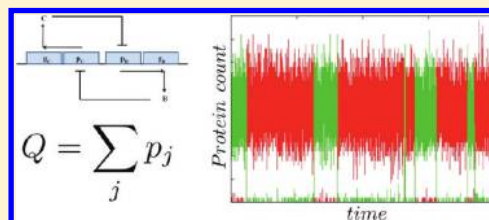


Modeling Stochastic Dynamics in Biochemical Systems with Feedback Using Maximum Caliber

S. Pressé,^{*,†} K. Ghosh,[‡] and K. A. Dill[†][†]Department of Pharmaceutical Chemistry, University of California, San Francisco, California, United States[‡]Department of Physics and Astronomy, University of Denver, Denver, Colorado, United States

Supporting Information

ABSTRACT: Complex feedback systems are ubiquitous in biology. Modeling such systems with mass action laws or master equations requires information rarely measured directly. Thus rates and reaction topologies are often treated as adjustable parameters. Here we present a general stochastic modeling method for small chemical and biochemical systems with emphasis on feedback systems. The method, Maximum Caliber (MaxCal), is more parsimonious than others in constructing dynamical models requiring fewer model assumptions and parameters to capture the effects of feedback. MaxCal is the dynamical analogue of Maximum Entropy. It uses average rate quantities and correlations obtained from short experimental trajectories to construct dynamical models. We illustrate the method on the bistable genetic toggle switch. To test our method, we generate synthetic data from an underlying stochastic model. MaxCal reliably infers the statistics of the stochastic bistability and other full dynamical distributions of the simulated data, without having to invoke complex reaction schemes. The method should be broadly applicable to other systems.



INTRODUCTION

We describe a method for modeling stochastic chemical and biochemical reactions that have feedback. By feedback, we mean situations where the output of one reaction drives forward or inhibits another reaction. Feedback is an important component of bistability, oscillations, signal amplification, noise mitigation, and biological regulation. More particularly, we focus on few-particle systems, where the raw experimental data is in the form of stochastic trajectories, and for which dynamical fluctuations can be significant. Our approach to modeling such systems aims to make minimal assumptions about the mathematical forms of the feedback mechanism, and to put maximal reliance on the experimental data itself.

Key examples of stochastic feedback in biology include the bifurcation points on fitness landscapes that account for the origins of new biological species, the mosaic patterning of the cells within the eye of the fruit fly,^{1,2} bacterial evolution through a process called bet-hedging,^{3,4} and many others.^{5–10} One important class of problem is stochastic bistable systems.^{11,12} These systems are driven by small fluctuations and feedback to toggle between stable states.

Often an underlying feedback mechanism is unknown or difficult to determine. In dynamical modeling, it is common to assume an arbitrary nonlinear mathematical form, for example, a Hill-type model of cooperativity in a mass-action model. There are two problems with such approaches. First, often the assumed mechanism cannot be validated independently. Second, even when such a model captures the average dynamical behavior, there is no guarantee that it will accurately describe the fluctuations.

There are methods that go beyond macroscopic mass-action models to capture dynamical fluctuations to model few-particle systems. Such methods are relevant in studying the stochastically induced switches of bistable systems mentioned earlier. A principal approach involves using master equations¹³ or simulations of them using Gillespie's algorithm.¹⁴ However, the main requirement of this approach is prior knowledge of (1) the "topology" of the reaction (i.e., a set of "process" arrows interrelating "states", such as substrates, products, and catalysts), and (2) the values of the corresponding rate coefficients.

Many methods are thus focused on extracting rates, network topologies or other features of chemical reactions. In the method of correlated metric construction (CMC),^{15,16} time correlations in the concentrations of chemical species are used to infer the reaction topology at the mass action level. Species concentrations are perturbed randomly, and correlations of relaxations are monitored. The method, however, does not treat feedback, which involves more complex correlations that can arise from underlying physical time delays or memory or multistability.¹⁵ In other cases, the aim of dynamical modeling is to assume a reaction topology and to find rates that best fit some existing experimental data trajectories.¹⁷ For example, maximum-likelihood (ML) methods^{18–21} are commonly used to analyze ion-channel or fluorescence experiments.²² There are methods to discriminate between assumed model topologies.²⁰ One

Received: November 21, 2010

Revised: March 15, 2011

Published: April 27, 2011

disadvantage is that often a large number of different models will fit the same data equally well.²³

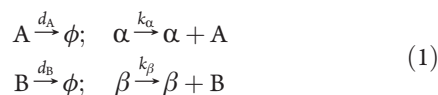
We describe a different (but related) approach called the Maximum Caliber (MaxCal) method, which is the dynamical analogue of Maximum Entropy.^{25,24,26–34} The MaxCal method has the advantage that the model building is data-driven and provides an unambiguous recipe for incorporating information on correlations between observable species. MaxCal bypasses the need to know the functional form or the exact details of feedback mechanism in advance, and it constructs a minimal stochastic model from the data. To illustrate the MaxCal method, we focus on an engineered cell-biological system called the genetic toggle switch,^{35–38} which has been explored in recent experiments by Gardner et al.³⁵ The toggle-switch has two very different time scales: fast time-scale fluctuations can trigger the system to switch from one state to the other, and a longer time scale over which the system remains stable in a single state. Recently Otten and Stock^{32,33} used MaxCal to infer the most likely discrete Markov model giving rise to a particular signal decay. Here our goal is different; we are interested in building an approximate model for complex systems, usually involving feedback, that will infer long-time statistical properties of the system given short-time trajectories.

METHODS

A Bistable Biological Circuit: The Genetic Toggle Switch.

Our basic approach here is to create an underlying model of the genetic toggle switch, simulate it using Gillespie's algorithm, and then treat those trajectories as our "data". We use the generated data, without further inputs about the model used to generate the data, to build a MaxCal model that will make predictions regarding quantities that have not been used to parametrize it.

We begin by introducing the detailed model of the toggle switch we use to simulate the data. The genetic toggle switch is a bistable biological circuit whose basic topology can be described using two promoters, p_A and p_B , and their respective transcription products, A and B. When the gene promoter p_A is activated, it produces a transcription factor, protein A, which then inhibits the promoter of a second transcription factor, protein B. Alternatively, when the promoter region p_B is activated, it produces a transcription factor protein B, which then inhibits the production of protein A. The feedback network is described in Figure 1. What follows is a proposed set of chemical reactions for a genetic toggle switch:^{39,40}



This system has three types of reactions: (1) $A \rightarrow \phi$ indicates that A molecules are degraded with rate d_A ; similarly for B. (2) $\alpha \rightarrow \alpha + A$ indicates that protein A is produced at rate k_α if the promoter α is active. Similarly $\beta \rightarrow \beta + B$ indicates that protein B is produced at rate k_β if the promoter β is active. (3) $A + \beta \leftrightarrow \beta^*$ indicates that protein A acts as a catalyst to convert B's active promoter β to inactive promoter β^* with rate f_A . The inactivation is released with rate r_A . Similarly, $B + \alpha \leftrightarrow \alpha^*$ indicates that

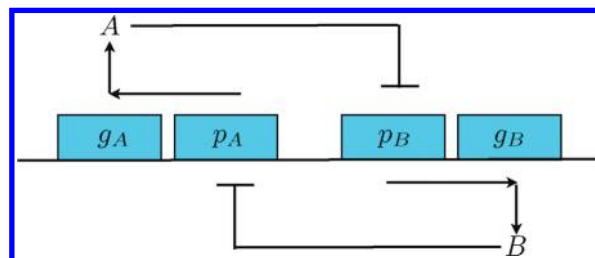


Figure 1. The genetic toggle switch. DNA plasmid is shown with promoters p_A and p_B of genes g_A and g_B that, when transcribed, produce proteins A and B, respectively. The gene–promoter complex for species A(B) is denoted $\alpha(\beta)$ in eq 1. In this gene circuit, production of A inhibits B or production of B inhibits A. In eq 2, this is shown by having A bind to the gene–promoter complex of B, and vice versa.

protein B converts A's active promoter α to an inactive form α^* with rate f_B . The inactivation is released with rate r_B . Conservation requires that $[\alpha] + [\alpha^*] = 1$ and $[\beta] + [\beta^*] = 1$, where the brackets indicate concentrations of the promoters. For the exclusive toggle switch (ETS), we assume the additional constraint $[\alpha^*] + [\beta^*] \leq 1$. With the latter constraint, the only possible steady macroscopic states are (high A and low B) or (low A and high B).⁴¹

On one hand, the experiments on the toggle switch reveal that it is a bistable system. On the other hand, despite the fact that this is an engineered system, the physical basis for the underlying bistability in experiments is not known. At the mass action level, Gardner et al.³⁵ first invoked cooperative binding of transcription factors to promoters to account for the observed bistability. However, there is no independent experimental validation of that mechanism. Later stochastic modeling, based on master equations, showed that invoking binding cooperativity was unnecessary. The bistability can result from fluctuations alone.^{39,40} In this case, as in many other cases of dynamical modeling, it is not known *a priori* what the physical basis for complex dynamical behaviors is. Dynamical modeling is often done by first assuming a particular set of reactions (the topology) or by assuming mathematical functional forms for the rate equations (such as Michaelis–Menten or Hill terms), and then by choosing parameters for those models that best fit the experimental data. The downside to this approach is that it provides no validation for the underlying model, because the parameters are often not obtainable independently.

We note particular advantages in using MaxCal to analyze the data generated for the ETS: (1) the MaxCal model is parametrized from the data directly, (2) the mathematical forms of the rate expressions need not be known in advance, (3) the fluctuations and higher moments of the dynamical distribution function are determined directly from the model—they provide additional information for validating the model against the experimental data, and (4) the long-time dynamical behavior can be inferred from shorter-time trajectories. In this case we will assume that the simulated data shows the production and disappearance of A and B as a function of time. Without asserting a particular mechanism, i.e., reaction network topology, MaxCal uses this information to build a minimal model.

Modeling Bistable Systems Using MaxCal. The method of MaxCal is described in detail elsewhere^{24,27,29,30,28,32,31,33} and in the Supporting Information of this paper. We summarize it briefly here. In equilibrium statistical mechanics, Jaynes showed²⁶ that the Boltzmann distribution law can be derived

from the principle (Maximum Entropy) that all the equilibrium microstates of a system should be taken to be equivalent, except insofar as the system must satisfy some constraint on an average equilibrium quantity (in the canonical ensemble, the fixed quantity is the temperature, or equivalently, the average energy, $\langle U \rangle$). In nonequilibrium statistical mechanics, MaxCal asserts that all the trajectories (pathways) of a system should be taken as equivalent, except insofar as the system must satisfy some constraint on an average dynamical quantity, such as an average rate. In equilibrium statistical mechanics, the central quantity is a partition function, which is a weighted sum over all the accessible microstates. In MaxCal, the central quantity is a dynamical partition function, which is a weighted sum over all the possible pathways.

We briefly outline a sketch of how one computes the dynamical partition function in MaxCal and later discuss the particular case of the ETS. We begin by defining $p_\Gamma(t)$, or simply p_Γ as the probability of observing trajectory, or pathway, Γ within some small time interval $[t, t + \delta t]$.³² We assume that all possible trajectories within the interval $[t, t + \delta t]$ are equally likely until a measurement says otherwise. Our trajectory ensemble is therefore maximally uncertain until data is used to constrain this ensemble.

In order to compute the trajectory ensemble, we maximize the caliber, $C(p_\Gamma)$, which is the measure $-\sum_\Gamma p_\Gamma \ln p_\Gamma$, subject to constraints on average observables, $\langle A_n \rangle$,

$$C = -\sum_\Gamma p_\Gamma \ln p_\Gamma + \alpha \sum_\Gamma p_\Gamma + \sum_{n,\Gamma} \lambda_n A_{n,\Gamma} p_\Gamma \quad (3)$$

where λ_n and α are Lagrange multipliers. In the absence of data, the first term in the above, $-\sum_\Gamma p_\Gamma \ln p_\Gamma$, simply asserts all trajectories are equally likely upon maximizing the caliber. Otherwise, given the data, this term allows trajectories to be reweighted in the least biased way possible according to Jaynes' prescription.²⁵

Maximizing C with respect to $\{p_\Gamma\}$, yields

$$p_\Gamma = Q^{-1} e^{\sum_n \lambda_n A_{n,\Gamma}} \quad \text{with} \quad Q \equiv \sum_\Gamma e^{\sum_n \lambda_n A_{n,\Gamma}} \quad (4)$$

where Q is the dynamical analogue of the equilibrium partition function. The dynamical partition function infers cumulants, subscripted c , of dynamical observables as follows:

$$\langle A_n^m \rangle_c = \partial^m \log Q / \partial \lambda_n^m \quad (5)$$

where m is the cumulant order.

We now treat the ETS. The raw data takes the form of observed time traces of numbers of A and B molecules. The numbers of A or B molecules go up when such molecules are produced and down when such molecules are degraded. Figure 3a shows an example time trace. In our MaxCal treatment, we begin by subdividing the time trace into small time slices of length δt . We do not know *a priori* the underlying nature of the feedback between A and B. We define two binary indicator variables for each species: one for production and one for degradation. Let ℓ_α be 0 when A is not produced within a small interval of length δt , or 1 otherwise. Correspondingly, ℓ_β is 0 when B is not produced, and ℓ_β is 1 otherwise. We choose the time interval, δt , to be sufficiently small so that no more than a single A or B is produced within it. Next, we assign variables $\ell_{i,A}$ and $\ell_{i,B}$ to each of the i individual protein molecules of type A and B, respectively. Let the variable $\ell_{i,B}$

equal 1 when the i th particle is not degraded within interval δt or 0 otherwise.

We have introduced four types of indicator variables. We therefore require four constraints to represent the observable average protein number rise and decrease, for both types of protein, in the time trace. These observables are the production rates and degradation rates of A and B. Lagrange multipliers (h_α, h_A) enforce the average observed production and degradation rates of protein A while (h_β, h_B) enforce the average observed production and degradation rates of protein B. For example, the statistical weight for production of a single A, say, within δt , is $\exp(h_\alpha)$. (Since the time interval is chosen to be short, this statistical weight for a single event is small.) Now, we construct the dynamical partition function as the weighted sum over all the possible microtrajectories within interval δt :

$$Q = \sum_{\substack{\ell_\alpha, \ell_\beta \\ \{\ell_A\}, \{\ell_B\}}} \mathcal{Z}(\ell_\alpha, \ell_\beta, \{\ell_A\}, \{\ell_B\}) \exp \left(h_\alpha \ell_\alpha + h_\beta \ell_\beta + h_A \sum_{i=1}^{N_A} \ell_{i,A} + h_B \sum_{i=1}^{N_B} \ell_{i,B} \right) \quad (6)$$

where the function \mathcal{Z} captures the so-far-unspecified correlations between proteins of type A and B, and the exponential term enforces the four average observables from the time traces. We direct the reader to the Supporting Information for a more detailed derivation of the dynamical partition function as well as for a discussion regarding why first moments, instead of higher order moments, of the production and degradation rates of A and B are constrained.

Prior to defining \mathcal{Z} , we first note that this MaxCal partition function gives the correct master equation for uncorrelated systems ($\mathcal{Z} \equiv 1$). In that case, it is shown (see Supporting Information for detailed derivation as well as ref 32) that eq 6 reduces to the master equations for the A and B birth–death process shown in eq 1:

$$\begin{aligned} \dot{P}(N_A, t) &= k_\alpha P(N_A - 1, t) + d_A(N_A + 1)P(N_A + 1, t) \\ &\quad - k_\alpha P(N_A, t) - d_A N_A P(N_A, t) \\ \dot{P}(N_B, t) &= k_\beta P(N_B - 1, t) + d_B(N_B + 1)P(N_B + 1, t) \\ &\quad - k_\beta P(N_B, t) - d_B N_B P(N_B, t) \end{aligned} \quad (7)$$

where

$$\exp(h_X) \sim \delta t k_X \quad (8)$$

is small (see above) and where $X = (\alpha, \beta)$. Similarly, $\exp(-h_A) = \delta t d_A$ (note the sign difference from eq 8) and $\exp(-h_B) = \delta t d_B$.

Introducing Correlation. We need to specify \mathcal{Z} . From the experimental time traces, it is clear that when A is high, B is low, and vice versa. Therefore the variables which coincide to birth and death events of A ($\ell_{\alpha/\ell_{i,A}}$) are coupled to those of B ($\ell_{\beta/\ell_{i,B}}$). Possible couplings are $\{\ell_{\alpha/\beta}, \ell_{i,A}/\ell_{j,B}, \ell_{\alpha/\ell_{i,B}}, \ell_{\beta/\ell_{i,A}}\}$. Analyses of the ETS time traces (“the data”), depicting levels of A and B as a function of time, show that the more A is present, the less B is produced (and vice versa) within a small time interval. Therefore, birth of A and birth of B event coupling, $\ell_{\alpha/\beta}$, cannot be the only, or dominant, contribution to the coupling between A and B species. Furthermore, analyses of the traces shows that the survival probability of a particular A in the low state (in other words, how long a particular A survives in the low state after being produced) is independent of the number of B's in the high state (and vice versa), thus $\ell_{i,A}/\ell_{j,B}$ also cannot be the dominant

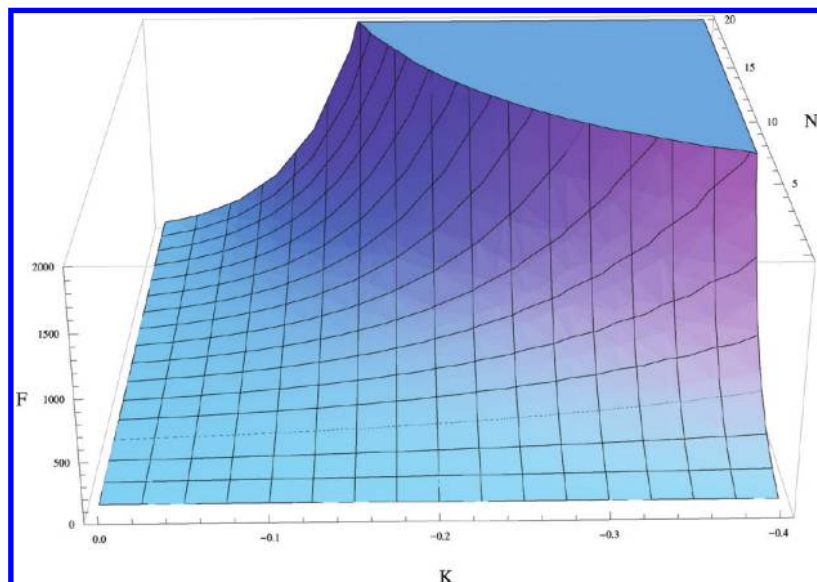


Figure 2. Effect of negative feedback strength on production fluctuations of A. Here we look at how F (production rate fluctuations normalized by its mean squared) varies with N and K using $h_\alpha = h_\beta = -4.605$, $h_A = h_B = 7.60$.

type of coupling between A and B species. We only have one type of coupling left to consider: $\{\langle \ell'_{\alpha/i,B} / \beta'_{i,A} \rangle\}$. The following definition of \mathcal{Z} is the simplest expression consistent with both observations:

$$\mathcal{Z} = \exp\left(K_{A\beta} \sum_i \ell'_{\beta/i,A} + K_{B\alpha} \sum_i \ell'_{\alpha/i,B}\right) \quad (9)$$

where $K_{A\beta}$ and $K_{B\alpha}$ are the correlation coupling parameters. As with the other Lagrange multipliers, these two quantities are not known until the data extraction is performed. The formulation of eq 9 does not require invoking otherwise unknown functional forms for the rate equations.

We describe two different methods of determining the Lagrange multipliers $K_{A\beta}$ and $K_{B\alpha}$ from observed trajectories. First, we can measure the correlation $\langle \sum_i \ell'_{\beta/i,A} \rangle$ for every interval following a time point where $\langle N_A \rangle = k/d$ and $\langle N_B \rangle = 0$, for example. This average is obtained from eq 6 as follows:

$$\langle \sum_i \ell'_{\beta/i,A} \rangle = \frac{\partial}{\partial K_{A\beta}} \ln Q\left(N_A = \frac{k}{d}, N_B = 0\right) \quad (10)$$

Second, we can compute the average production of B, say, within the time interval following every data point where N_A is in the high state (specified by $N_A = k/d$, and $N_B = 0$). For k/d not an integer, we can round to the nearest integer. We select to average the production in B (or A) over the interval following a data point where $N_A = k/d$, and $N_B = 0$ (or $N_B = k/d$, and $N_A = 0$) to get good statistics. For a discussion on the sensitivity of the correlation coupling parameter to the choice of particle number for species A and B, see the first part of the Supporting Information.

The average production rate of B is given by

$$\langle \ell'_{\beta} \rangle = \frac{\partial}{\partial h_\beta} \ln Q\left(N_A = \frac{k}{d}, N_B = 0\right) \quad (11)$$

From an observed trajectory, we harvest these average quantities

to find the values of the coupling parameter Lagrange multipliers, $K_{A\beta}$ and $K_{B\alpha}$ that are consistent with the data.

RESULTS/DISCUSSION

Testing the MaxCal Approach. How can we determine whether or not the MaxCal procedure correctly extracts unique model parameters from trajectory data? We built a computer model of a system having an ETS-type bistability, with parameters that we fix in advance. From that underlying model, we generate time traces of A and B numbers using the Gillespie algorithm.¹⁴ For simplicity, we assume that the reactions are symmetric in A and B. That is, we choose parameters $d = d_A = d_B$, $k = k_\alpha = k_\beta$ and $f = f_A = f_B$, $r = r_A = r_B$ and simulate eqs 1 and 2 with the restrictions that apply to the ETS. For the symmetric case, we therefore have $h_\alpha = h_\beta$, $h_A = h_B$, and $K = K_{A\beta} = K_{B\alpha}$.

The asymmetric case is no more difficult to treat than the symmetric case, but it is more cumbersome. For example, to compute K for a given time trace, we measured $\langle \sum_i \ell'_{\beta/i,A} \rangle$ and $\langle \sum_i \ell'_{\alpha/i,B} \rangle$ and used eq 10 to find two slightly different K 's, which we averaged to find the symmetric K . For the asymmetric case we would not average both those equations. We would use both equations to solve for our two different Lagrange multipliers.

Further, we note that the ETS is one particular model of the toggle switch introduced by Lipshtat et al.³⁹ to explain features of the experiments of Gardner et al.³⁵ Other models of the toggle could be constructed that would show behavior qualitatively similar to that of the ETS. We take the simulations of the ETS as if they represented some true, albeit particular, experimental system. We then use MaxCal to build an effective model of this system, and test whether MaxCal correctly deduces the properties of the underlying system, which would require more data to determine than is needed to parametrize the model.

To apply MaxCal, we first subdivide the trajectories into time intervals of length δt . Since the steady-state of the toggle switch involves a high level of production of one of the proteins, it means that protein production must be faster than protein degradation. We select a value of δt smaller than the fastest *observable* time scale, namely, protein production. This ensures that we avoid

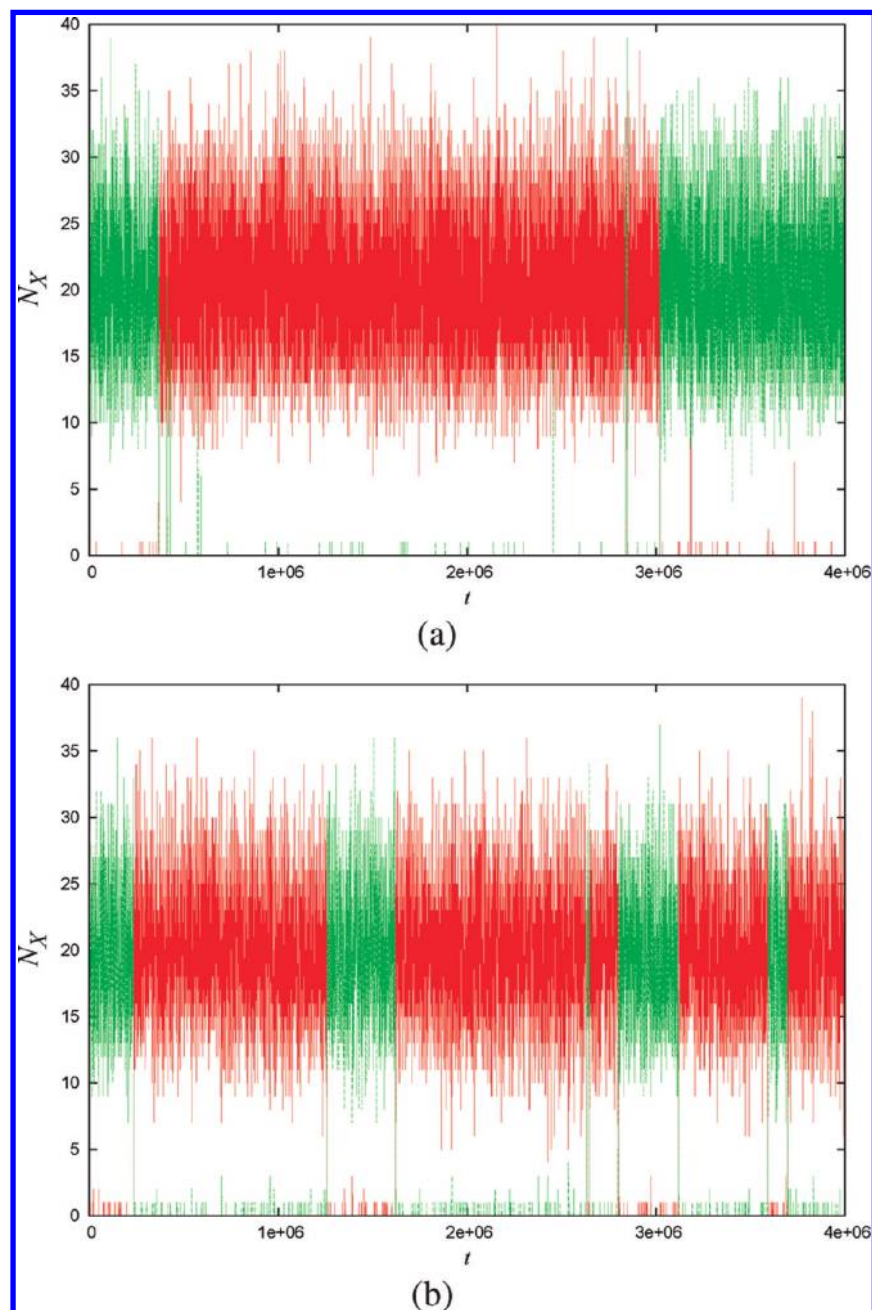


Figure 3. Levels of promoter unbound protein A (red) and levels of unbound protein B (green). (a) Toggle-switch trajectory generated by the Gillespie model. Computer-generated trajectories using the parameters $d = 0.005$, $k = 0.1$, $f = 2$, and $r = 0.01$. The time trace was broken up into time intervals $\delta t = 0.1$. Applying MaxCal to the trajectories from (a), we extracted the Lagrange multiplier $K = -0.380$ and use $h_\alpha = h_\beta = -4.605$ and $h_A = h_B = 7.6$. (b) MaxCal model with parameters extracted from the simulations. Representative trajectories generated using those values in the MaxCal model.

having multiple production events within a single δt . Thus, from the protein-number time trace, at each time point, indexed i , we have a specified number of particles of type A and B, $\{N_A(t_i), N_B(t_i), t_i\}$. From these time points, we must determine the production, degradation, and correlation Lagrange multipliers. In the absence of species B, for example, the Lagrange multipliers for production and degradation of species A are readily extracted from A-only time traces following a recipe analogous to that outlined in refs 31 and 32, for determining forward and backward hopping rates in two-state systems. As an example, this is done by breaking the A-only time trace into small time intervals and

computing h_α and h_A from average production and degradation probabilities of A within δt . Since short trajectories are sufficient to compute accurate h_α and h_A , and this is straightforward to accomplish, as outlined in Stock et al.,³² here we simply use $-h_A = \ln(\delta t d_A)$ and $h_\alpha = \ln(\delta t k_A)$ from eq 8 (see Supporting Information). Similarly for B, where $-h_B = \ln(\delta t d_B)$ and $h_\beta = \ln(\delta t k_B)$. We use the time traces in the presence of both A and B to compute K .

In Figure 2, we look at properties of our model and how sensitive our model predictions are to the numerical value of K in general. We plot the variance over mean production squared of species A, a quantity we call F , within the interval $[t, t + \delta t]$,

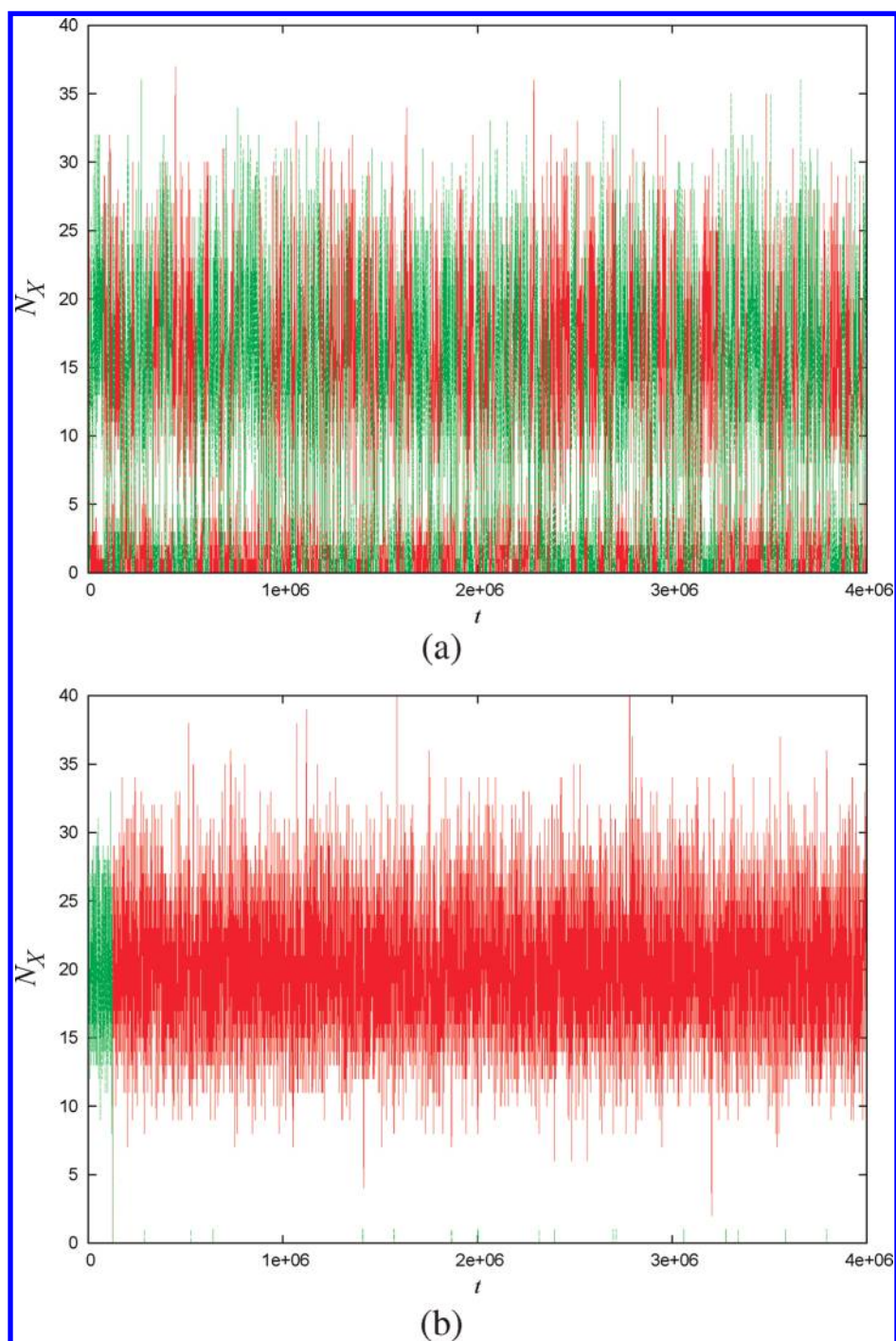


Figure 4. Correlation strength changes hopping dynamics. Given the same Gillespie model as in Figure 3a, if we now use a value of (a) K 50% smaller (-0.19) or (b) 50% larger (-0.57) than those extracted by MaxCal, the predicted model behavior is very different.

given $N_A = 0$ at time t . We define the mean production as $\langle \delta N_A \rangle = \langle N_A(t + \delta t) - N_A(t) \rangle$ and the variance as $\langle \delta N_A^2 \rangle = \langle (N_A(t + \delta t) - N_A(t))^2 \rangle_c$, where the subscript c denotes cumulants. These cumulants can be computed analytically (for example, see eq 5). We plot $F \equiv \langle \delta N_A^2 \rangle / \langle \delta N_A \rangle^2$, as a function of the number of $N = N_B$ proteins and K . It shows that fluctuations in the numbers of protein A grow dramatically larger either as the number, N , of proteins B increases or as the negative feedback is made stronger. The latter is expected for a bistable system.

Figure 3a shows the Gillespie trajectory of the populations of states A and B over time, for a particular set of parameters. We then applied MaxCal to those trajectories to extract our Lagrange multipliers by the methods described above. Figure 3b shows the MaxCal model that we extracted. At a qualitative level, MaxCal captures well the two very different time scales: the fast fluctuations within each stable state of the system, and the much longer time scale over which transitions occur from one state to the other. It also captures the sharpness of the transitions between the two stable states. Comparing Figure 4 to Figure 3 shows how even small changes in K in this model can

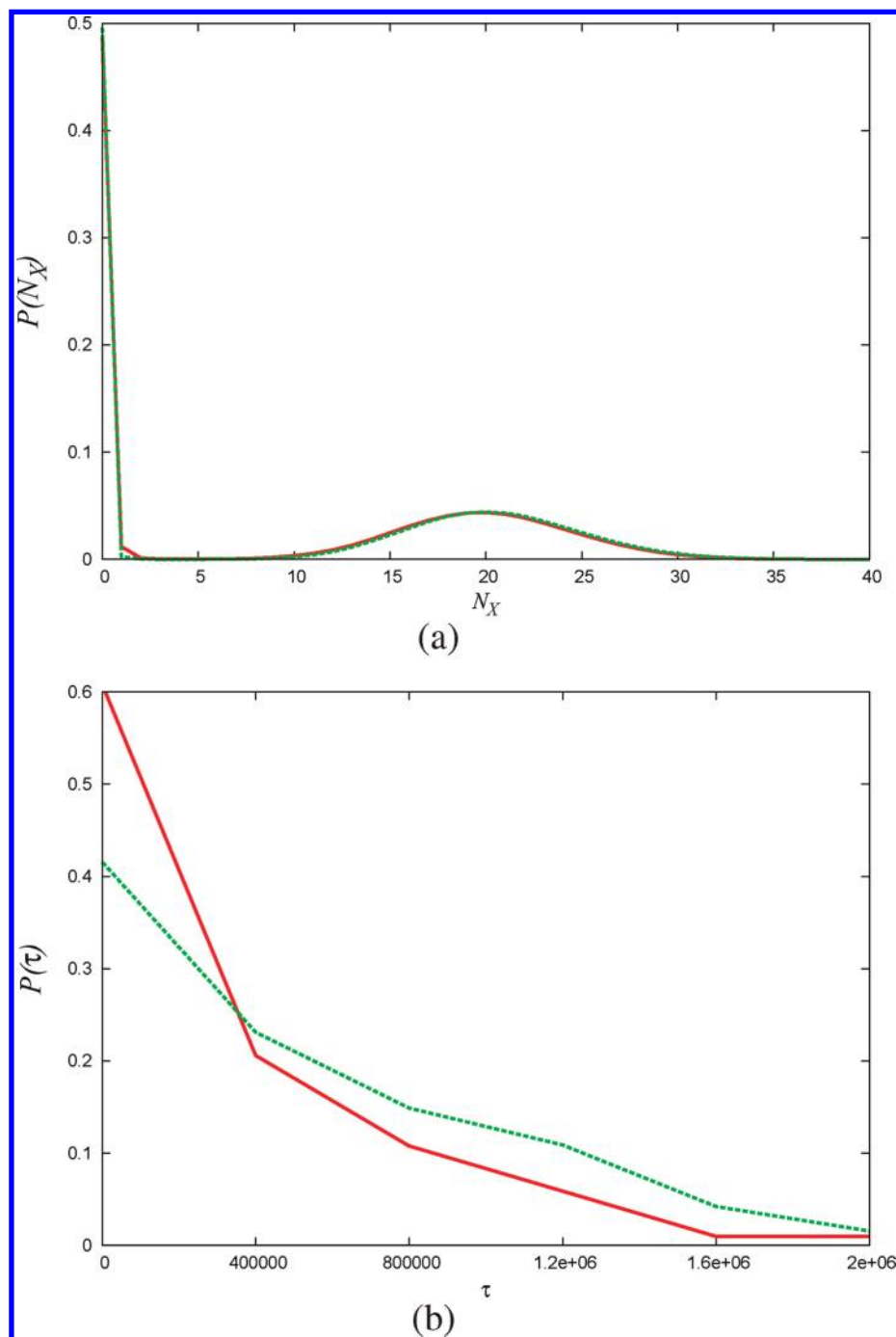


Figure 5. MaxCal predicts static and dynamical distribution functions. (a) Distribution of particle numbers. (b) Distribution of dwell times. (Green) Gillespie “raw data”. (Red) MaxCal model. Parameters are those used in Figure 3.

dramatically change the qualitative system behavior. For example, as K becomes more positive (weakening the negative feedback), the transitions become much less sharp.

As a more quantitative test, Figure 5a shows that the MaxCal model correctly computes static properties such as the particle-number distribution functions around the two states. For the dynamics, Figure 5b shows that MaxCal also gives approximately correct dwell-time distributions, even over time scales that are much longer than those used to parametrize our model.⁴²

For instance, MaxCal predicts an average dwell time of $\langle \tau \rangle = 0.44 \pm 0.43 \times 10^6$ s, while the simulated data reveals average

dwell times of $\langle \tau \rangle = 0.74 \pm 0.70 \times 10^6$ s. The dwell time uncertainty is related to the inherent dwell-time-distribution variance, not a sampling error.

Figures 6 and 7 are based on different parameter values: $d = 0.005$, $k = 0.1$, $f = 100$, and $r = 2$. In this case, the value of $K = -0.295$ that is extracted by MaxCal corresponds to a larger number of transitions between states per unit time. The average dwell time found by MaxCal is $\langle \tau \rangle = 0.11 \pm 0.11 \times 10^6$ s, while for the simulated data it is $\langle \tau \rangle = 0.16 \pm 0.17 \times 10^6$ s. As the binding–unbinding rates to the gene–promoter complex (f and r) are made large compared to typical birth–death rates, the predictions of MaxCal improve.

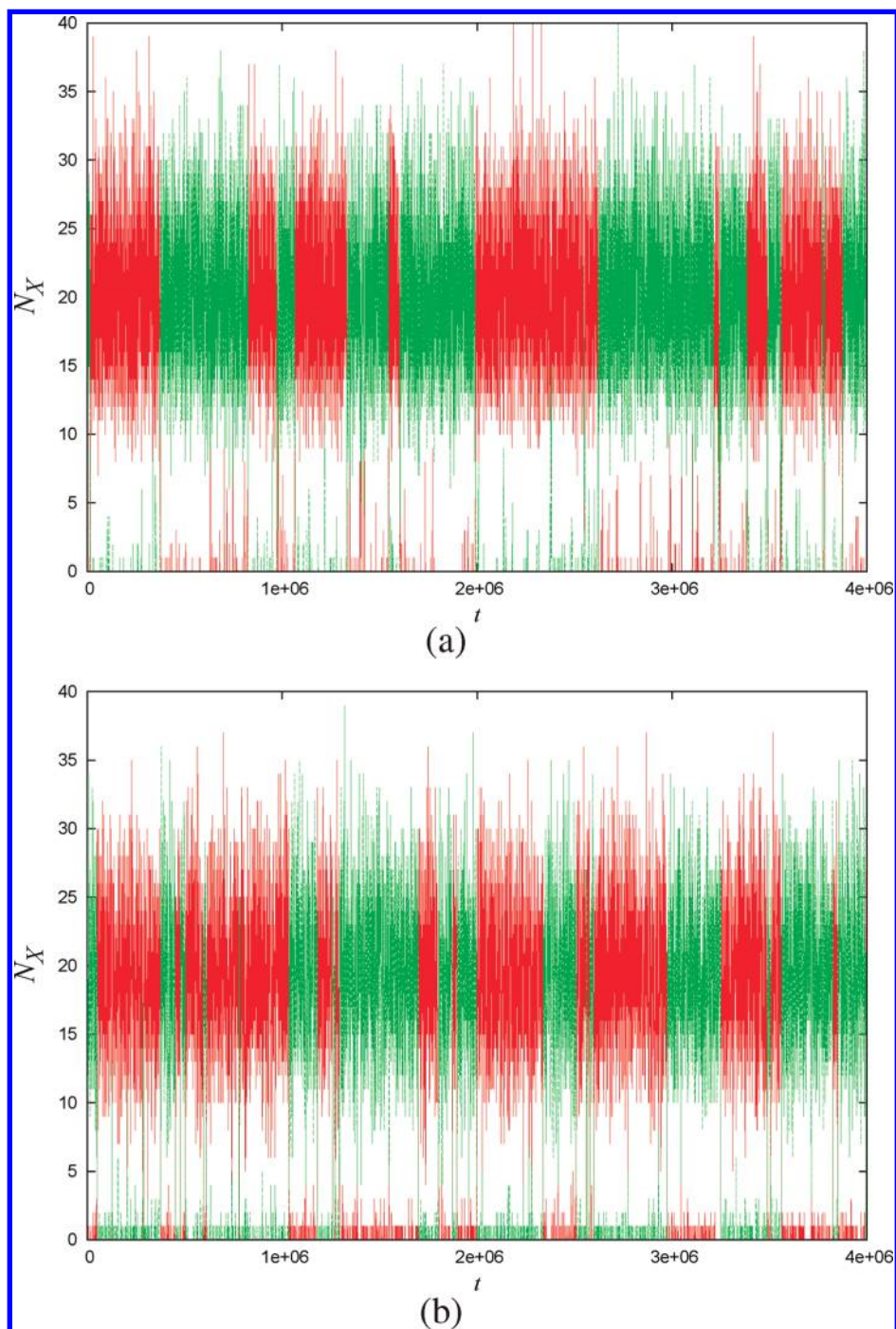


Figure 6. Gillespie and MaxCal time traces for a different parameter regime. (a) Gillespie time trace. Gillespie model, using $d = 0.005$, $k = 0.1$, $f = 100$, and $r = 2$. (b) MaxCal model time trace. Corresponding MaxCal model with the extracted parameter $K = -0.294$ using $h_\alpha = h_\beta = -4.605$ and $h_A = h_B = 7.6$.

Within our MaxCal model, we have assumed that correlations happen essentially instantaneously, without time delays. However, real systems will have time delays. The approximation made in our current MaxCal model qualitatively breaks down when binding–unbinding events from the gene–promoter complex are slow compared to typical birth–death rates. For example, the agreement of Figure 5b was worse than that of Figure 7. This is because there is a delay in the effect of species B on A, and vice versa, when binding–unbinding rates are not much faster than birth–death events.

We note here the importance of the dynamical fluctuations, relative to average properties. Suppose we take the average (mass-action) behavior of the same model of the ETS that we used in our Gillespie simulation above. The time evolution of the two species A and α would be given by

$$\langle \dot{A} \rangle = -d\langle A \rangle + k\langle \alpha \rangle \quad (12)$$

$$\langle \dot{\alpha} \rangle = r(1 - \langle \alpha \rangle) - f\langle B\alpha \rangle \quad (13)$$

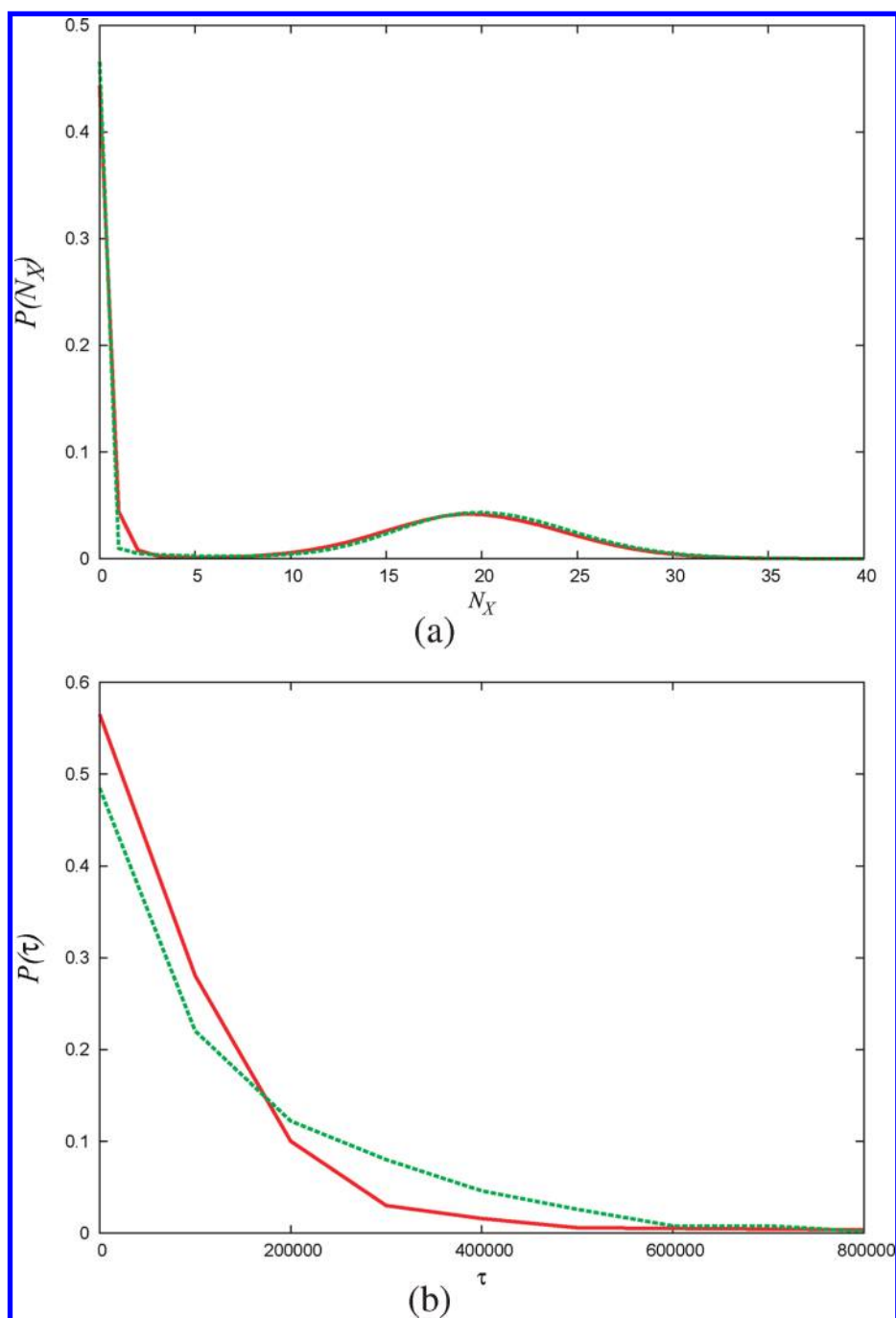


Figure 7. Prediction of dynamical and static distributions for a different parameter regime. (a) Distribution of particle numbers. (b) Distribution of dwell times. (Green) Gillespie simulations. (Red) MaxCal model. Parameters are those used in Figure 6.

We neglect here terms in eq 12 that are due to binding and unbinding of A in the ETS to β because we assume there are a large number of A's. In addition, we have assumed that there is a single α .

When binding and unbinding rates of A and B to the promoter (f and r , respectively) are faster than production and degradation of A and B, then we can assume that $\langle\alpha\rangle$ is in approximate equilibrium with $\langle\alpha^*\rangle$, so we have $\langle\dot{\alpha}\rangle = 0$. Combining eq 13 with eq 12 yields

$$\langle\dot{A}\rangle = k - d\langle A\rangle - \frac{fk}{r}\langle B\alpha\rangle \quad (14)$$

The key point is that the mass-action dynamics predicted from eq 14 shows no hopping between states.³⁹ The bistability in this system is a stochastic phenomenon that results from the small numbers of particles.

Our MaxCal approach bears some resemblance to ML methods. ML starts with knowledge of a reaction topology. ML expresses the probabilities of the various trajectories the system could take, given that reaction topology. ML then finds those parameters that maximize the probability of the experimentally observed path. MaxCal also enumerates all the possible trajectories of a system. However, MaxCal does not require prior

knowledge of the reaction topology. In addition, MaxCal chooses parameters to ensure that the average observables of the model are identical to those of the experiments and gives a minimal model consistent with this data.

CONCLUSIONS

We have described how to apply the method of MaxCal to model the time trajectories of bistable systems. As an illustration, we have considered the ETS,^{39,40} resembling the experimental system of Gardner et al.³⁵ To create the “raw experimental data” trajectories, we simulated a model ETS using the Gillespie algorithm. We then applied the method of MaxCal to see if it could properly extract the full dynamical and static behavior of the toggle switch.

MaxCal takes as input a few measured average quantities over the trajectories. It then computes a dynamical partition-function-like quantity. From that quantity, MaxCal then predicts other dynamical properties. We find that MaxCal produces a predictive underlying model for this stochastically driven bistable system. The advantages of the MaxCal approach are that (1) it is more directly data-driven than other approaches, (2) the model is built using a minimal set of parameters (in our case, we use less parameters than would have been used in the corresponding master equations for the toggle), (3) it is based on observable correlations between variables, rather than *ad hoc* functional forms, and (4) it captures the dynamical averages and fluctuations at both short and long time scales despite requiring only relatively short trajectories to predict quantities that would otherwise typically require much larger data sets. Also, as shown in Figure (2), we find that switching behavior of the toggle switch correlates with conditions under which the fluctuations become stronger.

ASSOCIATED CONTENT

S Supporting Information. The Supporting Information contains two sections. The first is a description of the method of MaxCal. In particular, it contains a detailed derivation of the dynamical partition function. The second section shows under which circumstances the master equation and the dynamical partition function are related. This material is available free of charge via the Internet at <http://pubs.acs.org>.

ACKNOWLEDGMENT

We appreciate the support of NIH Grant GM34993. S.P. acknowledges the FQRNT for its financial support, and K.G. acknowledges an FRF Grant from DU. We thank Adam Arkin for proposing the toggle switch as a test case, George Peterson, Julian Lee, and Hao Ge for proofreading our manuscript, and Attila Szabo and Rob Phillips for many interesting discussions.

REFERENCES

- (1) Wernet, M. F.; Mazzoni, E. O.; Çelik, A.; Duncan, D. M.; Duncan, I.; Desplan, C. *Nature* **2006**, *440*, 174–180.
- (2) Samoilov, M. S.; Price, G.; Arkin, A. P. *Sci. STKE* **2006**, *2006*, re17.
- (3) Smits, W. K.; Kuipers, O. P.; Veening, J. W. *Nat. Rev. Microbiol.* **2006**, *4*, 259–271.
- (4) Venning, J. W.; Smits, W. K.; Kuipers, O. P. *Ann. Rev. Microbiol.* **2008**, *62*, 193–210.

- (5) Huang, S.; Guo, Y.-P.; May, G.; Enver, T. *Dev. Biol.* **2007**, *305*, 695–713.
- (6) Maamar, H.; Raj, A.; Dubnau, D. *Science* **2007**, *317*, 526–529.
- (7) Paliwal, S.; Iglesias, P. A.; Campbell, K.; Hilioti, Z.; Groisman, A.; Levchenko, A. *Nature* **2007**, *446*, 46–51.
- (8) Veening, J. W.; Hamoen, L. W.; Kuipers, O. P. *Mol. Microbiol.* **2005**, *56*, 1481–1494.
- (9) Bagowski, C. P.; Ferrell, J. E., Jr. *Curr. Biol.* **2001**, *11*, 1176–1182.
- (10) Acar, M.; Mettetal, J. T.; van Oudenaarden, A. *Nat. Genet.* **2008**, *40*, 471–475.
- (11) Hooshangi, S.; Thiberge, S.; Weiss, R. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 3581–3586.
- (12) Huang, C. Y.; Ferrell, J. E., Jr. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 10078–10083.
- (13) van Kampen, N. G. *Stochastic Processes in Chemistry and Physics*; North Holland Publishing Company: Amsterdam, 1981.
- (14) Gillespie, D. T. *J. Phys. Chem.* **1977**, *81*, 2340–2361.
- (15) Arkin, A.; Ross, J. *J. Phys. Chem.* **1995**, *99*, 970–979.
- (16) Ross, J. *Acc. Chem. Res.* **2003**, *36*, 839–847.
- (17) Noé, F. *J. Chem. Phys.* **2008**, *128*, 244103.
- (18) Milescu, L. S.; Akk, G.; Sachs, F. *Biophys. J.* **2005**, *88*, 2494–2515.
- (19) McKinney, S. A.; Joo, C.; Ha, T. *Biophys. J.* **2006**, *91*, 1941–1951.
- (20) Witkoskie, J. B.; Cao, J. *J. Chem. Phys.* **2004**, *121*, 6361–6372.
- (21) Flomenbom, O.; Klafner, J.; Szabo, A. *Biophys. J.* **2005**, *88*, 3780–3783.
- (22) Gull, S. F. Bayesian Inductive Inference and Maximum Entropy. In *Maximum Entropy and Bayesian Methods in Science and Engineering*; Erickson, G. J., Smith, C. R., Eds.; Kluwer Academic Publishers: Boston, 1988; Vol. 1, pp. 53–74.
- (23) Kienker, P. *R. Soc. London, Ser. B* **1989**, *236*, 269–309.
- (24) Jaynes, E. T. Macroscopic Prediction. In *Complex Systems - Operational Approaches*; Haken, H., Ed.; Springer-Verlag: Berlin, 1985; p254.
- (25) Bretthorst, G. L.; Jaynes, E. T. *Probability Theory: The Logic of Science*; Cambridge University Press: Cambridge, U.K., 2003.
- (26) Jaynes, E. T. *Phys. Rev.* **1957**, *108*, 171–190.
- (27) Jaynes, E. T. *Annual Review of Physical Chemistry*; Rabinovitch, S., Ed.; Annual Reviews Ins.: Palo Alto, 1980; pp 579–601.
- (28) Ghosh, K.; Dill, K. A.; Inamdar, M. M.; Seitaridou, E.; Phillips, R. *Am. J. Phys.* **2006**, *74*, 123–133.
- (29) Seitaridou, E.; Inamdar, M.; Phillips, R.; Ghosh, K.; Dill, K. *J. Phys. Chem. B* **2007**, *111*, 2288–2292.
- (30) Wu, D.; Ghosh, K.; Inamdar, M.; Lee, H. J.; Fraser, S.; Dill, K.; Phillips, R. *Phys. Rev. Lett.* **2009**, *103*, 050603.
- (31) Pressé, S.; Ghosh, K.; Phillips, R.; Dill, K. *Phys. Rev. E* **2010**, *82*, 031905.
- (32) Stock, G.; Ghosh, K.; Dill, K. *J. Chem. Phys.* **2008**, *128*, 194102.
- (33) Otten, M.; Stock, G. *J. Chem. Phys.* **2010**, *133*, 034119.
- (34) Schneidman, E.; Berry, M. J., II; Segev, R.; Bialek, W. *Nature* **2006**, *440*, 1007–1012.
- (35) Gardner, T. S.; Cantor, C. R.; Collins, J. J. *Nature* **2000**, *403*, 339–342.
- (36) Tozaki, H.; Kobayashi, T. J.; Okano, H.; Yamamoto, R.; Aihara, K.; Kimura, H. *FEBS Lett.* **2008**, *582*, 1067–1072.
- (37) Atkinson, M. R.; Savageau, M. A.; Myers, J. T.; Ninfa, A. J. *Cell* **2003**, *113*, 597–607.
- (38) Kim, K.; Wang, J. *PLoS Comput. Biol.* **2007**, *3*, e60.
- (39) Lipshtat, A.; Loinger, A.; Balaban, N. Q.; Biham, O. *Phys. Rev. Lett.* **2006**, *96*, 188101.
- (40) Loinger, A.; Biham, O. *Phys. Rev. Lett.* **2009**, *103*, 068104.
- (41) In the experiments of Gardner et al.,³⁵ the *Escherichia coli* in which were injected the engineered plasmid containing both promoters replicated as the experiment was carried through. The set of chemical reactions above does not take this or other complications into account; rather the reactions put forth are a simple set of ingredients required to obtain bistable steady state as well as switching between such states.
- (42) The dwell times were obtained by determining levels of protein A and B at time intervals of length *T* in the time traces. If protein A had

the higher level, then a counter is set to 1; otherwise it is 0. A switch is indicated by a change in the counter in the next interval. For sufficiently large T , there is a range over which the dwell times are independent of T . If T is too large, short transitions are missed. If T is too small, fluctuations are picked up as switches between steady states. It is because of the difficulty in defining what is a true transition in the presence of the rare raggedy switches that we define a switch through the simple algorithm above. Since we are really interested in comparing the dwells in MaxCal and Gillespie traces, to avoid bias, we use the same algorithm throughout with T equal to 1000 Gillespie or MaxCal steps, a step being defined by a change in particle number of A or B. We also verified our distribution of dwell times in different ways, for example by averaging levels of A and B within the interval T and using this to determine whether our counter variable should be set to 0 or 1. Different specialized methods for computing dwell times for toggle switches are also available in the literature. See: Allen, R. J.; Warren, P. B.; ten Wolde, P. R. *Phys. Rev. Lett.* **2005**, *94*, 018104.