D = k$.

Equations (5) and (4) follow directly from (A8) through (A10).

[1] B. B. Kaufmann and A. van Oudenaarden, Curr. Opin. Genet. Dev. **17**, 107 (2007).
[2] A. Becskei and L. Serrano, Nature (London) **405**, 590 (2000).
[3] Y. Dublanche, K. Michalodimitriakis, N. Kümmerer, M. Foglierini, and L.Serrano, Mol. Syst. Biol. **2**, 41 (2006).
[4] J. Paulsson, Nature (London) **427**, 415 (2004).
[5] O. Radulescu, S. Lagarrigue, A. Siegel, P. Veber, and M. Le Borgne, J. R. Soc. Interface **3**, 185 (2006).
[6] D. Stekel and D. Jenkins, BMC Syst. Biol. **2**, 6 (2008).
[7] T. Marquez-Lago and J. Stelling, Biophys. J. **98**, 1742 (2010).
[8] M. L. Ferguson, D. Le Coq, M. Jules, S. Aymerich, O. Radulescu, N. Declerck, and C. A. Royer, Proc. Natl. Acad. Sci. **109**, 155 (2012).
[9] J. Prost, J-F. Joanny, and J. M. R. Parrondo, Phys. Rev. Lett. **103**, 90601 (2009).
[10] A. Crudu, A. Debussche, and O. Radulescu, BMC Syst. Biol. **3**, 89 (2009).
[11] J. E. M. Hornos, D. Schultz, G. C. P. Innocentini, J. Wang, A. M. Walczak, J. N. Onuchic, and P. G. Wolynes, Phys. Rev. E **72**, 051907 (2005).
[12] Daniel T. Gillespie, J. Chem. Phys. **113**, 297 (2000).
[13] R. Kubo, M. Toda, and N. Hashitsume, *Statistical Physics II* (Springer, Berlin, 1991), Chap. 4.
[14] M. C. Wang and G. E. Uhlenbeck, Rev. Mod. Phys. **17**, 323 (1945).
[15] T. Lu, J. Hasty, and P. G. Wolynes, Biophys. J. **91**, 84 (2006).
[16] L. Lopéz-Maury, S. Marguerat, and J. Bähler, Nature Rev. Genetics **9**, 583 (2008).
[17] A. Rocco, Phys. Biol. **6**, 016002 (2009).
[18] A. T. Barucha-Reid, *Elements of the Theory of Markov Processes and Their Applications* (McGraw-Hill, New York, 1960).