Frequency domain chemical Langevin analysis of stochasticity in gene transcriptional regulation

Michael L. Simpson\textsuperscript{a,b,c,*}, Chris D. Cox\textsuperscript{b,d}, Gary S. Sayler\textsuperscript{b,e}

\textsuperscript{a}Molecular Scale Engineering and Nanoscale Technologies Research Group, Oak Ridge National Laboratory, P.O. Box 2008, M.S. 6006, Oak Ridge, TN 37831-6006, USA
\textsuperscript{b}Center for Environmental Biotechnology, University of Tennessee, Knoxville, TN 37996, USA
\textsuperscript{c}Department of Materials Science and Engineering, University of Tennessee, Knoxville, TN 37996, USA
\textsuperscript{d}Department of Civil and Environmental Engineering, University of Tennessee, Knoxville, TN 37996, USA
\textsuperscript{e}Department of Microbiology, University of Tennessee, Knoxville, TN 37996, USA

Received 3 November 2003; received in revised form 1 March 2004; accepted 8 April 2004

Abstract

We present a frequency domain Langevin approach for stochastic analysis that remains valid for many important gene circuit elements even as molecular populations approach zero. We begin by considering the case of low-rate transcription and show that the previously reported shot noise representation is exact at all mRNA population levels for a constant transcription rate. Next, we consider transcriptional control through protein–DNA interactions at an operator site within the gene promoter region. This analysis results in expressions for the dynamics and noise behavior of this important gene sub-circuit, including the spectral density of the intrinsic operator noise and the processing of extrinsic noise by this transcriptional regulation system. This analysis shows that mRNA synthesis noise is composed of wideband shot noise and band-limited operator binding generated noise components. We find that the bandwidth of operator noise and its ultimate effect on total mRNA and protein noise is controlled by operator binding and unbinding dynamics. The most substantial impact of the operator noise is seen at transcription rates just above basal expression. This analysis captures the full behavior of this transcriptional regulation system, and points to potentially serious flaws in simplified mathematical relationships often used to model transcriptional regulation.

© 2004 Elsevier Ltd. All rights reserved.

Keywords: Gene circuit analysis; Stochastic simulation; Gene circuit noise; Chemical Langevin equation

1. Introduction

A new appreciation of the role of stochastic processes in decision making in biological systems is emerging, as it is now understood that these processes may play a pivotal role in gene circuit functionality (Rao et al., 2002). This is perhaps best illustrated by the \( \lambda \) phage lysis–lysogeny decision circuit (Arkin et al., 1998), where the final state of the circuit is arbitrated by the increasing concentration of two phage proteins engaged in a critical race (Mcadams and Shapiro, 1995; Ptashne, 1992). At the individual cell level, the outcome of this critical race, and therefore the lysis-lysogeny decision, is determined in large part by the stochastic variations (hereafter referred to as noise) in the protein concentrations. At the community level, this noise-mediated switch allows a nearly continuous selection of the fraction of cells that become lysogenic within a population, thus yielding precise control from imprecise components. A similar role for noise has been described for bacterial quorum sensing (Cox et al., 2003), infection (Connell et al., 1996), and other processes where cells gain an advantage by randomly expressing certain genes and silencing others as mediated by stochastic processes.

Abbreviations: Power Spectral Density—PSD Chemical Master Equation—CME Chemical Langevin Equation—CLE.

*Corresponding author. Molecular Scale Engineering and Nanoscale Technologies Research Group, Oak Ridge National Laboratory, P.O. Box 2008, M.S. 6006, Oak Ridge, TN 37831-6006, USA. Tel.: +865-574-8588; fax: +865-576-2813.
E-mail address: simpsonML1@ornl.gov (M.L. Simpson).
Stochastic behavior of gene circuits can be found through solution of the chemical master equation (CME) (Gillespie, 1992; Rao et al., 2002). However, as the CME usually can be solved only for very simple systems, noise analysis typically is performed using exact stochastic simulation (Gillespie, 1977), or approximate solutions are found from the chemical Langevin equations (CLEs) (Rao et al., 2002). Exact stochastic simulation is thought to provide the most accurate results as it explicitly deals with discrete molecules of each species, microscopic rate constants, and circuit nonlinearities. However, even for circuits of only modest complexity, simulation can be computationally demanding, and does not generally lead to an intuitive understanding of simulated noise behavior with specific circuit elements, parameters, and operational regimes.

Conversely, the CLEs represents stochastic processes in cells by adding noise terms to the ordinary differential equations associated with the chemical reactions (Rao et al., 2002). The noise behavior of relatively simple gene circuits has been solved through the time domain solution of the associated CLEs (Ozbudak et al., 2003). As the FD-CLEs are related to the CLEs through either the Fourier (steady state) or Laplace (transient) transform, the two methods share the same limitations and caveats.

Although the CLEs are much easier to deal with analytically than the CME and much less computationally demanding than exact stochastic simulation, important limitations of this approach are often expressed in the literature. Several authors express concern about the accuracy of the Langevin representation when applied to systems with low molecular populations (Arkin et al., 1998; Gillespie, 2000, 2002; Simpson et al., 2003), yet such systems are vitally important to nearly all gene circuits. Here we offer a formulation of the CLEs at low molecular populations that remains valid for many gene circuit elements even as molecular populations go to zero. First we show that the previously reported shot noise representation (Simpson et al., 2003) is valid for all mRNA population levels for a constant transcription rate. Next we consider the case of transcriptional regulation through protein–DNA interactions at an operator site within the gene promoter region. For a single operator site and a single copy of the operon, this system has only two allowable states (and therefore a molecular population of either one or zero), and would generally be considered a poor candidate for Langevin analysis. However, the FD-CLE analysis presented here provides a solution that is useful for most operational regimes of this important gene circuit element.

The results presented here yield a fundamental analytical model of gene circuit elements central to gene expression control, and elucidate circuit architectural pressures that may have led to the evolution of the transcriptional control dynamics observed in natural systems. These results have important implications in areas beyond gene circuit analysis, including the development of more efficient computational algorithms for simulation, and the generation of design equations for synthetic gene circuit development (Elowitz and Leibler, 2000; Gardner et al., 2000; Judd et al., 2000; McMillen et al., 2002).

2. Chemical Langevin equations and low molecular populations

Genetic and biochemical circuits are systems composed of various chemical species specified by state vector $X(t)$, which indicates the number of molecules of each species. The chemical species react through $M$ chemical reactions (indexed $j$), affecting the populations of molecules in each species according to a stoichiometry matrix $v_j$ and the propensity of each reaction to occur, $a_j(x)$, is a function only of the system state $X(t)$. Starting from the CME, Gillespie has provided a rigorous derivation of the CLEs for this system yielding (Gillespie, 2000)

$$\frac{dX_i(t)}{dt} = \sum_{j=1}^{M} v_{ij}a_j(X(t)) + \sum_{j=1}^{M} v_{ij\gamma_j}^{1/2}(X(t))\Gamma_j(t),$$

where the $\Gamma_j(t)$ are statistically independent Gaussian white noises associated with each reaction type and $\Delta t$ is a positive macroscopically infinitesimal time increment (Gillespie, 2002). Eq. (1) approximates a jump Markov process as a continuous Markov process represented by the sum of drift-compensator and Gaussian noise fluctuation terms. Gillespie (2000, 2002) identified two conditions that must be satisfied in order for this approximation to be valid: (1) $\Delta t$ must be of short enough duration that the $a_j$ terms do not change significantly during $\Delta t$ and (2) $\Delta t$ must be of long enough duration that each reaction channel $M_j$ fires several times during $\Delta t$. In many important gene circuit elements these two conditions cannot be satisfied simultaneously. Consider the case of transcription shown in Fig. 1. The mRNA molecules are synthesized at an average rate of $2g$ and decay at an average rate of $7g r(t)$, where $r(t)$ is the population of mRNA molecules.
Fig. 1. Transcription in a single gene circuit where mRNA molecules are synthesized from the template DNA strand at a constant rate, $g$, and decay at an average rate of $\gamma_R$, where $r(t)$ is the population of mRNA molecules at time $t$.

at time $t$. The CLE for this system is

$$\frac{dr(t)}{dt} = -\gamma_R r(t) + g_R(t) + \theta(t),$$

(2)

where $\theta(t)$ is a random variable representing the noise. However, at low values of $r(t)$ a suitable value of $\theta(t)$ cannot be found as the propensity function for decay ($\gamma_R r(t)$) changes significantly after each synthesis and decay event.

To address this problem we define a new variable, $\overline{r}(t)$, which we conceptualize as a continuous (not constrained to integer values) signal encoded within the discrete (constrained to integer values) random process $r(t)$. $r(t)$ and $\overline{r}(t)$ are related by

$$r(t) = \overline{r}(t) + \eta(t),$$

(3)

where $\eta(t)$ is a zero mean random noise term that is the noise penalty for encoding $\overline{r}(t)$ within $r(t)$. Unlike the situation described in Eq. (1), $\eta(t)$ is neither constrained nor likely to have a Gaussian distribution, especially at low molecular populations.

We find $\overline{r}(t)$ by considering a series of discrete random processes, $r_n(t)$ ($n = 1, 2, 3, ..., N$) which are equivalent to $r(t)$ in a stochastic sense with uncorrelated $\eta_n(t)$ terms. Then,

$$\overline{r}(t) = \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} r_n(t)$$

(4)

and

$$\frac{d\overline{r}(t)}{dt} = -\gamma_R \overline{r}(t) + \overline{g_R(t)},$$

(5)

where $\overline{g_R(t)}$ is the instantaneous rate of mRNA synthesis averaged over the $N$ different processes, and $dr$ is a positive microscopically infinitesimal time increment as usually defined for the derivative. The noise generated by mRNA synthesis and averaged by mRNA decay (Simpson et al., 2003) has been absorbed into $\eta(t)$ in Eq. (3). Eqs. (3) and (5) form a modified CLE equation set, where the solution to Eq. (5) defines the dynamics (i.e. frequency and transient response) of the system, and the difficulty has been moved from finding a suitable $\overline{r}$ in Eq. (1) to finding an appropriate expression for the noise in Eq. (3).

We now define the noise relationship for unregulated transcription below and for transcription regulated through protein–DNA interactions later in this paper. Defining the noise relationship for higher-order processes, such as dimerization, is more difficult and may limit range of systems to which the approach may be applied.

We find an appropriate expression for the noise in Eq. (3) as described in Appendix A by direct calculation of the autocorrelation function of the random process and subsequent Fourier transformation to obtain the power spectral density (PSD). The PSD gives the frequency distribution of the noise, which is required to determine how the noise in one molecular species is processed as it cascades through subsequent gene circuits (Simpson et al., 2003). The single-sided (positive frequency only) PSD, $S_{\overline{r}r}(f)$, for the intrinsic noise of this process is found from Eq. (A.17) of Appendix A as

$$S_{\overline{r}r}(f) = \frac{4\overline{r}(t)}{(\gamma_R)^2 \left(1 + (2\pi/f)/(\gamma_R)^2\right)^2},$$

(6)

where $\overline{r}(t)$ is the time average of $r(t)$.

3. Transcriptional regulation

A more realistic description of transcription would have switching between discrete high and low transcription rates with the average rate determined by the fractional amount of time spent in each of the two states. This model is consistent with transcription controlled through protein–DNA interactions at an operator site within the gene promoter region. We consider the case of a single operator site within a single copy of the operon as shown in Fig. 2a. The operator has only two possible states, which we denote here as $O$ (unbound) and $O'$ (bound), and transition between these states is described by

$$O \xrightarrow{k_{d}} O',$$

(7)

where $k_f$ and $k_r$ are rate constants for the forward and reverse reactions, respectively, and $d$ is the population of the molecular species (inducer) that binds to $O$. Eq. (7) applies to both positive and negative regulation as the fully induced state may be either $O$ or $O'$. For the sake of brevity, we consider only the positively regulated case ($O' \rightarrow$ fully induced) here. The negatively regulated derivation follows the same reasoning.

As there is only one operator site, the population of the bound operator, $O'(t)$, is either one or zero at all times (Fig. 2b). For positive regulation we define

$$x_R = x_0, \quad O' = 0,$$

(8a)

$$x_R = x_1, \quad O' = 1,$$

(8b)
where we have included the noise of mRNA decay (equal in magnitude, but uncorrelated to synthesis shot noise (Simpson et al., 2003)), and the wideband white noise of the constant term (i.e. basal gene expression) in Eq. (8c). Eq. (10) gives the PSD of the noise in the mRNA synthesis rate. The PSD of the mRNA population is found below: see discussion.

To define the second of the FD-CLEs for this system, we find \( \overline{\sigma}(t) \) as described in the previous section as

\[
\overline{\sigma}(t) = \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \sigma_{n}(t) = p_{1}(t),
\]

where \( p_{1}(t) \) is the marginal probability that \( O'(t) = 1 \).

From the CME we write

\[
\frac{d\overline{\sigma}(t)}{dt} = -k_{r}\overline{\sigma}(t) + k_{f} d(t)(1 - \overline{\sigma}(t)).
\]

We analyse the dynamics of Eq. (12) by considering the response to a unit step change (i.e. addition of one molecule to \( d' \) at \( t = 0 \) ) such that \( d(t) = d_{0} + U(t) \), where \( U(t) \) is a unit step function, yielding

\[
\frac{d\overline{\sigma}(t)}{dt} = -\overline{\sigma}(t)(k_{r} + k_{f}d_{0} + k_{f} U(t)) + k_{f}d_{0} + k_{f} U(t).
\]

Eq. (14) can be solved to show that

\[
\overline{\sigma}(t) = \frac{k_{f}d_{0}}{(k_{r} + k_{f}d_{0})} + \frac{(k_{f}k_{r})}{(k_{r} + k_{f}d_{0} + k_{f})} \times (1 - e^{-(k_{r}+k_{f}(d_{0}+1))t})U(t).
\]

4. Discussion

Neglecting operator contributions, the noise in the mRNA population given in Eq. (6) is exactly that found from the shot noise formulation previously used in gene circuit analysis (Simpson et al., 2003). This analysis demonstrates that the shot noise representation and the FD-CLEs are valid even as \( \overline{\sigma}(t) \to 0 \), despite concerns about their use in low molecular population situations (Arkin et al., 1998; Gillespie, 2000; Gillespie, 2002; Simpson et al., 2003). The validity of the FD-CLE at low molecular populations for this case is confirmed by the plot in Fig. 3 that shows excellent agreement between the simulated and calculated mRNA PSDs for the single gene circuit of Fig. 1 with an average population of only 0.1 mRNA molecules.

Including the operator effects, we find the single-sided PSD of noise in the mRNA population, \( S_{n}(f) \), by incorporating the gain, noise, and frequency response effects of mRNA decay (Simpson et al., 2003) into
Eq. (10) yielding

\[
S_n(f) = \frac{4(x_0 + \mathcal{O}(t)(x_f - x_0))}{(\gamma R)^2 \left(1 + \frac{2\pi f}{\gamma R} \right)^2} 
+ \frac{4(x_f - x_0)^2(\mathcal{O}(t) - (\mathcal{O}(t))^2)}{(\gamma R)^2(k_r + k_f d)} 
\times \left(1 + \frac{2\pi f}{(k_r + k_f d)} \right)^2 \left(1 + \frac{2\pi f}{(k_r + k_f d)} \right).
\]  

(15)

The first term on the right-hand side of Eq. (15) is the mRNA noise found in previous gene circuit noise analysis (Ozbudak et al., 2002; Simpson et al., 2003; Thattai and van Oudenaarden, 2001), and the second term is an additional noise component generated by the operator binding control of the mRNA synthesis rate. This noise is generated by operator binding events that cause bursts of transcription that pull the mRNA population away from the baseline level. Previously, Kepler and Elston provided a CME-based derivation of the variance in protein concentration generated by this operator noise (Kepler and Elston, 2001). However, Eq. (15) goes further by providing the frequency distribution of this noise, allowing the analysis of its effect within its own autoregulated feedback loop, or as it cascades through subsequent gene circuits (Simpson et al., 2003). As Kepler and Elston reported, fast operator transitions (i.e. \(k_r + k_f d \gg \gamma_R\)) reduce the effect of the operator generated noise. Eq. (15) demonstrates that this is accomplished by spreading the same total operator noise over a wider bandwidth where the higher frequency noise is filtered out by the mRNA filter (Fig. 4). In other words, when \(k_r + k_f d \gg \gamma_R\), the transcriptional bursts are short compared to the time constant of mRNA population change, resulting in only small excursions from the mRNA population baseline and a small noise penalty. However when \(\gamma_R \gg k_r + k_f d\), the transcriptional bursts are long compared to the time constant of mRNA population change, and the excursions from the mRNA baseline may be as large as \((x_f - x_0)/\gamma_R\), resulting in a substantially larger low-frequency noise component.
Some of the effects of operator noise are seen in Fig. 5, which shows simulated and calculated mRNA PSDs (shot, operator, and total noise) at 50% of full induction \((kr = kfd)\). For the case shown \((fr \gg kr + kfd)\), the operator noise is concentrated in the frequency band below the mRNA noise bandwidth (case (a) in Fig. 4), and completely dominates the low-frequency mRNA PSD. However, a clear inflection point is seen in Fig. 5 where the dominant noise mechanism switches from operator to shot noise, as operator noise is subject to greater attenuation at higher frequencies than shot noise (see Eq. (15)). For this case, slow stochastic fluctuations in the mRNA population are due to operator dynamics, while fast stochastic fluctuations are due to the random time and discrete nature of mRNA synthesis.

Most often it is only the slow mRNA fluctuations that affect protein noise, as the protein decay dynamics usually dominate the frequency response (Simpson et al., 2003). In this case, a single low-frequency point in the mRNA PSD taken from the horizontal portion of the total noise curve in Fig. 5 is sufficient to determine the effect of mRNA noise on protein noise. In engineering systems this measurement of noise is known as ‘spot noise’ as it is the PSD at a given ‘spot’ in frequency, as shown in Fig. 5. A criterion for conditions in which operator noise is insignificant can be derived from Eq. (15) as

\[
\alpha_I \ll \alpha_0 + \frac{O(t) + \sqrt{(O(t))^2 + 4g_0 (O(t))^2}_{kr + kfd}}\frac{\frac{O(t) - (O(t))^2}{k + kfd}}{2}. \tag{16}
\]

Examination of Eq. (16) shows that fast operator kinetics (large \(kr + kfd\)), low induced expression level \((\alpha_I)\), high induction levels \((kr + kfd \gg kr)\), and large basal expression level \((\alpha_0)\) all de-emphasize the importance of operator noise compared to shot noise. The mRNA decay rate does not show up in Eq. (16) as we are now assuming that protein dynamics dominate the frequency response. Simulated and calculated low-frequency spot noise is plotted as a function of induction level in Fig. 6a for several different values of \(kr\). For all cases the total noise approaches the shot noise limit as the induction level approaches either 0% or 100%. However, at induction levels between these two extremes, operator noise is a significant contributor to total noise.

As suggested above, for smaller values of \(kr\), the operator noise dominates total noise, which approaches a maximum at induction levels near 50%. Interestingly, even for relatively large values of \(kr\), the operator contribution pulls total noise well above the shot noise limit at low induction levels (Fig. 6a), where the average mRNA population, \(\langle R \rangle\), is small. As shown in Fig. 6b, the spot noise strength (spot noise/average mRNA population) increases quickly as induction begins, reaching maxima at induction levels of just a few percent. All of the observations above are based on the values of basal \((0.002/s)\) and induced \((0.5/s)\) transcription used here.

Smaller differences in these two transcription levels results in a smaller operator noise effect compared to shot noise as described in Eq. (15).

Equally important, from Eq. (14) the gain and frequency response of the operator circuit can be found (see Appendix B), which allows the incorporation of noise in molecular species \(d\) into mRNA noise behavior analysis. Using the relationships in Eq. (8c) and B2 from the appendix and including the gain and bandwidth effects of the transcription circuit (Simpson et al., 2003), we find that the PSD in mRNA population due to noise in \(d\), \(Sn-d(f)\), is given by

\[
Sn-d(f) = \left(\frac{(kr)(kr + kfd + kr)}{(kr + kfd)(\alpha_I - \alpha_0)}\right)^2 \times \left(\frac{Sn-d(f)}{1 + \left(\frac{2\pi f}{kr + kfd + kr}\right)^2} \cdot \left(1 + \left(\frac{2\pi f}{\alpha_0}\right)^2\right)\right)^2. \tag{17}
\]

Fig. 6. (a) Spot noise (low-frequency mRNA noise PSD) vs. induction level for the transcriptional circuit of Fig. 2 with \(fr = 0.002\) molecules/s, \(ta = 0.5\) molecules/s, and \(ta = 0.01/s\). The calculated curves (shot noise and total noise with \(kr = 0.002, 0.01, 0.05, 0.5/s\) were found using Eq. (15), while the simulated curve \((kr = 0.002/s)\) was found using the Exact Stochastic Simulator software described in Appendix C. The induction level was varied from 0% to nearly 100% by changing the value of \(kr\). (b) The normalized spot noise strength (spot noise/average mRNA population) for the curves in (a). All curves were normalized to shot noise, making the spot shot noise strength \(1\) for all levels of induction.
where $S_{n-d}(f)$ is the PSD of the noise in $d$. The total noise PSD in the mRNA population is given by the sum of Eqs. (15) and (17).

Since both the gain and frequency response of the operator circuit depends on $d_0$, Eq. (17) is only strictly valid for small variations around $d_0$. However, at relatively low populations where $k_r > k_d d_0$, Eq. (17) reduces to the constant gain and bandwidth relationship

$$S_{n-d}(f) = \left(\frac{k_r}{k_d}(a_l - a_0)\right)^2 \times \left(\frac{S_{n-d}(f)}{1 + \left(\frac{2\pi f}{\nu_R}\right)^2} \times \left(1 + \left(\frac{2\pi f}{\nu_R}\right)^2\right)ight),$$

(18)

and the small signal noise analysis using Eqs. (15) and (17) is nearly exact. At the other extreme where $k_d d_0 > k_r$, the operator noise in Eq. (15) and the gain term in Eq. (17) become negligibly small. In this state, the transcription rate is always at the fully induced level, and the noise in $d$ is largely ignored by the transcriptional circuitry. Between these two extremes, positive variations from $d_0$ see a compressed gain while negative variations see an extended gain. In this regime, the accuracy of the small signal noise analysis is suspect and must be carefully applied and interpreted.

Transcriptional regulation is often modeled using the Hill equation where (Elowitz and Leibler, 2000; Gardner et al., 2000; Ozbudak et al., 2002; Simpson et al., 2003; Thattai and van Oudenaarden, 2001)

$$a_R(t) = a_0 + \frac{a_l}{1 + (k_D/p(t))^n},$$

(19)

where $p(t)$ is the population of the transcriptional regulatory protein monomer, $k_D$ is the protein monomer population for 50% induction, and $n$ is the Hill coefficient that models cooperative binding to the operator site. The analysis here and other recently reported results demonstrate that the Hill equation neglects important circuit behavior. The Hill relationship assumes an instantaneous change in the transcription rate in response to a change in the regulatory protein population. In contrast, the transfer function in Eq. (B.2) (see Appendix B) has frequency dependency that affects circuit dynamics. This frequency dependency may be especially important for negatively autoregulated gene circuits, as it affects the closed-loop stability even if protein decay dynamics are considerably slower than operator dynamics (Simpson et al., 2003). Furthermore, Hill kinetics neglect the dynamics of protein polymerization (e.g. dimer formation), which may have a large and protein population dependent effect on frequency response and produces noise terms that may not be negligible (Cox et al., 2003). Finally, operator noise is not modeled by the Hill relationship. The advantage of the Hill relationship is that several reaction steps are collapsed into a single simple relationship. This simplification is especially useful in simulation as the required computational time and resources increase dramatically with the number of modeled reaction steps. However, to maintain accuracy in simulations, any simplified relationship describing transcriptional regulation must include the effects described above. It remains a challenge to obtain an expression that both provides the simplicity of the Hill equation and captures the noise and dynamics of transcriptional regulation.

There are limitations of the analysis techniques presented here. The random partitioning of molecules between daughter cells after cell division is not included. The biochemical models were constructed as birth and death processes with first-order rates. Higher-order interactions between biochemical species may introduce more complex gain, frequency, and noise relationships. Also, multiple operator sites within the promoter region (Shea and Ackers, 1985) were not modeled, although the analysis provided here can be applied to such systems.

However as previously observed, in spite of these limitations the frequency domain analysis tools may be used to provide insights into gene circuit architectures and to elucidate the selective pressures that influenced natural genetic circuit topology (Simpson et al., 2003). Through the analysis here and other recent results (Cox et al., 2003; Kepler and Elston, 2001) a much more detailed understanding of the dynamical and noise behavior of transcriptional regulatory circuits is emerging. The complete circuit functionality of transcriptional regulation cannot be grasped without gaining an appreciation of the frequency domain behavior of this system. Gene circuits that have evolved to minimize noise may favor fast operator binding dynamics to limit operator noise (Fig. 6) and, in negatively autoregulated circuits, to preserve the stability and functionality of the feedback loop (Simpson et al., 2003). However noise optimization may be more complex, as noise in the inducer ($d$ in our analysis) is reduced (i.e. low-pass filtered) by slower operator binding dynamics (see Eq. (17)).

Conversely, in view of the emerging appreciation of the functional role of stochastic processes in genetic circuits, perhaps in some cases the operator noise, particularly at low induction levels, plays an important role in circuit operation. Small values of $k_r$ are found in natural genetic systems (Streaker and Beckett, 2003), and Kepler and Elston have shown that operator noise can lead to either the emergence or elimination of bifurcations (Kepler and Elston, 2001). Although somewhat counterintuitive, we speculate that operator noise could play a role in very sensitive cellular molecular detection strategies. Consider a hypothetical case where the protein product of an inducible single gene circuit controls a genetic latch. In the absence of any inducer molecules, the average mRNA population would be
very small, and would only display the shot noise associated with basal gene expression. The addition of a few inducer molecules, enough to achieve just a few percent of full induction, would make a modest increase in the average mRNA population, but could lead to a substantial increase in mRNA and the resulting protein noise, perhaps large enough to trigger the genetic latch. Although such a circuit could be extremely sensitive, it would be immune to false noise firings as the excessive operator noise would only be generated in the presence of the inducer.

Finally, gene circuits and networks make radical and abrupt changes in functionality in response to changes in the populations of the molecules controlling them, such as switching from positive to negative autoregulation as the concentration of regulatory elements move from a low to high concentration (Cox et al., 2003; Dunlap and Greenberg, 1988; Ptushne, 1992). Very different sub-circuit behaviors are required to optimize or even allow these different modes of operation at the circuit or network level. A closer examination of the total circuit architecture is likely to demonstrate that the population dependent changes in the dynamical and noise properties of the sub-circuits controlling gene expression (protein polymerization (Cox et al., 2003) and operator binding) are essential elements for the proper function of gene circuits as they display these different behaviors in different operational regimes. The analytical methods presented here provide important tools needed to allow this examination.

Acknowledgements

We gratefully acknowledge funding support from the National Science Foundation (NSF) and the Defense Advanced Research Projects Agency through NSF Grant number EIA-0130843 and the Laboratory Directed Research and Development Program of Oak Ridge National Laboratory (ORNL). We are indebted to M.J. Roberts, M.J. Doktycz, G.D. Peterson, J.M. Lancaster, and M.S. Allen for several fruitful discussions related to the topics presented here. This work was partially performed at the Oak Ridge National Laboratory, managed by UT-Battelle, LLC for the U.S. DOE under Contract No. DE-AC05-00OR22725.

Appendix A. Noise in discrete random processes

A.1. Case 1: Random pulse train with no pulse pile-up

Consider the random process pulse train, $X(t)$, shown in Fig. 7 Pulses make high-to-low ($H$–$L$) and low-to-high ($L$–$H$) transitions at random times, but there is no pulse pile-up as there is a $H$–$L$ transition between each pair of $L$–$H$ transitions. The conditional probability density functions, $\rho_{H-L}(\tau_H-L/X = 1)$ and $\rho_{L-H}(\tau_L-H/X = 0)$, of the transition times are given by

$$
\rho_{H-L}(\tau_L-H/X = 1) = k_H-L e^{-k_H-L \tau_L-H},
$$

$$
\rho_{L-H}(\tau_L-H/X = 0) = k_L-H e^{-k_L-H \tau_L-H},
$$

where $1/k_H-L$ and $1/k_L-H$ are the average transition times, and $\tau_H-L$ and $\tau_L-H$ represent the time since the last $L-H$ or $H-L$ transition, respectively. We seek the autocorrelation function, $\phi_X(\tau)$, of the random process of $X(t)$ where

$$
\phi_X(\tau) = E[X(t)X(t+\tau)],
$$

$E[\cdot]$ denotes the expected value of the function within the brackets, and we have assumed that $X(t)$ is wide-sense stationary.

Assuming that $X(t)$ is an ergodic process, we evaluate Eq. (A.2) by slicing $X(t)$ into $N(\rightarrow \infty)$ intervals and displacing each interval to $t = 0$ as shown in Fig. 8. Then,

$$
\phi_X(\tau) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^{N} X_n(0)X_n(\tau).
$$

Since the only possible values for $X_n(t)$ are 1 and 0 (see Fig. 7), the probabilities, $p_1$ and $p_0$, of finding $X_n(0) = 1$ or $X_n(0) = 0$ are

$$
p_1 = E[X(t)] = \overline{X(t)},
$$
\[ p_0 = 1 - X(t). \]  \hspace{1cm} \text{(A.4)}

Members of the ensemble where \( X_n(0) = 0 \) do not contribute to \( \Phi_X \), so Eq. (A.3) becomes

\[
\Phi_X(t) = p_1 \left( \lim_{M \to \infty} \frac{1}{M} \sum_{m=1}^{M} X_m(t) \right) = \frac{X(t)}{p_1(\tau/X_m(0) = 1)},
\]

where \( X_m \) terms are the members of the \( X_n \) ensemble where \( X_n(0) = 1 \). \( p_1(\tau/X_m(0) = 1) \) is the conditional probability that \( X_m(\tau) = 1 \) given that \( X_m(0) = 1 \), and can be found using

\[
\frac{dp_1(\tau)}{d\tau} = -k_{H-L}p_1(\tau) + k_{L-H}p_0(\tau) = -k_{H-L}p_1(\tau) + k_{L-H}(1 - p_1(\tau)), \hspace{1cm} \text{(A.6)}
\]

by setting the initial condition \( p_1(0) = 1 \). Eq. (A.6) is easily solved to show that

\[
p_1(\tau/X_m(0) = 1) = p_1(\tau) |_{p(0)=1} = X(t)(1 - X(t))e^{-(k_{H-L}+k_{L-H})\tau}. \hspace{1cm} \text{(A.7)}
\]

Combining the results of Eqs. (A.5) and (A.7) we get

\[
\Phi_X(t) = \frac{(X(t) - (X(t))^2)e^{-(k_{H-L}+k_{L-H})|\tau|} + (X(t))^2}{\sigma_X^2} \]

where \( \sigma_X^2 \) is the variance of \( X(t) \), and we have used the relationship that the first and second (as well as higher) moments of \( X(t) \) are equal since one and zero are the only allowed values of \( X(t) \). The result of Eq. (A.8) was extended to negative values of \( \tau \) by noting that autocorrelation is an even function (use of \(|\tau|\) in the exponential). If we assume that the signal is contained in the average value, then the autocorrelation function of just the noise in \( X(t) \), \( \Phi_{n-X}(\tau) \), is

\[
\Phi_{n-X}(\tau) = \frac{(X(t) - (X(t))^2)e^{-(k_{H-L}+k_{L-H})|\tau|}}{\sigma_X^2}. \hspace{1cm} \text{(A.9)}
\]

It is important to note that the variance does not depend on the time-scale of the transitions; the same variance will be found from either rapid or slow transitions between states. However, the frequency content of the noise varies considerably with the time-scale of the transitions.

The power spectral density (PSD), \( S_{n-X}(f) \), is given by

\[
S_{n-X}(f) = \int_{-\infty}^{\infty} \Phi_{n-X}(\tau)e^{-i2\pi ft} d\tau = \sigma_X^2 \int_{-\infty}^{\infty} e^{-(k_{H-L}+k_{L-H})|\tau|} e^{-i2\pi ft} d\tau
\]

\[
= \frac{2(\overline{X(t)} - (\overline{X(t)})^2)}{(k_{H-L} + k_{L-H})} \left( 1 + \frac{2\pi ft}{(k_{H-L} + k_{L-H})} \right)^2, \hspace{1cm} f \geqslant 0. \hspace{1cm} \text{(A.10)}
\]

which includes both negative and positive frequency components. For convenience, we will use the single-sided (i.e. all the noise represented in the positive frequency regime) PSD, in which case Eq. (A.10) becomes

\[
S_{n-X}(f) = \frac{4(\overline{X(t)} - (\overline{X(t)})^2)}{(k_{H-L} + k_{L-H})} \times \left( 1 + \frac{2\pi ft}{(k_{H-L} + k_{L-H})} \right)^2, \hspace{1cm} f \geqslant 0. \hspace{1cm} \text{(A.11)}
\]

This equation shows that the total noise (i.e. the variance) is only a function of \( \overline{X(t)} \), but this noise is concentrated within an ever decreasing frequency band as the time scale of the transitions becomes longer.

A.2. Case 2: Random pulse train with pulse pile-up

Next we consider the case where pulse pile-up is allowed as shown in Fig. 9. Once again, in a random process, \( Y(t) \), pulses make \( H-L \) and \( L-H \) transitions at random times. However, we allow multiple \( H-L \) and \( L-H \) transitions between each \( L-H \) transition. For reasons that are clear in the biological context of the main text, we define \( k_{L-H} \) as a constant, while \( k_{L-H} = \gamma Y(t) \), where \( \gamma \) is a constant. Once again we assume this random process is wide-sense stationary and ergodic.

To simplify this analysis by making use of the treatment above, we conceptualize this process as the

![Fig. 9. Example random process for Case 2.](image-url)
Since the amplitude of each pulse is unity, all the component processes \( Y_1(t), Y_2(t), \ldots, Y_N(t) \) is characterized with rate constants given by

\[
k_{L-H,n} = \frac{k_{L-H}}{N},
\]

\[
k_{H-L,n} = \gamma Y_n(t) \tag{A.12}
\]

and

\[
Y_n(t) = \frac{Y(t)}{N} = k_{L-H}\gamma N.
\]

\[
Y^2_n(t) = \frac{Y^2(t)}{N} = \frac{k_{L-H}}{\gamma N} = Y_n(t),
\]

\[
\sigma^2_{Y_n} = \frac{k_{L-H}}{\gamma N} - \left( \frac{k_{L-H}}{\gamma N} \right)^2 \xrightarrow{N \to \infty} \overline{Y_n(t)} \tag{A.13}
\]

Since the amplitude of each pulse is unity, all the moments, and in particular the second moment in Eq. (A.13), of \( Y_n(t) = \overline{Y_n(t)} \). If we allow \( N \to \infty \), there are no pulse pile-ups in the component processes, and with the proper substitutions of variables, the autocorrelation functions (total process and noise only) of each \( Y_n(t) \) are given by Eqs. (A.8) and (A.9) as

\[
\phi_{Y_n}(\tau) = \frac{Y(t)}{N} e^{-(\gamma + (k_{L-H} - \gamma)/N)|\tau|} + \left( \frac{Y(t)}{N} \right)^2 \tag{A.14a}
\]

\[
\phi_{n-Y_n}(\tau) = \frac{\overline{Y(t)}}{N} e^{-(\gamma + (k_{L-H} - \gamma)/N)|\tau|}. \tag{A.14b}
\]

Care must be exercised when obtaining the autocorrelation functions for the total process by summation of the component autocorrelation functions. The noise of the \( Y_n \) processes are uncorrelated, and the \( \phi_{n-Y_n}(\tau) \) terms may be summed directly to obtain

\[
\phi_{n-Y}(\tau) = \lim_{N \to \infty} N\phi_{n-Y_n}(\tau) = \overline{Y(t)} e^{-(\gamma)|\tau|}. \tag{A.15}
\]

However, the constant components (i.e. the baseline as \( |\tau| \to \infty \)) in the \( \phi_{Y_n} \) terms are correlated, and must be added linearly, rather than in quadrature. So,

\[
\phi_Y(\tau) = \lim_{N \to \infty} \left( N \frac{Y(t)}{N} e^{-(\gamma + (k_{L-H}-\gamma)/N)|\tau|} + N^2 \left( \frac{Y(t)}{N} \right)^2 \right) = \overline{Y(t)} e^{-(\gamma)|\tau|} + \left( \overline{Y(t)} \right)^2 \tag{A.16}
\]

With the appropriate change of variables the single-sided PSD of the noise in \( Y(t) \) is given by Eq. (A.11) as

\[
S_{n-Y}(f) = \frac{4\overline{Y(t)}}{\gamma} \frac{1}{1 + (2\pi f/\gamma)^2}, \quad f \geq 0. \tag{A.17}
\]

A.3. Case 3: Random pulse train of Dirac delta functions

As it is important in determining the noise in molecular synthesis rates, we consider a special case of a random train of pulses with amplitude \( A \) and fixed width \( T \). Since we have now fixed the interval for the \( H-L \) transition and assuming no pulse pile-up, the autocorrelation function falls linearly as \(|\tau|\) increases from zero, becoming

\[
\phi_Y(\tau) = (ATk_{L-H})^2, \quad |\tau| > T, \tag{A.18}
\]

In the limit as \( A \to \infty, T \to 0, \) and \( AT \to 1, \) the pulses become a randomly spaced train of impulses and validates our assumption of no pulse pile-up. In this limit the autocorrelation function becomes

\[
\phi_Y(\tau) = k_{L-H}^2 \delta(\tau) + (k_{L-H})^2, \tag{A.19a}
\]

\[
\phi_{n-Y}(\tau) = k_{L-H}^2 \delta(\tau), \tag{A.19b}
\]

where \( \delta(\tau) \) is the Dirac delta function. We find the single-sided PSD of the noise by Fourier transform of Eq. (A.19b) to obtain

\[
S_{n-Y}(f) = 2k_{L-H}. \tag{A.20}
\]

A.4. Case 4: Random pulse train of gated Dirac delta functions

As a final special case important in gene circuit analysis, we consider a pulse train of Dirac delta functions that are gated (i.e. turned on and off) by a random train of pulses like those described in Case 1 above (Fig. 11). As long as the two processes are statistically independent, the autocorrelation function of the composite function, \( \phi_{XY}(\tau) \), is found as

\[
\phi_{XY}(\tau) = \phi_X(\tau)\phi_Y(\tau). \tag{A.21}
\]
which we find using Eqs. (A.8) and (A.19a) as
\[
\phi_{XY}(\tau) = ((X(t) - \langle X(t) \rangle)^2 e^{-(k_{XH-L} + k_{XL-H})\langle \tau \rangle} + (X(t))^2) 
\times (k_{YH-L} \delta(\tau) + (k_{YH-L})^2) 
= (\langle X(t) \rangle k_{YH-L} \delta(\tau) + (\langle X(t) \rangle)^2) 
\times e^{-(k_{XH-L} + k_{XL-H})\langle \tau \rangle} (k_{YH-L})^2 + (\langle X(t) \rangle)^2 (k_{YH-L})^2.
\]
(A.22)

As before, we find the autocorrelation function of just the noise as
\[
\phi_{n,XY}(\tau) = \langle X(t) \rangle k_{YH-L} \delta(\tau) + (\langle X(t) \rangle)^2 
\times e^{-(k_{XH-L} + k_{XL-H})\langle \tau \rangle} (k_{YH-L})^2 
\]
and the single-sided PSD of the noise is
\[
S_{n,XY}(f) = 2 \langle X(t) \rangle k_{YH-L} \times \frac{4k_{YH-L}^2 (\langle X(t) \rangle)^2}{(k_{XH-L} + k_{XL-H})^2} 
\times \left(\frac{1}{1 + (2\pi f / (k_{XH-L} + k_{XL-H}))^2}\right).
\]
(A.23)

\(\phi_{n,XY}(\tau)\) and \(S_{n,XY}(f)\) are the frequency domain (FD) CLEs for transcription rate in frequency space. Eqs. (B.1) and (6) taken together are the frequency domain (FD) CLEs for this system.

In the frequency domain, Eq. (5) describes a low-pass filter with a transfer function, \(H_R(f)\) (\(df/d\gamma_R\) as a function of \(f\)), given by (Simpson et al., 2003)
\[
H_R(f) = \left(\frac{1}{\gamma_R}\right) \left(\frac{1}{1 + (2\pi f / \gamma_R)^2}\right).
\]
(B.1)

where \(f\) is the frequency in Hertz. This function describes how the mRNA population varies with changes in the transcription rate in frequency space. Eqs. (B.1) and (6) taken together are the frequency domain (FD) CLEs for this system.

In the frequency domain the step response of Eq. (14) describes a single pole system with a small signal transfer function, \(H_D(f)\) (change in \(\bar{C}(t)\) in response to a unit step change in \(d\) as a function of \(f\)), given by
\[
H_D(f) = \left(k_i k_f / (k_i + k_f d_0 + k_f)\right) 
\times \left(\frac{1}{1 + (2\pi f / (k_i + k_f d_0 + k_f))^2}\right).
\]
(B.2)

Eqs. (B.2) and (14) are the FD-CLEs describing this system.

**Appendix C. Simulations and parameter selection**

Stochastic simulations shown in Figs. 3, 5 and 6 were conducted using an optimized version (Gibson and Bruck, 2000) of Gillespie’s algorithm (Gillespie, 1977) via Exact Stochastic Simulator (ESS) software available for download at biocomp.ece.utk.edu. In Fig. 3, the circuit shown in Fig. 1 was simulated using values of \(x_R\) and \(\gamma_R\) (0.02 molecules/s and 0.2/s, respectively) selected to yield a mean mRNA population of 0.1 molecules. PSDs of the time series were calculated via fast Fourier transformation of 2000 autocorrelation values with a sampling time of 1 s. In Figs. 5 and 6, the gene circuit shown in Fig. 2 was simulated. The selected value of \(\gamma_R\) (0.01/s) corresponds to a half-life of 69.3 s, which falls within the range of 1–3 min reported for prokaryotes (Stephanopoulos et al., 1998). The value of \(k_i\), selected to be a factor of 5 smaller than \(\gamma_R\), yields a half-life similar to those reported for operator–protein complexes in the literature (Streaker and Beckett, 2003). Values of \(z_0\) and \(z_1\) were selected to yield basal and fully induced average mRNA populations of 0.2 and 50, respectively. PSDs of the time series were found via fast Fourier transformation of 32,000 autocorrelation values found from the simulated time series data with a sampling time of 0.5 s.

**References**


