Noise in Gene Regulatory Networks

Ioannis Lestas, Member, IEEE, Johan Paulsson, Nicholas E. Ross, and Glenn Vinnicombe

Abstract—Life processes in single cells and at the molecular level are inherently stochastic. Quantifying the noise is, however, far from trivial, as a major contribution comes from intrinsic fluctuations, arising from the randomness in the times between discrete jumps. It is shown in this paper how a noise-filtering setup with an operator theoretic interpretation can be relevant for analyzing the intrinsic stochasticity in jump processes described by master equations. Such interpretation naturally exists in linear noise approximations, but it also provides an exact description of the jump process when the transition rates are linear. As an important example, it is shown in this paper how, by addressing the proximity of the underlying dynamics in an appropriate topology, a sequence of coupled birth–death processes, which can be relevant in gene expression, tends to a pure delay; this implies important limitations in noise suppression capabilities. Despite the exactness, in a linear regime, of the analysis of noise in conjunction with the network dynamics, we emphasize in this paper the importance of also analyzing dynamic behavior when transition rates are highly nonlinear; otherwise, steady-state solutions can be misinterpreted. The examples are taken from systems with macroscopic models leading to bistability. It is discussed that bistability in the deterministic mass action kinetics and bimodality in the steady-state solution of the master equation neither always imply one another nor do they necessarily lead to efficient switching behaviors: the underlying dynamics need to be taken into account. Finally, we explore some of these issues in relation to a model of the lac operon.

I. INTRODUCTION

THE importance of intrinsic noise within gene regulatory networks has long been appreciated within the mathematical biology community with increasing interest in recent years (e.g., [1]–[4]). The “stochastic chemical kinetics” that arise due to random births and deaths of individual molecules give rise to jump Markov processes, which can be analyzed by means of master equations and simulated with stochastic simulation algorithms [5]. This kind of intrinsic fluctuation is undoubtedly a major factor that needs to be taken into account when one attempts to address the functionality or even characterize the optimality of the regulatory network under study: large variability is likely to deteriorate regulation performance and increase metabolic burden.

Nevertheless, if one tries to adopt a control theoretic approach in order to reverse engineer or justify an optimal (in some sense) control strategy for a particular regulation problem in a biological network, this intrinsic stochasticity turns out to introduce a major complication in the mathematical analysis. Unlike more conventional approaches in control theory, the noise that is introduced to the system is not additive, but is a result of the randomness in the time between jumps, with the jump rates depending on other species’ concentrations. In addition, the feedback mechanism itself will have to be a jump Markov process, since it will correspond to a number of other chemical reactions.

Finding ways to analyze this kind of intrinsic noise is a problem that has received considerable attention by physicists from an early stage. Those go back to Einstein’s relations [6], indicating the correspondence of the damping coefficient in a fluid with the mean square of the fluctuations. This is a special case of the fluctuation-dissipation theorem which relates macroscopic parameters with intrinsic variability. An exact manifestation can be given for an Ornstein–Uhlenbeck process, where it is known that the probability density function satisfies a linear Fokker–Planck equation. In this case, the diffusion and drift coefficients are related, by means of a Lyapunov equation, to the steady-state covariance matrix.

One of the main aims in this paper is to provide a link between operator theoretic approaches employed in robust control theory and methods of quantifying the intrinsic noise in jump processes described by master equations. For an Ornstein–Uhlenbeck process, the fact that the Fokker–Planck equation for the probability density function leads to a Lyapunov equation for the covariance matrix is not surprising from a systems theoretic perspective, since the noise in this case is additive and white, and the steady-state variance is hence given by the $H_2$ norm of a corresponding system driven by the noise. The $H_2$ norm is known to be related to the controllability or observability gramian, which in turn satisfies a Lyapunov equation involving the system dynamics.

Such an analytically tractable Fokker–Planck approximation for the master equation can be obtained by means of van Kampen’s $\Omega$ expansion [7], in the limit of a large volume and large number of molecules. However, what is more striking is the fact that the corresponding Lyapunov equation for the covariance matrix is exact, without requiring an assumption of a limiting process, when the transition rates are linear functions of the species concentrations. Moreover, it turns out in this case that the jump process has the same power spectrum with a corresponding linear system driven by additive white noise (even though the two processes are not identical). This implies that more conventional tools from control theory, such as Bode integral formulas and norm-type cost functions to quantify robustness and performance, could contribute in reverse engineering the role of feedback. As an example of such an operator theoretic approach, we show in this paper how, by addressing the proximity of the underlying dynamics within the graph topology, a sequence of coupled birth–death processes, which are relevant in gene expression, tends to a pure delay in the limit.

Manuscript received January 23, 2007; revised September 1, 2007.

I. Lestas, N. E. Ross, and G. Vinnicombe are with the Department of Engineering, University of Cambridge, Cambridge CB21PZ, U.K. (e-mail: ic120@cam.ac.uk; ner25@cam.ac.uk; gv@eng.cam.ac.uk).

J. Paulsson is with the Department of Systems Biology, Harvard University, Boston, MA 02115 USA (e-mail: johan.paulsson@hms.harvard.edu).

Digital Object Identifier 10.1109/TAC.2007.911347

© 2008 IEEE
of a large number of such processes. This illustrates how delays can arise in feedback configurations (thus implying important limitations in noise-suppression capabilities) and complexity can be reduced when modeling stochastic gene expression, since each individual step during transcription or translation is itself a stochastic jump process.

When transition rates are nonlinear, an exact solution to the master equation is generally not possible, and different approximation schemes are often used for analysis. For example, the exactness in a linear regime (even far from the macroscopic limit) of the noise-filtering setup described above can provide a significant insight when transition rates are weakly nonlinear or fluctuations are small; also, mass action kinetics and nonlinear Langevin approximations can be formally justified in large molecule numbers [8], [9]. If intrinsic stochasticity is not small in a highly nonlinear regime (such as in cases where the corresponding macroscopic model leads to bistability and limit cycles), the underlying dynamics should carefully be taken into account in order to avoid potential pitfalls and misinterpretations of steady-state solutions. In particular, bistability is often encountered in deterministic models, and this is associated with the behavior of a genetic switch. In the lac operon, for example, it has been conjectured that this provides hysteresis during transitions between the induced and uninduced states of the cell. However, a phenomenological paradox is that the corresponding master equation has in most physical models a unique stationary solution [7], [10]; bistability of the deterministic mass action kinetics is thus often associated with bimodality of the steady-state solution of the master equation. It is discussed in this paper that the steady-state solutions of either the master equation or the deterministic macroscopic model neither always imply one another, nor do they necessarily lead to efficient switching behaviors. For the latter, one needs to consider the transient response of the system through the expected time it takes to have a transition from one of the equilibrium points to the other and compare this with the time scale of the underlying dynamics.

If, for example, the former is of the order of many thousands of years, as is the case for a transistor in a conventional microprocessor, then this is indeed a good switch! On the other hand, if this expected time is a couple of seconds, then obviously the switch is not effective despite the bistability predicted in the deterministic model. At the same time, in small molecule numbers, the mass action kinetics provide a less accurate description of the jump process, and a bimodal probability distribution can be observed even when the deterministic model has only one stable equilibrium.

This paper is structured as follows. We discuss first how a noise-filtering setup can become relevant for a jump process described by a master equation. This is used as a basis for illustrating how a sequence of coupled birth–death processes relevant in gene expression tends to a pure delay. We then discuss the effects of noise in cases of bistable/bimodal systems. The lac operon is finally discussed as an example.

II. THEORY

Here, we discuss cases where an exact solution to the master equation is possible [12], analogies and differences with linear noise approximations [7] are illustrated, and an operator theoretic interpretation of these results is pointed out.

We consider a biochemical system with \( n \) molecular species and \( m \) elementary reactions

\[
 x(t) \xrightarrow{W_i(x(t))} x(t) + r_i, \quad i = 1, 2, \ldots, m
\]

where random variable \( x(t) \) is the vector of molecule numbers at time \( t \), i.e., \( x(t) = [x_1(t), x_2(t), \ldots, x_n(t)]^T \), where \( x_j(t) \) denotes the number of molecules of species \( j \) at time \( t \). \( W_i(x(t)) \) is the rate of reaction \( i \), and the \( j \)th element of \( r_i \) gives the number of molecules by which element \( j \) changes due to reaction \( i \). The master equation for this system is

\[
 \frac{dP(k, t/k_0, t_0)}{dt} = \sum_i W_i(k - r_i)P(k - r_i, t/k_0, t_0) - \sum_i W_i(k)P(k, t/k_0, t_0)
\]

where \( P(k, t/k_0, t_0) \) is the probability \( x(t) = k \) given that \( x(t_0) = k_0 \). The master equation is a version of the Chapman–Kolmogorov equation for Markov processes, and, from now on, we simplify the notation by denoting \( P(k, t/k_0, t_0) \) with \( P(k, t) \) and considering appropriate initial conditions. We also assume that the underlying dynamics are such that \( W_i(x(t)) \geq 0 \). (1)

\[
 f(x(t)) = \sum_i r_i W_i(x(t)).
\]

In the linear case, we can write

\[
 f(x(t)) = Ax(t) + f_0, \quad A \in \mathbb{R}^{n \times n}, \quad f_0 \in \mathbb{R}^n
\]

leading to the explicit equation

\[
 \frac{d \langle x(t) \rangle}{dt} = A \langle x(t) \rangle + f_0.
\]

(2)

Similarly for the second moment, we have

\[
 \frac{d \langle x(t)x^T(t) \rangle}{dt} = \sum_k kk^T \frac{dP(k, t)}{dt}
\]

\[
 = \langle x(t)x^T(t) \rangle + \{f(x(t))x^T(t)\} + \left( \sum_i r_ir_i^TW_i(x(t)) \right).
\]

Writing this in terms of the covariance

\[
 \Sigma(t) := \langle x(t)x^T(t) \rangle - \langle x(t) \rangle \langle x^T(t) \rangle
\]
and noting that
\[
\langle f(x(t)) x^T(t) \rangle = \left( \langle f(x(t)) \rangle - \langle f(x(t)) \rangle \right) \left( x(t) - \langle x(t) \rangle \right)^T
\]
leads to
\[
\frac{dX(t)}{dt} = \left( \langle f(x(t)) \rangle - \langle f(x(t)) \rangle \right) \left( x(t) - \langle x(t) \rangle \right)^T
+ \left( \langle x(t) \rangle - \langle x(t) \rangle \right) \left( f(x(t)) - \langle f(x(t)) \rangle \right)^T
+ \sum_i r_i W_i \langle x(t) \rangle r_i^T.
\]
In the linear case, we have \( f(x(t)) = A(x(t) - \langle x(t) \rangle) \), giving
\[
\frac{dX(t)}{dt} = AX(t) + X(t) A^T + \sum_i r_i W_i \langle x(t) \rangle r_i^T.
\]  
Similarly, we denote \( P(k, t_1; l, t_2) \) as the joint probability \( x(t_1) = k \) and \( x(t_2) = l \), and we have
\[
\frac{\partial \langle x(t_1) x^T(t_2) \rangle}{\partial t_1} = \sum_{k,l} k l \frac{\partial P(k, t_1; l, t_2)}{\partial t_1}
= \sum_{k,l,i} (k + r_i) T W_i(k) P(k, t_1; l, t_2)
- \sum_{k,l,i} k l W_i(l) P(k, t_1; l, t_2)
= \sum_{k,l,i} r_i W_i(l) P(k, t_1; l, t_2)
= \langle f(x(t_1)) x^T(t_2) \rangle.
\]  
This can be written, as before, in terms of the autocorrelation \( R_{xx}(t_1, t_2) := \langle x(t_1) x^T(t_2) \rangle - \langle x(t_1) \rangle \langle x^T(t_2) \rangle \) using
\[
\frac{\partial R_{xx}(t_1, t_2)}{\partial t_1} = \frac{\partial \langle x(t_1) x^T(t_2) \rangle}{\partial t_1} - \frac{\partial \langle x(t_1) \rangle}{\partial t_1} \langle x^T(t_2) \rangle
\]  
and an analogous expression with (4) for \( \langle f(x(t_1)) x^T(t_2) \rangle \) leads to
\[
\frac{\partial R_{xx}(t_1, t_2)}{\partial t_1} = \left( \langle f(x(t_1)) - \langle f(x(t_1)) \rangle \right) \left( x(t_2) - \langle x(t_2) \rangle \right)^T.
\]  
In the linear case, this becomes
\[
\frac{\partial R_{xx}(t_1, t_2)}{\partial t_1} = AR_{xx}(t_1, t_2).
\]  
At steady state, \( R_{xx}(t_1, t_2) \) depends only on \( \tau = t_1 - t_2 \) and is denoted as \( R_{xx}(\tau) \). We also denote \( D(t) := \sum_i r_i W_i \langle x(t) \rangle r_i^T \). Note that \( D(t) = D^T(t) \geq 0 \), and it can be factorized as
\[
D(t) = B(t) B^T(t), \quad B(t)_{jk} = \sum_i r_i^j r_i^k W_i \langle x(t) \rangle.
\]  
\(^1\)Note the importance of the linearity of \( f(x) \) for the moment closure. If \( f(x) \) is nonlinear, then each moment depends also on higher moments (moment truncation could be used under certain conditions to get approximate solutions [13]).

and denote \( D \) and \( B \) as the steady-state values of \( D(t) \), \( B(t) \), respectively. Hence, (5) and (7) become at steady state
\[
AR_{xx}(0) + R_{xx}(0)A^T + D = 0
\]  
\[
\frac{\partial R_{xx}(\tau)}{\partial \tau} = AR_{xx}(\tau).
\]  
This provides the basis for the following theorem.

**Theorem 1:** Consider the jump process described by (1), (2), and (8) such that a wide-sense stationary process is reached at steady state. Then, the power spectrum is given by
\[
S(\omega) = G(j\omega)G^*(j\omega)
\]
where \( G(s) \) is the transfer function from \( y \) to \( z \) of the system \( z(t) = Az(t) + Bu(t) \).

**Proof:** The spectrum is given by
\[
S(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-j\omega \tau} R_{xx}(\tau) d\tau
= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-j\omega \tau} R_{xx}(\tau) + \frac{1}{2\pi} \int_{-\infty}^{0} e^{-j\omega \tau} R_{xx}(\tau) d\tau.
\]  
The first part is \( R_{xx}(j\omega) \), where \( R_{xx}(s) \) is the Laplace transform of \( R_{xx}(\tau) \). From (10), \( R_{xx}(s) = [sI - A]^{-1}R_{xx}(0) \). Also, using \( R_{xx}(-\tau) = R_{xx}(\tau)^T \) yields
\[
\frac{1}{2\pi} \int_{-\infty}^{0} e^{j\omega \tau} R_{xx}(\tau) d\tau
= \frac{1}{2\pi} \int_{-\infty}^{0} e^{j\omega \tau} R_{xx}(\tau) d\tau^T
= \left[ \frac{1}{2\pi} \int_{-\infty}^{0} e^{j\omega \tau} R_{xx}(\tau) d\tau \right]^T
= R_{xx}(-s).
\]  
So, \( S(\omega) = [j\omega I - A]^{-1}R_{xx}(0) + R_{xx}(0)[-j\omega I - A^T]^{-1} \) and \( [j\omega I - A]S(\omega) = -AR_{xx}(\tau) + R_{xx}(\tau)A^T = D \) using (9). Hence, \( S(\omega) = [j\omega I - A]^{-1}BB^T[-j\omega I - A^T]^{-1} = G(j\omega)G^*(j\omega) \).

**Remark 1:** If the input \( u(t) \) is a white-noise process, then \( z(t) \) and the jump process \( x(t) \) have the same power spectrum, even though the two processes are not identical.\(^2\)

**Remark 2:** By the definition of the power spectrum, the variance of species \( j \), \( Var[j] = \int_{-\infty}^{\infty} CS(\omega)\omega^2 d\omega \), where \( C \in R^m \) is a row vector s.t. \( C_j = 1 \), \( C_k = 0 \) for \( k \neq j \). Hence, it follows from the theorem \( Var[j] = \int_{-\infty}^{\infty} CG(j\omega)G^*(j\omega)\omega^2 d\omega = ||T_{u_{t-\omega}}||^2 \), where the latter is the \( H_2 \) norm of the transfer function from \( u \) to \( z_j \).

**Remark 3:** The Lyapunov equation in (9) is often called the fluctuation dissipation theorem, and it is encountered in many applications in physics and engineering where there is a balance between the tendency of particles to diffuse and mechanisms such as negative feedback that bring them back to equilibrium.

**Remark 4:** Matrix \( B \) in (8) can be factorized as \( B = \text{Re} \text{diag}(\sqrt{W_i}) \) with \( R_{ji} = r_i^j \), i.e., \( R \in R^{n \times m} \) can
be seen as a “routing” matrix that distributes, within the network, the intrinsic noise from the various reactions.

The exactness of Theorem 1, when transition rates are linear functions of the species concentrations, suggests that it could give approximately the second-order properties of the jump process, when the transition rates are weakly nonlinear over the range of fluctuations about equilibrium.

An alternative approach is to use linear noise approximations. These are based on a systematic approximation of the master equation by a Fokker–Planck equation, in the large volume limit, by means of van Kampen’s $\Omega$ expansion. More precisely, following the procedure in [7], each reaction is assigned a rate $\Omega f_j((x/\Omega), \Omega)$ where $\Omega$ denotes the volume of the system. The master equation hence takes the form

$$\frac{dP(x,t)}{dt} = \sum_i \Omega f_i \left( \frac{x - r_i}{\Omega}, \Omega \right) P(x - r_i, t)$$

$$- \sum_i \Omega f_i \left( \frac{x}{\Omega}, \Omega \right) P(x, t).$$

The linear noise approximation is obtained by decomposing $x$ as $x = \Omega \psi(t) + 1/\sqrt{\Omega} \xi$, where $\psi(t)$ satisfies the deterministic equations

$$\frac{d\psi(t)}{dt} = \sum_i \mathbf{r}_i^T \mathbf{f}_i (\psi(t))$$

with $\mathbf{f}_i(\psi(t)) = \lim_{\Omega \to \infty} \hat{f}_i(\psi(t), \Omega)$. Note that $\psi$ corresponds to the molecular concentrations at high volumes and let $\hat{\psi}$ be the steady-state solution of (11), which is assumed to be unique. A second-order expansion of (11) about $\hat{\psi}$ leads to a linear Fokker–Planck equation for the probability density $\Pi(\xi, t)$ as

$$\frac{\partial \Pi(\xi, t)}{\partial t} = -\sum_{i,j,k} a_{ij,k} \frac{\partial (\xi_i \Pi(\xi, t))}{\partial \xi_j} + \frac{1}{2} \sum_{i,j,k} d_{ij,k} \frac{\partial^2 \Pi(\xi, t)}{\partial \xi_j \partial \xi_k}.$$

$$A = [a_{ij,k}], \quad a_{ik} = \sum_{i=1}^{m} \mathbf{r}_i \mathbf{f}_i(\partial \psi/\partial \psi)_{\hat{\psi}} \mathbf{r}_k$$

is the Jacobian for (11) about $\hat{\psi}$ and is referred to as the drift matrix. $D = [d_{ij,k}]$ is referred to as the diffusion matrix and is given by $d_{ij,k} = \sum_{i=1}^{m} \mathbf{r}_j \mathbf{r}_i \mathbf{f}_k$. It also turns out ([(7)]) that the stationary solution of (12) is a zero-mean multivariate normal distribution with covariance matrix $\Xi$ satisfying the Lyapunov equation [cf. (9)]

$$A \Xi + \Xi A^T + D = 0. \quad \text{(13)}$$

It is important to note that the macroscopic law in (11) is not in general an equation for the mean of $x$ unless the transition rates are linear [as in (3)]. The validity of the macroscopic law has been shown in [8] in the sense that, over finite times, the trajectories $x(t)/\Omega$ of the jump Markov process converge in probability to those of the macroscopic law as $\Omega \to \infty$. Nevertheless, even though the linear noise approximation can be very convenient in the derivation of analytical results, it has certain limitations. For example, an infinite variance will be predicted for locally unstable macroscopic laws, even though these could lead to a limit cycle behavior. Also, in the case of multiple equilibria, transitions between these equilibria needs to be taken into account (see Section IV). However, it still is a good approximation for local analysis about a particular equilibrium point in cases where the fluctuations are small relative to the number of species present and provides a good basis for getting an insight into how various mechanisms interact with one another [1, 14, 15].

It is well known that that a Fokker–Planck equation for the probability density function can be deduced using Itô’s rule, from a stochastic differential equation for the corresponding random process (e.g., [16]). In this case, note that, similarly with (8), $D$ can be factorized as $D = B B^T$, where $B = R \log(\sqrt{T})$. Hence, the corresponding Langevin equation is $d\xi(t) = \mathbf{L}(\xi(t)) dt + \mathbf{B} \sqrt{W(t)}$, where $W(t)$ is an $n$-variable Wiener process. This illustrates the analogy of the linear noise approximation with the exact solution of the master equation (Theorem 1, Remark 1), as far as second-order moments are concerned. However, note a clear distinction: the jump process is not identical to a process resulting from a linear system driven by additive white Gaussian noise, which is what we obtain as an approximation in the large volume limit. It agrees, though, with such a process up to second-order moments when transition rates are linear.

Finally, we show for completeness how the variance of a species can be given as the 2-norm of an appropriate linear operator by considering the solution of the Lyapunov equation, without going through the corresponding stochastic differential equation and/or the resulting power spectrum (as in Remark 2). We use standard notation from systems theory, $\mathcal{R}_{2m}^{x} = \{ G \in \mathcal{R}_{2m}^{x} \}$ and state space realization $\mathcal{E} = \mathcal{A} \mathcal{E} + \mathcal{B} \mathcal{U}$, $y = \mathcal{C} \mathcal{E}$. It is known that $\| G \|_2^2 = \text{Tr}[\mathcal{C}^T \mathcal{X} \mathcal{C}]$, where $\mathcal{X}$ is the controllability Gramian that is the unique solution to the Lyapunov equation $\mathcal{A} \mathcal{X} \mathcal{C} + \mathcal{X} \mathcal{C}^T + \mathcal{B} \mathcal{B}^T = 0$. Comparing with (13), we can choose the system above such that $\mathcal{A} \equiv \mathcal{A}, \mathcal{B} \mathcal{B}^T \equiv D$, and $\mathcal{C} \equiv \mathcal{N}_{m,x}$. In this way, $\Xi = \mathcal{X}$, and $\| G \|_2^2 = \sum_{i,j} \text{Var}(x_{ij})$. Note also that $\| \text{Tr}[\mathcal{X}_{m,x}] \|_2^2 = \sum_{i,j} \text{Var}(x_{ij})$ and hence the variance in each of the molecular species can be given an operator norm type of characterization within a deterministic system.

This norm characterization of the variance, together with the more general equivalence of the power spectrum of a jump process with that of a system driven by noise (which is exact for linear transition rates as pointed out in Theorem 1), means that classical tools from robust control theory can be used to analyze the role of feedback. For example, the Bode integral formula can help to reveal various tradeoffs in a feedback configuration. Also, a power spectral analysis can improve our understanding of the functionality of gene regulatory networks, such as noise suppression and shifting of noise to higher frequencies in order to reduce its effect in following gene circuits in a cascade [17]. Nevertheless, one should carefully consider the underlying limitations if an $\mathcal{H}_2$ optimal control problem is posed, as the control laws we are considering do not have arbitrary structure. This is because each additional state corresponds to a molecular species which interacts with other species with chemical reactions and, hence, there is noise associated with those. However, we can still optimize over a number of parameters certain performance criteria for different regulation mechanisms, the complexity of which is often naturally limited due to the associated metabolic burden [18].
III. DELAY AS THE LIMIT OF A SEQUENCE OF BIRTH–DEATH PROCESSES

We show in this section how, by employing an operator theoretic interpretation of the variance in a jump process, as discussed in Section II, one can deduce that a sequence of coupled jump processes tends to a delay in the limit of many such processes.

Consider the following set of coupled birth–death processes:

\[
\begin{align*}
   &x_1 \xrightarrow{R_1} x_1 + 1 \\
   &\text{for } k=1, \ldots, n, \\
   (x_1, x_2) \xrightarrow{x_1n/T} (x_1 - 1, x_2 + 1) \\
   &\vdots \\
   (x_n, x_{n+1}) \xrightarrow{x_{n+1}n/T} (x_n - 1, x_{n+1} + 1)
\end{align*}
\]

(14)

i.e., species \( x_k, k = 1, \ldots, n \), is converted spontaneously to species \( x_{n+1} \) at rate \( \eta_i/T \). \( x_1 \) is constitutively formed at rate \( R_1 \), and \( x_{n+1} \) spontanteously decays at rate \( R_d \). This could correspond to a multistage process, such as transcription or translation, where there is an overall rate \( 1/T \) and a large number of intermediate steps. The process is initiated with a rate \( R_1 \), and the final product (e.g., mRNA or protein) decays with rate \( R_d \). Consider the block diagram in Fig. 1, where \( \eta_1, \ldots, \eta_{n+2} \) are white-noise processes, \( K \) is a linear time-invariant (LTI) system with transfer function \( K(s) \), and \( 1/(s+n/T) \) is the transfer function of a first-order system. In the case \( K = 0 \), then \( x_1 \ldots x_{n+1} \) in Fig. 1 has an identical power spectrum to that of the jump process in (14) (Theorem 1.1).\(^3\) We also denote the following open-loop (\( K = 0 \)) transfer functions:

\[
\begin{align*}
   G_{22}^n &:= T_{u^n \rightarrow y^n} = \left( \frac{1}{s^2 + 1} \right)^n \frac{1}{s + R_d} \\
   G_{11}^n &:= T_{n \rightarrow y^n} = \sqrt{R_1 G_{22}^n} \\
   G_{12}^n &:= T_{n \rightarrow y^n} \\
   G_{13}^n &:= T_{n \rightarrow y^n} = \sqrt{R_1} \frac{1}{s + R_d} \\
\end{align*}
\]

where \( u = [\eta_2 \eta_3 \ldots \eta_{n+1}] \) and define \( G_{22}^n := e^{-sT}/(s + R_d) \).

\( G_{11} := \sqrt{R_1 G_{22}} \).

Theorem 2 shows that the process in Fig. 2 is a limiting process with respect to the steady-state variance of \( y^n \). The proof of the Theorem is based on Lemmas 1–3 that are included at the end of the section.

**Theorem 2:** \( \text{Var}[y^n] \rightarrow \text{Var}[y_1] \) as \( n \rightarrow \infty \) for all LTI systems \( K \) s.t. the interconnection in Fig. 2 is stable.

**Proof:** Note that \( \text{Var}[y_1] = \text{Var}[u] \) with \( y \) as in Fig. 3. This holds since \( \|T_{y_1 \rightarrow y^n}\|_2^2 = \|T_{\eta_{n+2} \rightarrow y^n}\|_2^2 = 1/2 \|T_{\eta_1 \rightarrow y_1}\|_2^2 \).

\[
\text{Var}[y^n] = \|T_{y_1 \rightarrow y^n}\|_2^2 = \|T_{\eta_{n+2} \rightarrow y^n}\|_2^2 + \|T_{\eta_{n+2} \rightarrow y^n}\|_2^2.
\]

It is thus sufficient to show that for all \( K \) s.t. the interconnection in Fig. 3 is stable:

1) \( \lim_{n \rightarrow \infty} \|T_{u^n \rightarrow y^n}\|_2^2 = 0 \).
2) \( \lim_{n \rightarrow \infty} \|T_{\eta_{n+2} \rightarrow y^n}\|_2^2 = \|T_{\eta_1 \rightarrow y_1}\|_2^2 \).
3) \( \lim_{n \rightarrow \infty} \|T_{\eta_{n+2} \rightarrow y^n}\|_2^2 = \|T_{\eta_{n+2} \rightarrow y^n}\|_2^2 \).

We use the following:

\[
\begin{align*}
   \lim_{n \rightarrow \infty} \|G_{12}^n\|_2 &= 0 \quad \text{(Lemma 3)} \\
   \lim_{n \rightarrow \infty} \|G_{22}^n\|_2 &= 0 \quad \text{(Lemma 1)} \\
   \lim_{n \rightarrow \infty} \|G_{22}^n\|_2 &= 0 \quad \text{(Lemma 2)}
\end{align*}
\]

where \( \delta_1 (P_1, P_2) \) is the \( \nu \)-gap [19] between transfer functions \( P_1 \) and \( P_2 \).

**Part 1:**

\[
\|T_{u^n \rightarrow y^n}\|_2 \leq \left\| \left[ I - G_{22}^n K \right]^{-1} G_{12}^n \right\|_2
\]

\( ^3 \) The superscript of \( n \) in \( y, u, \) and \( G \) refers to the fact that we are considering an \( n \)-stage process as in Fig. 1 (\( x_{n+1} \) in Fig. 1 is also denoted by \( y^n \) for convenience in the notation).
where \( \| \cdot \|_\infty \) denotes the \( \mathcal{H}_\infty \) norm of the corresponding transfer function. From (21), \( \| (I - G_{22}^n K)^{-1} \|_\infty \to \| (I - G_{22} K)^{-1} \|_\infty \) as \( n \to \infty \) [19], and, using (19), we deduce that \( \| T_{\eta_\rightarrow \eta'} \|_2 \to 0 \) as \( n \to \infty \).

Part 2: It is sufficient to show that \( \| T_{\eta_\rightarrow \eta'} \|_2 \to 0 \) as \( n \to \infty \). Now,

\[
\| T_{\eta_\rightarrow \eta'} \|_2 = \left\| (I - G_{11}^n K)^{-1} G_{11} - (I - G_{22} K)^{-1} G_{11} \right\|_2 \\
= \left\| (I - G_{11}^n K)^{-1} G_{11} - (I - G_{22}^n K)^{-1} G_{11} \right\|_2 \\
- \left\| (I - G_{22}^n K)^{-1} (G_{11} - G_{11}^n + G_{11}) \right\|_2 \\
\leq \left\| (I - G_{11}^n K)^{-1} (G_{11} - G_{11}^n + G_{11}) \right\|_2 \\
+ \left\| (I - G_{22}^n K)^{-1} (G_{11} - G_{11}^n + G_{11}) \right\|_2 \\
\leq \left\| (I - G_{11}^n K)^{-1} \right\|_\infty \left\| (G_{11} - G_{11}^n + G_{11}) \right\|_2 \\
+ \left\| (I - G_{22}^n K)^{-1} \right\|_\infty \left\| (G_{11} - G_{11}^n + G_{11}) \right\|_2.
\]

From (21), we have \( \| (I - G_{22}^n K)^{-1} \|_\infty \to 0 \) and from (20) and (16) we have \( \| G_{11}^n - G_{11} \|_\infty \to 0 \). Also, from (15) and (16), we have

\[
\| G_{11}^n \|_2 \leq \sqrt{R_t} \left( \frac{1}{\sqrt{\frac{T}{n} + 1}} \right)^n \left\| \frac{1}{s + R_d} \right\|_2 \\
\leq \sqrt{R_t} \left\| \frac{1}{s + R_d} \right\|_2.
\]

Hence, \( \| G_{11}^n \|_2 \) is upper bounded by a finite number for all \( n \), and \( \| (I - G_{22}^n K)^{-1} \|_\infty \) is also finite since \( K \) is stabilizing. Therefore, we deduce \( \| T_{\eta_\rightarrow \eta'} \|_2 \to 0 \).

Part 3:

\[
\lim_{n \to \infty} \left\| T_{\eta_\rightarrow \eta'} \right\|_2 = \left\| T_{\eta_\rightarrow \eta'} \right\|_2
\]

using the same arguments as in Part 2 with \( G_{11}^n \) and \( G_{11} \) replaced by \( G_{11}^n \).

Lemma 1: \( \lim_{n \to \infty} \| G_{22}^n - G_{22} \|_2 = 0 \).

Proof: Let \( P_\Omega : \mathbb{R} \to \mathbb{R}, P_\Omega : \mathbb{R} \to \mathbb{R} \) as

\[
P_\Omega(\omega) := \begin{cases} 
1, & |\omega| \leq \Omega \\
0, & \text{otherwise}.
\end{cases}
\]

Now, \( \| G_{22}^n - G_{22} \|_2^2 = \| P_\Omega(\omega) [G_{22}^n(j\omega) - G_{22}(j\omega)] \|_2^2 \\
+ \| P_\Omega(\omega) [G_{22}(j\omega) - G_{22}(j\omega)] \|_2^2.
\]

From (15) and the way \( G_{22} \) is defined, we have

\[
\left\| P_\Omega(\omega) [G_{22}^n(j\omega) - G_{22}(j\omega)] \right\|_2^2 \\
= \left\| P_\Omega(\omega) \left[ \frac{1}{j\omega + R_d} \left( \left( \frac{1}{j\omega + R_d} \right)^n - e^{j\omega T} \right) \right] \right\|_2^2 \\
\leq 8 \int_\Omega^{\infty} \frac{1}{\omega^2 + R_d^2} d\omega \to 0
\]
as \( \Omega \to \infty \) (uniformly in \( n \))

and

\[
\left\| P_\Omega(\omega) [G_{22}(j\omega) - G_{22}(j\omega)] \right\|_2 \\
\leq \sqrt{2} \sup_{\omega \in [0,\Omega]} \frac{1}{\omega^2 + R_d^2} \left( \frac{1}{\omega^2 + R_d^2} \right)^n - e^{-j\omega T} \to 0
\]
as \( n \to \infty \) (for given \( \Omega \)).

Thus, given \( \varepsilon > 0 \), one can choose \( \Omega \) (independently of \( n \)) s.t. \( \| P_\Omega(\omega) [G_{22}^n(j\omega) - G_{22}(j\omega)] \|_2 < \varepsilon /2 \) and \( n \) s.t. \( \| P_\Omega(\omega) [G_{22}(j\omega) - G_{22}(j\omega)] \|_2 < \varepsilon /2 \).

Lemma 2: \( \lim_{n \to \infty} \delta_\nu(G_{22}^n, G_{22}) = 0 \).

Proof: We show first that \( \exists N \) s.t. the winding number condition for \( \delta_\nu \) [19] is satisfied for \( n \geq N \). Since \( G_{22}, G_{22} \in \mathcal{H}_\infty \) for all \( n \), the winding number condition for \( \delta_\nu(G_{22}^n, G_{22}) \) is

\[
\text{w} \left( 1 + G_{22}(j\omega) G_{22}^n(j\omega) \right) = 0
\]

where \( \text{w}(\cdot) \) denotes the winding number with the function evaluated on the Nyquist contour. Now,

\[
G_{22}^n(j\omega) G_{22}(-j\omega) = \frac{1}{\omega^2 + R_d^2} e^{-j n \arctan(\omega T/n)}.
\]

Note that \( G_{22}^n(j\omega) G_{22}( -j\omega) \) is a decreasing function of \( \omega^2 \) and \( \omega^2 = n \arctan(\omega T/n) \) is an increasing function of \( \omega \) for \( \omega > 0 \) and a decreasing function of \( n \). Thus, one can choose \( \Omega \) s.t. \( G_{22}^n(j\omega) G_{22}( -j\omega) < 1 \) and then choose \( n = N \) s.t. \( \omega T = n \arctan(\omega T/n) < \pi \). Therefore, \( G_{22}^n(j\omega) G_{22}( -j\omega) \) does not encircle the point \(-1\) for \( n \geq N \).

Since (22) is satisfied for \( n \geq N \), we have

\[
\delta_\nu(G_{22}^n, G_{22}) = \sup_{\omega} \frac{\| G_{22}^n(j\omega) - G_{22}(j\omega) \|}{\sqrt{1 + |G_{22}^n(j\omega)|^2} \sqrt{1 + |G_{22}(j\omega)|^2}} \\
\leq \| G_{22}^n - G_{22} \|_\infty.
\]

Using similar arguments as in Lemma 1 given \( \varepsilon > 0 \), one can choose \( \Omega \) (independently of \( n \)) s.t. \( \sup_{|\omega| > \Omega} |G_{22}^n(j\omega) - G_{22}(j\omega)| < \varepsilon \) and \( n \) s.t. \( \sup_{|\omega| \in [0,\Omega]} |G_{22}^n(j\omega) - G_{22}(j\omega)| < \varepsilon \).

Lemma 3: \( \lim_{n \to \infty} \| G_{22}^n \|_2 = 0 \).

Proof: \( \| G_{22}^n \|_2^2 = \| T_{\eta_\rightarrow \eta'} \|_2^2 = \sum_{k=2}^{n+1} \| T_{\eta_\rightarrow \eta'} \|_2^2 \\
\) where \( T_{\eta_\rightarrow \eta'} = \left( \frac{1}{\sqrt{n+1} \times \frac{1}{s + R_d}} \right)^n + 1 \frac{1}{s + R_d}.
\]
Thus, writing the 2-norm in the frequency domain and simplifying the summation yields
\[
\left\| T_{\omega \rightarrow n} \right\|_2^2 = \int_{-\infty}^{\infty} \frac{1}{(\omega^2 + I_2)} \left[ 1 - \left( \frac{\omega^2 I_2}{n^2} + 1 \right)^n \right] d\omega
\]
\[
= \int_{-\Omega}^{\Omega} f(\omega) d\omega + 2 \int_{-\Omega}^{\Omega} f(\omega) d\omega.
\]

Hence, using similar arguments as in Lemma 1, given \( \epsilon > 0 \), one can choose \( \Omega \) (independently of \( n \)) s.t. \( I_2(\Omega) < \epsilon/2 \) and \( n \) s.t. \( I_1(\Omega) < \epsilon/2 \).

IV. BISTABILITY VERSUS BIMODALITY

In the previous sections, we considered cases where an exact solution to the master equation is possible and showed that a noise-filtering setup with an operator theoretic interpretation can be relevant not only in the large volume limit but also in small molecule numbers when transition rates are linear. When transition rates are nonlinear, then an exact solution to the master equation is generally not possible, and different approximation schemes are often employed for analysis. For example, mass action kinetics and Langevin approaches (nonlinear analogs of the exact additive noise model discussed in Section II) can be formally justified when molecule numbers are large [8], [9]. Our aim in this section is to emphasize the importance of taking into account the underlying dynamics in the presence of intrinsic noise, since otherwise steady-state solutions of the master equation and the mass action kinetics can be misinterpreted, particularly when molecule numbers are small. The examples are taken from systems where the macroscopic model exhibits bistability.

It is often assumed that bistability of deterministic mass action kinetics is associated with bimodality in the steady-state solution to the master equation. However, this is often not the case—we can have bistability without obvious bimodality and bimodality without bistability. In fact, steady-state solutions to either the mass action kinetics or the master equation can be very misleading—we cannot ignore the dynamics.

Consider the single species system with jumps
\[
x \xrightarrow{f(x)} x + 1, \quad x \xrightarrow{\beta x} x - 1 \tag{24}
\]
and let \( f(\cdot) \) take the form of a Hill equation with Hill coefficient \( n, f(x) = \frac{v(\alpha + kx^n)}{(1 + kx^n)} \) (such equations are typical in gene regulation—an example will be given in Section V). We regard \( \beta \) as time varying—possibly representing varying growth rates or else availability of some substrate, as will in the example in Section V. There are many ways of analyzing this system. Mass action kinetics suggest the ordinary differential equation (ODE)
\[
\dot{x} = f(x) - \beta x. \tag{25}
\]

Consider the parameters \( \alpha = 0.125, n = 3, k = 0.003750 \times 10^{-3}, \) and \( v = 40 \). In this case, the steady-state solutions to (25) exhibits hysteresis. Fig. 4 depicts \( f(x) \) and \( \beta x \) for \( \beta = 0.95 \) and shows three intersections—the two outer ones stable and the inner one unstable.

In Fig. 5, the solid deterministic/steady-state curve shows the equilibria of this equation. The dash–dot deterministic/dynamic curve shows the dynamic solution for \( \beta = 1 + 0.5 \cos(2\pi t/T) \), with \( T = 500 \), a sinusoidally varying \( \beta \) between 1.5 and 0.5. The hysteresis loop lies outside the steady-state hysteresis loop because of the dynamics—i.e., as \( \beta \) reduces below 0.8, the lower equilibrium becomes unstable, but it takes time for the state to move towards the upper equilibrium. Is it possible for the stochastic solutions to exhibit hysteresis?

The steady-state solution to the master equation does not exhibit hysteresis, of course, for each value of \( \beta \), there is a unique equilibrium distribution. The mean of this distribution is shown as the solid stochastic/steady-state curve). So, is it possible for the real, stochastic, system to exhibit hysteresis? In particular, can the real system ever behave like solid deterministic/steady-state curve. The answer is no, at least for the molecule numbers here. The dashed stochastic/dynamic curve shows the mean of the solution to the time varying master equation. If the rate of variation of \( \beta \) were any slower then these curves would lie closer to the solid stochastic/steady-state curve; any faster and they would move towards the dash–dot deterministic/dynamic curve. To get the standard hysteresis curve (deterministic/
steady-state), there has to be a time-scale separation between the deterministic dynamics and the relaxation to steady state of the master equation. This requires large numbers of molecules. In electronics, this is not an issue—for example, in a standard Schmitt trigger, the deterministic dynamics are of the order of milliseconds or faster. However, the time taken to flip from a one to a zero with fixed input is probably of the order of many thousands of years, at least. In gene regulation, we would argue that this is not the case; there is no such time-scale separation, and so the standard hysteresis curves for electronic circuits are irrelevant.

Of the curves in the figure, it is the dashed stochastic/dynamic ones that are the most difficult to calculate. As we increase molecule numbers it should be easier to find time scales for which the curve lies closest to the ideal hysteresis—however, the master equation becomes increasingly difficult to solve. It is though possible to estimate the time taken to jump from the on to the off state by approximating the master equation with a quasi-linear Fokker–Planck equation

\[
\frac{\partial P(x,t)}{\partial t} = \frac{\partial}{\partial x} \left[ U'(x) P(x,t) \right] + \theta \frac{\partial^2 P(x,t)}{\partial x^2}
\]

which is satisfied by the probability density function of the stochastic differential equation \(dx(t) = -U'(x(t))dt + \sigma dW(t)\) where \(\sigma = \sqrt{2\theta}\) and \(-U'(x)\) is the vector field in the mass action equation (i.e., \(U'(x) = [f(x) - \beta x]\) in the example). If \(U'(x) = 0\) at \(x = a, b, c\), with stable equilibria \(a\) and \(c\) where \(a < b < c\), then an explicit value of the mean first passage time from \(a\) to \(c\) can be derived as\(^7\)

\[
\tau_{ac} = \frac{1}{\theta} \int_a^c \frac{U(x)}{\theta} dx dy.
\]  

Using a parabolic approximation for the potential well \(U(x)\) around each of the equilibrium points (i.e., a first-order approximation for \(U'(x)\) as in the linear noise approximation), the integral can be simplified\(^6\) to

\[
\tau_{ac} = \frac{2\pi}{\sqrt{U''(a)U''(b)}} \exp \left[ \frac{U(b) - U(a)}{\theta} \right].
\]  

\(^6\)The on and off states in the quasi-linear Fokker–Planck approximation are the two stable equilibria in the mass action kinetics. In the master equation, one would normally use the peaks of the bimodal distribution, though a time-scale separation is also needed for these to make sense, as previously discussed.

\(^7\)One assumes that \(e^{-U(x)/\theta}\) has a significant contribution only round the peak at \(b\), in which case the inner integral is approximately constant since \(e^{-U(y)/\theta}\) is small round \(b\). Then similarly \(e^{-U(y)/\theta}\) has a significant contribution to the inner integral only round the potential well at \(a\). Therefore, using a parabolic approximation of \(U'(x)\) about \(a\) and \(b\) the double integral is approximated with

\[
\tau_{ac} \approx \frac{2\pi}{\sqrt{U''(a)U''(b)}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{U(x) - U(a)}{\theta}} \frac{U'(y)}{\theta} dy dx
\]

which is of the same form as the Arrhenius equation \(\kappa = A \exp(-E_A/RT)\) relating the rate constant \(\kappa\) of a chemical reaction to the activation energy \(E_A\) and the temperature \(T\) (i.e., the height of the potential well \(U(b) - U(a)\) is analogous to the activation energy \(E_A\)). This implies that the expected time for transitions between the two equilibria depends on the shape and height of the potential well, as well as the power of the noise. Noting that the time scale of convergence to the stable equilibrium point is of the order of the factor multiplying the exponential term, \(\exp((U(b) - U(a))/\theta)\) gives roughly the time-scale separation, i.e., it increases exponentially with the height of the potential well divided by half the noise power.

For the parameters given above, the function \(U(x)\) is shown in Fig. 6. The estimated transition times may be calculated from (27) and (26), respectively,\(^7\) as

\[
\text{approximate : } \tau_{ac} = 96.9 \quad \tau_{ca} = 30.6
\]

\[
\text{“exact” : } \tau_{ac} = 122.7 \quad \tau_{ca} = 28.3.
\]

A typical sample path (simulated using the Gillespie algorithm) is shown in Fig. 7. The average, from simulated trajectories, of one hundred transition times was calculated, respectively, as \(\tau_{ac} = 91.7\) and \(\tau_{ca} = 29.9\). Note that the model-based estimates of \(\tau_{ca}\) are closer to the average from the exact simulations because of the larger molecule numbers present at \(c\), which makes the quasi-linear Fokker–Planck equation a better approximation.

Finally, Fig. 8 shows the propensities and master equation solutions for two different sets of parameters in (24) and (25);

\(^7\)The noise power for the escape from \(a\) is approximated as \(\sigma^2 = 2/\theta\) (similarly, for escape from \(b\), \(\sigma^2 \approx 2/\theta\)).
showing that you can have bimodality without bistability (first column of Fig. 8, which only has one intersection) and bistability without any obvious bimodality (second column of Fig. 8—there is a small peak associated with the higher equilibrium), i.e., steady-state solutions can be misleading on their own. In general, for low molecule numbers, steady-state bimodality is common, however, mass action kinetics are less accurate in this regime and bistability might not necessarily be present. For higher molecule numbers, the mass action kinetics and nonlinear Langevin approximations are more relevant, but steady-state bimodality is rare (i.e., occurs over a very narrow range of parameters)—for most parameters, one or other of the potential wells will dominate. In this regime, any observed bimodality will either be transient or due to other factors. This is discussed in detail for the lac operon in Section V.

V. Lac Operon

As one of the most studied systems in bacteria, the lac operon in Escherichia coli has long been a paradigm for gene regulation in prokaryotic organisms [20]. The central dogma of molecular biology is that genes which, along with all of an organism’s heritable information, are found on the DNA are transcribed (or copied) by molecules called RNA polymerase into single-stranded messenger RNA (mRNA). The mRNA is then translated by ribosomes into proteins. The lac operon controls the expression of enzymes for the metabolism of lactose in the cell, and consists of four genes: lacI, lacZ, lacY, and lacA which encode for four proteins: the lac repressor, β-galactosidase, lactose permease, and thiogalactoside acetyltransferase, respectively. lacZ, lacY, and lacA are all co-transcribed onto the same mRNA molecule.

In the absence of an inducer, such as allolactose—the natural inducer—or a gratuitous inducer such as methyl-β-D-thiogalactopyranoside (TMG), the repressor molecules bind to specific sites on the DNA, called operator sites, and inhibit transcription of the lacZYA mRNA product by RNA polymerase at the lac promoter. Inducers bind to the repressor and decrease the affinity of the repressor for the operator by approximately a thousand fold, freeing the operator and allowing transcription to proceed. The mRNA is then translated by ribosomes leading to the formation of proteins. β-galactosidase is responsible for hydrolysing lactose into glucose and galactose, while the permease is a membrane transport protein that accumulates lactose and other galatosides inside the cell by active transport. Thiogalactoside acetyltransferase (LacA) has no obvious physiological function associated with the operon.

Hysteresis was first observed in the lac operon induced with TMG by Novick and Weiner [11]. TMG is a gratuitous inducer, so-called because although it is transported by lactose permease and binds to the repressor, it is not hydrolyzed by β-galactosidase. Since then, bimodal behavior has been described in numerous experiments involving gratuitous inducers (e.g. [21] and [22]).

A. Deterministic Model

The model of the lac operon with TMG is simpler than the full model in the presence of lactose (e.g., [23]), since β-galactosidase can be ignored, and only transcription, translation, and transport through the membrane and degradation and dilution need be considered.

Assuming repressor-operator interactions proceed at a fast rate, transcription can be modelled by means of a Hill equation with Hill coefficient \( n = 2 \) [23], [24]. Bi-directional transport through the membrane can be approximated by Michaelis–Menten kinetics, and assuming linear rates for translation and degradation, the state of the system—i.e., the concentration of internal inducer \( I \), mRNA \( M \) and permease \( P \), for any external inducer concentration, \( I_e \)—can be described by

\[
\frac{dM}{dt} = \alpha_M \frac{1 + K_1 L^2}{K + K_1 L^2} - (\gamma_M + \mu)M
\]

\[
\frac{dP}{dt} = \alpha_P M - \mu P
\]

\[
\frac{dL}{dt} = P \left( \frac{\alpha_L L_e}{k_{LE} + I_e} - \frac{\beta_L L}{k_L + L} \right) - \mu L
\]

Data are available to estimate the parameters in this model. Noting that the expression is thousandfold less under repression [20], we get \( K = 1000 \) and can determine \( K_1 = 3.1 \times 10^4 \text{mM}^{-2} \) by fitting the Hill equation to an induction curve in [25].

During exponential growth in rich media and under near optimal conditions, \( E. coli \) doubles approximately every 22.5 min, giving a growth rate \( \mu = 0.0308 \text{ min}^{-1} \). The rates of transcription and translation are taken from [26] and adjusted for the growth rate to give \( \alpha_M = 60 \text{ min}^{-1} \) and \( \alpha_P = 2.25 \text{ min}^{-1} \). Taking the volume of an \( E. coli \) cell to be \( 2 \times 10^{-15} \text{L} \) [20], we get \( \alpha_M = 4.98 \times 10^{-5} \text{ mM min}^{-1} \). lacZYA mRNA has a half-life of 1.7 min [27], giving \( \gamma_M = 0.41 \text{ min}^{-1} \).

Finally, \( \alpha_L, k_{LE}, \), and \( k_L \) are given in [28] as 2220 \text{ min}^{-1}, 0.87 \text{ mM}, and 42 \text{ mM}. At an external concentration of 0.5 mM, TMG is reported to accumulate in nongrowing cells to 180 times the external concentration [29], from which it is possible to calculate \( \beta_L = 1188 \text{ min}^{-1} \).

Solving (28)–(30) at steady state with these parameters corresponds to the intersections in Fig. 9, which is analogous to...
the fast species are modeled using ordinary differential equations, is used (see the Appendix).

When moving from the deterministic to a stochastic model, it is straightforward to calculate the propensities for translation and degradation, but some thought needs to be given to transcription. If we calculate the propensity of transcription $W_{M \rightarrow M+1}$ as

$$W_{M \rightarrow M+1} = \frac{1 + K_1 L^2}{K + K_1 L^2}$$

then we ignore the underlying interactions that are masked by the Hill equation. To allow for the leaky expression in the absence of inducer, we propose to calculate the propensity of transcription by

$$W_{M \rightarrow M+1} = \begin{cases} 0, & \text{if } X = 1 \\ \frac{\alpha M}{K}, & \text{if } X = 0 \end{cases}$$

where $X = 1$ when the operator is occupied ($O$) and $X = 0$ when the operator is free ($F$). This will give the same average operator occupancy as (31) in the limit, provided that

$$\frac{1}{K} O + F = \frac{1 + K_1 L^2}{K + K_1 L^2}.$$  

If the operator state changes by $O \rightarrow F$ and noting that $F + O = 1$ and $\alpha \beta = F/O$, (33) is satisfied if we assign $\alpha$ and $\beta$ such that $\alpha \beta = K_1 L^2 / K$. Assuming that the operator becomes free in the absence of inducer due to the effects of DNA replication, it makes sense to choose $\alpha$ on the order of $\mu$ ($\alpha \approx 4 \mu$ was used in the simulations).

While the analysis described in Section IV for transition times can be extended to the multivariate case, we have used stochastic simulations to investigate the average time it takes to transition from induced to uninduced states (and vice versa) at different inducer concentrations. This is done by starting at either induced or uninduced conditions, and then running the simulation until a threshold has passed, indicating that the state has changed. A threshold on the number of permease molecules $P$ in a cell (uninduced: $P < 250$; induced: $P > 2000$) is used. The average time to transition is taken from the average of up to 100 simulations at any one concentration in any one direction.

A study of the average transition times of the model (Fig. 11) reveals that, in spite of being deterministically bistable, the model would not exhibit any significant bimodality. At low TMG concentrations ($L_e < 0.8 \mu M$), the time to become uninduced is many orders of magnitude less than the time to become induced, so we would expect to only find uninduced cells. Similarly, at higher concentrations of inducer ($L_e > 1.5 \mu M$), we would only expect to find induced cells.

In the range $0.8 \mu M < L_e < 1.5 \mu M$, the transition times in either direction are so large ($> 10^4$ min) that no transitions would be expected over the time of a typical experiment investigating bistability, and so cells would remain in their initial condition at these concentrations unless there is additional noise introduced. This hysteresis region is in agreement, in principle, with the observation of Novick and Weiner [11] that, for intermediate TMG concentrations, neither induction nor loss of induction occurs.
VI. CONCLUSION

It has been illustrated in the paper how an operator theoretic approach can be relevant in quantifying the intrinsic stochasticity in jump processes described by master equations. Such an interpretation of the noise was used as a basis for showing that a sequence of coupled birth/death processes relevant in gene expression tends to a delay in the limit of many such processes. The effect of noise in the case of nonlinear mass action kinetics leading to bistabilities was also considered. It was discussed that, in order to avoid potential pitfalls and misinterpretations in the functionality of the network, the steady-state solution of the deterministic macroscopic model or the master equation should be analyzed in conjunction with the underlying dynamic behavior.

As a final comment, it should not be neglected that a major uncertainty in any model-based prediction lies in the validity of the model itself. Are there any intermediate reactions that contribute to intrinsic stochasticity that have not been taken into account? Is the Markovian assumption always valid? How reliable are the parameter estimates? This emphasizes the need for experimental verification as well as fundamental principles and limitations that supplement any complex model-based calculations. The presence of delays, as well as the fact that feedback affects propensities and not actual molecular levels, do pose such fundamental limitations for an arbitrary feedback policy (which will be discussed in an upcoming research paper).

APPENDIX

EXACT HYBRID SIMULATION

Here, we discuss the exact hybrid stochastic/deterministic simulation used to obtain the \( \text{lac} \) simulations. The main idea is to partition reactions into slow and fast subsets, and to solve the fast ones deterministically between jumps of the slow ones. The difficulty lies in choosing the time for which to integrate the differential equations. A number of such algorithms have been published. Notable among those is that in [31], which avoids having to solve the ODEs over a longer period than necessary. In comparison to that in [31], the novelty in the algorithm below is that, by changing variables (from time to integrated propensities), we can solve the ODEs up to precisely the time of the next jump in one shot. This is both more convenient and faster than incorporating stopping criteria into the ODE solver.

As before, we consider the system with jumps

\[
W_i(x) \rightarrow x + r_i, \quad i = 1, 2, \ldots, m
\]

where \( x \) is now a vector of molecule numbers, i.e.,

\[
x = [x_1, x_2, \ldots, x_n]^T
\]

and \( x_i \) is the number of molecules of species \( i \).

We partition the reactions into two sets \( \mathcal{C} \), the fast ones to be treated as continuous, and \( \mathcal{D} \), the slow ones to be treated as discrete. Fast and slow reactions should normally differ by at least an order of magnitude (an exact simulation is often first helpful to see which reactions occur more frequently). Also, different partitions could apply in different regimes depending on molecule numbers (the partition need only be kept unchanged between successive discrete jumps).

We now solve

\[
\dot{x} = \sum_{i \in \mathcal{C}} r_i W_i(x) \tag{34}
\]

which is correct under the assumption that no discrete reactions occur, until that next reaction occurs. Define

\[
\lambda(t) = \sum_{i \in \mathcal{D}} W_i(x(t))
\]

as the total time-varying propensity of the discrete reactions, and let \( u \) be sampled from a uniform distribution on \([0,1]\). For notational simplicity, let \( t = 0 \) denote the current time. The next discrete reaction is taken to occur at time \( T \) such that

\[
\int_0^T \lambda(\tau) d\tau = -\ln u.
\]

Let \( I(t) = \int_0^t \lambda(\tau) d\tau \). We could append the equation \( dI/dt = \lambda(t) \) to (34) and integrate them until \( I \) hits \((-\ln u)\), but a neater way is to integrate with respect to \( I \) instead—noting that \( dI = \lambda dt \). Changing variables, we obtain

\[
\frac{dx}{dI} = \sum_{i \in \mathcal{C}} W_i(x) \sum_{i \in \mathcal{C}} r_i W_i(x)
\]

which we integrate up to \( I = -\ln u \). We should append the equation

\[
\frac{dt}{dI} = \frac{1}{\sum_{i \in \mathcal{D}} W_i(x)}
\]

in order to keep track of time.

Finally, we choose which reaction occurs in the usual way, with the probability of reaction \( j \) being

\[
\frac{W_j(x(T))}{\sum_{i \in \mathcal{D}} W_i(x(T))}.
\]

One further call to a uniform random number generator suffices for this, by breaking the interval \([0,1]\) into sections of length proportional to \( W_i(x(T)) \).

Thus, we make precisely two calls to a uniform random number generator and one to a standard ODE solver for each discrete event.
REFERENCES


Johan Paulsson received the B.A. degrees in mathematics and biology and the Ph.D. degree from Uppsala University, Uppsala, Sweden, in 1996 and 2000, respectively. From 2000 to 2003, he was a Lewis–Thomas Fellow with Princeton University, Princeton, NJ, and then moved to the University of Cambridge, Cambridge, U.K., where he held a faculty position with the Department of Applied Mathematics and Theoretical Physics between 2003 and 2006. He is currently on the faculty of Harvard University’s Department of Systems Biology, Cambridge, MA, and remains affiliated with the Faculty of Mathematics with the University of Cambridge, where he teaches Part III. His research focuses on stochastic processes in biology, developing analytical theory and experimental methods.

Nicholas Ross received the B.Sc.(Eng) degree in electrical engineering and the M.Sc.(Eng) degree in image recognition from the University of the Witwatersrand, Johannesburg, South Africa, in 2002 and 2004, respectively. He is currently working toward the Ph.D. degree in control and systems biology at the University of Cambridge, Cambridge, U.K.

Glenn Vinnicombe received the B.A. degree in engineering and the Ph.D. degree from the University of Cambridge, Cambridge, U.K., in 1984 and 1993, respectively. From 1984 to 1987, he was with British Aerospace, working primarily on the design and flight test of control systems on the Airbus A320. He returned to the University of Cambridge in 1987 as a College Lecturer at Churchill College. He has held faculty positions with the Department of Mechanical Engineering, University of Minnesota, and is currently on the faculty of the Department of Engineering, University of Cambridge.