2-20-2018

Hill Functions for Stochastic Gene Regulatory Networks from Master Equations with Split Nodes and Time-Scale Separation

Ovidiu Lipan
University of Richmond, olipan@richmond.edu

Cameron Ferwerda

Recommended Citation

Lipan, Ovidiu, and Cameron Ferwerda. "Hill Functions for Stochastic Gene Regulatory Networks from Master Equations with Split Nodes and Time-Scale Separation." *Physical Review E* 97, no. 2 (February 20, 2018): 022413. https://doi.org/10.1103/PhysRevE.97.022413.

Hill functions for stochastic gene regulatory networks from master equations with split nodes and time-scale separation

Ovidiu Lipan*
Department of Physics, University of Richmond, 28 Westhampton Way, Richmond, Virginia 23173, USA

Cameron Ferwerda
Department of Mathematics, King's College London, Strand, London WC2R 2LS, United Kingdom

(Received 14 November 2017; revised manuscript received 25 January 2018; published 20 February 2018)

The deterministic Hill function depends only on the average values of molecule numbers. To account for the fluctuations in the molecule numbers, the argument of the Hill function needs to contain the means, the standard deviations, and the correlations. Here we present a method that allows for stochastic Hill functions to be constructed from the dynamical evolution of stochastic biocircuits with specific topologies. These stochastic Hill functions are presented in a closed analytical form so that they can be easily incorporated in models for large genetic regulatory networks. Using a repressive biocircuit as an example, we show by Monte Carlo simulations that the traditional deterministic Hill function inaccurately predicts time of repression by an order of two magnitudes. However, the stochastic Hill function was able to capture the fluctuations and thus accurately predicted the time of repression.

DOI: 10.1103/PhysRevE.97.022413

I. INTRODUCTION

What is the stochastic version of the deterministic Hill function? To search for an answer we start with the well-understood deterministic approach. In 1910 Hill was studying the nature of the laws that govern colloidal solutions [1]. He developed a theory for dissociation curves for hemoglobin in solutions of various salts. The function that Hill used to fit the experimental dissociation curves,

$$H(X) = \frac{aX^n}{K^n + X^n}, \tag{1}$$

has become widely used and, at present, is part of almost any theoretical model of a genetic regulatory network [2–9]. The theoretical models for these networks are bound to be phenomenological because many details of the regulation processes are missing. It is of note that Hill's objective was phenomenological too, as he explains in Ref. [1]: "my object was rather to see whether an equation of this type can satisfy all the observations, than to base any direct physical meaning on n and K." Although the physical meaning may be absent, the phenomenological interpretation of these parameters is simple, K and n represent a threshold of the process and the switching behavior around the threshold, respectively, [10–12]. The effectiveness of Hill functions in modeling regulatory networks stems from the fact that many biocircuits act like switches. Some of these switches are vital to extending lifetimes, fighting obesity, and curing Alzheimer's disease [13–15].

The derivations of the Hill function (1) employ diverse biocircuit topologies such as Michaelis-Menten, sequential binding of ligands to receptors, competition, DNA looping, etc. [5,8,10,11,16–21]. Depending on the topology used, the Hill function may appear differently than that of (1), but it remains a rational function of its argument. A noteworthy example is the rational function that describes repression, which tends to zero as its argument increases. In a typical theoretical model, the Hill function controls the rate of production. For example, molecule X drives the time evolution of another molecule Y through a Hill function $H(X)$, $dY/dt = H(X) +$ other terms. The other terms may contain Hill functions that describe the coupling of Y with other molecules.

Deterministic models do not explain all the phenomena exhibited by molecular interactions. Stochastic effects, reviewed in Ref. [22], play an important role in many gene regulatory networks in living cells. In what follows we present a method to derive stochastic Hill functions from the master equation that is commonly used to describe the intrinsic noise [22]. As we transition from a deterministic to a stochastic model, the Hill function needs to gain more arguments. The argument X in (1) only represents a mean value. A stochastic Hill function needs additional arguments such as the standard deviation. Without additional arguments, the stochastic fluctuations associated with the information that flows through a genetic regulatory network are ignored. If multiple molecules cooperate to regulate Y, then the arguments of the Hill function need to contain their means, variances, and covariances. Choosing an arbitrary positive rational function of these arguments does not lead to a valid stochastic Hill function. The Hill function needs to express an input-output relation of a stochastic biocircuit. Thus, the goal for this study is to create stochastic biocircuits with such topologies that allow for the derivation of stochastic rational Hill functions. Additionally, another necessary goal is to obtain stochastic Hill functions in closed analytical forms. The analytical forms help to reduce the computational expense for large regulatory

*olipan@irichmond.edu

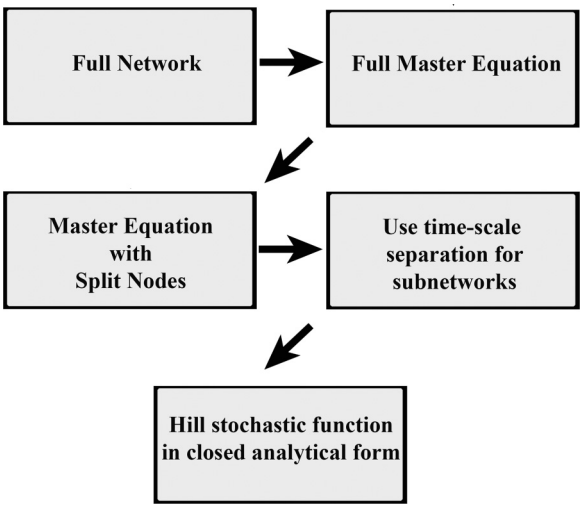

FIG. 1. Flowchart for obtaining stochastic Hill functions.

networks. The main challenge of working with stochastic biocircuits is that the joint probability distribution is usually inaccessible in an analytical form. This becomes more apparent for large genetic regulatory networks that contain hundreds of molecular species. To overcome this challenge, we work with a system of equations derived from the master equation that models the stochastic regulatory processes. The equations solve directly for the factorial moments of the random variables and thus avoid the joint probability distribution. However, for a nonlinear master equation this system of equations is infinite and does not close at any order of factorial moments. Closing the system of equations at second order is another challenge, which is addressed through the loop-closing (LC) method from Ref. [23]. This method is easily scalable, allowing it to be used for either small or large genetic networks and, specifically for this study, generates stochastic Hill functions in a closed analytical form. The flowchart representing the main steps of the method is presented in Fig. 1.

The full network is composed of N molecule types described by stochastic time-dependent levels of molecular numbers, $q = (q_k)$, $k = 1,\ldots,N$. The molecular interactions are described by a set of transition probabilities per unit time $T_\epsilon(q,t)$. Molecules can jump from one state q to another $q + \epsilon = (q_k + \epsilon_k)$, where ϵ is a vector of dimension N where $\epsilon_k \in \{\pm 1,0\}$ representing the jumps that either increase, decrease, or do not change the molecule number q_k. The master equation

$$\frac{\partial P(q,t)}{\partial t} = \sum_\epsilon T_\epsilon(q-\epsilon,t)P(q-\epsilon,t) - P(q,t)\sum_\epsilon T_\epsilon(q,t) \tag{2}$$

expresses this time evolution of the network.

The notation for the first- and second-order factorial moments generated from $F(z,t) = \sum_{q_1=0,\ldots,q_N=0}^{\infty} z_1^{q_1}\ldots z_N^{q_N} P(q_1,\ldots,q_N,t)$, is $F_k = \partial_{z_k} F|_{z=1}$ and $F_{jk} = \partial_{z_j,z_k} F|_{z=1}$, where $z \equiv (z_1,\ldots,z_N)$. At this point we arrived at the full master equation, which constitutes the second step of Fig. 1. The next step is to linearize the master equation through splitting the complex formation nodes.

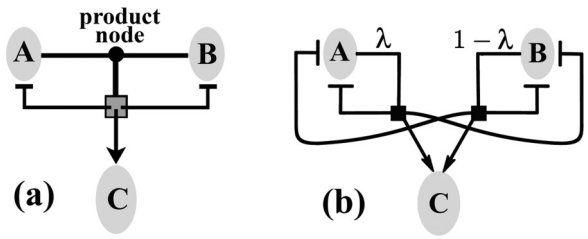

FIG. 2. Splitting the product node.

In short, the LC method focuses on the nonlinear complex formation process $A + B \to C$ where molecules A and B bind to form the complex C.

This irreversible complex formation runs with a transition probability per unit time $T(q,t) = aq_Aq_B$ represented in Fig. 2(a) by the product node, where q_A and q_B are the number of molecules of A and B, respectively. This process is expressed in the master equation by the following term:

$$\partial_t F = \left(z_A^{-1}z_B^{-1}z_C^{+1} - 1\right) a\, z_A z_B \partial_{z_A z_B} F. \tag{3}$$

The lack of linearity comes from the second derivative, $\partial_{z_A z_B} F$, which is derived from the nonlinear product of q_Aq_B. The LC method linearizes this term by splitting the product node, Fig. 2(b), as follows:

$$\begin{aligned}\partial_t F = {} & \lambda\left(z_A^{-1}z_B^{-1}z_C - 1\right)\frac{F_{AB}(t)}{F_A(t)} a\, z_A \partial_{z_A} F \\ & + (1-\lambda)\left(z_A^{-1}z_B^{-1}z_C - 1\right)\frac{F_{AB}(t)}{F_B(t)} a\, z_B \partial_{z_B} F.\end{aligned} \tag{4}$$

The splitting is akin to a convex combination of terms through the λ parameter, which is a positive number between zero and one. In addition, after splitting, the transition probabilities depend on the factorial moments, which are the unknowns to be solved for. Details on application of the LC method to different systems, including bistable systems, are presented in Ref. [23].

The third step in the flow chart of Fig. 1 uses two subnetworks. To control the value of the Hill coefficient, n, we will first construct a biocircuit with a cascade topology. The output of the cascade is a complex, C, which becomes the input of a second biocircuit, called the rational transfer function (RTF), Fig. 3. The factorial moments of C are polynomials of the factorial moments of input molecules 1 and 2. The degrees of these polynomials constitute the Hill coefficients. The threshold, K, is implemented by the second biocircuit, the RTF. The polynomials constructed by the cascade form the input of the RTF, which delivers the stochastic Hill rational function at its output. This output Hill function controls the dynamics of A through $dF_A/dt =$ Hill function $(F_1,F_2,F_{11},F_{22},F_{12})$ where the argument of the Hill function depends on the factorial moments of the input molecules 1 and 2.

It is relevant at this point to compare the approach we described above with the approaches of Refs. [24] and [25]. The authors of Ref. [24] clearly show that it is incorrect to describe the intrinsic noise by (1). Importantly, one cannot take the deterministic Hill function and use it to construct a microscopic transition probability by replacing X with the fluctuating number of molecules, q. The reason is that the

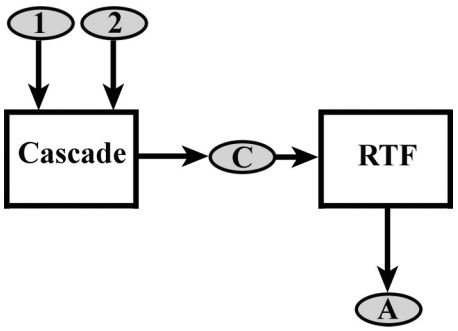

FIG. 3. The factorial moments of the complex, C, have a polynomial dependence on the input molecules' (1 and 2) factorial moments: $F_1(t)$, $F_2(t)$, $F_{11}(t)$, $F_{12}(t)$, and $F_{22}(t)$. The output, $F_A(t)$ and $F_{AA}(t)$, are rational functions of 1 and 2.

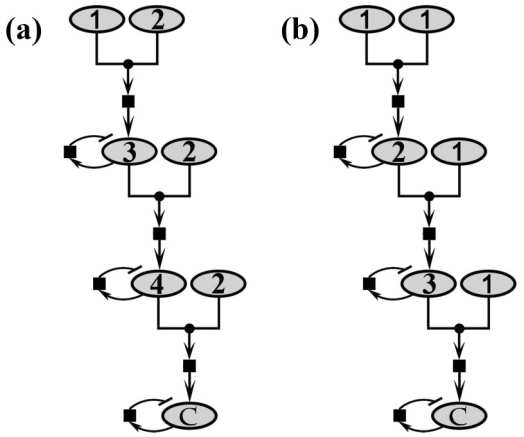

FIG. 4. (a) Molecules 1 and 2 combine to form the complex C, which is the output of the cascade. Each node represents an intermediate complex formation, denoted by 3 and 4, Ref. [23]. Consecutive additions of 2 increase the power of the polynomial that represent the state of C

transition probability is not derived from a master equation under conditions of time-scale separation, but rather it is a heuristic approach. The linearized master equation used by Ref. [24] is based on the system-size expansion of van Kampen [26] whereas we use the LC method, cf. Fig. 1 and Fig. 2 from Ref. [24].

The topologies of the cascade and RTF were inspired by ideas that are shared by many biocircuits from works such as Ref. [25]. One idea is that cooperativity in allosteric biocircuits gives rise to polynomials in the variable connected to the molecule that recurrently participates in molecular interactions. Thus, a cascade will have a polynomial output in the recurrent molecule. The polynomials easily become cumbersome and Ref. [25] makes an effort to reduce these polynomials to monomial powers by taking different limits in the parameter space. To quickly reveal the monomial structure, we will choose a simple topology for the cascade. The same goal of simplicity will be applied to the RTF. In general, the deterministic steady-state enzyme-substrate kinetics (ranging from Michaelis-Menten to competitive and noncompetitive inhibition as well as the kinetics of multireactant enzymes) are rational functions of molecule concentrations. We use a modified Michaelis-Menten to obtain a simple rational function.

The analytical form for a stochastic Hill function substantially reduces the computational cost for large genetic regulatory networks and so, in what follows, we proceed to solve the input-output relations first for the cascade and then for the RTF.

II. POLYNOMIAL TRANSFER FUNCTION FROM A STOCHASTIC CASCADE

Two molecules, 1 and 2, collide and start the cascade of Fig. 4(a) by forming the complex, 3. Then 3 collides with 2 to form 4 and the cascade continues by adding a new molecule 2 at every subsequent collision with the previously formed complex. The total number of collisions, n, determines the length of the cascade, the degree of the polynomial transfer function, and thus governs the Hill coefficient. For longer cascades, a more abrupt switchlike behavior is seen in the sigmoidal response of the Hill function. In Fig. 4(a), autodegradation processes on each of the resulting complexes ensures that the cascade reaches an equilibrium state. The equilibrium state of the cascade is then used to relate the state of C with the states of 1 and 2, which establishes the transfer function for the cascade. Other cascades with different transfer functions can be obtained by changing the network topology of the biocircuit. For example, the deterministic cascades of enzymatic reactions for a substrate molecule with n phosphorylation sites were studied in Ref. [27] where a reverse cascade is used to attain an equilibrium state.

The deterministic case is quite simple since the rate equations for the concentrations X are described by

$$\begin{aligned}\frac{dX_3}{dt} &= a_1 X_1 X_2 - b_1 X_3\\ \frac{dX_{i+2}}{dt} &= a_i X_{i+1} X_2 - b_i X_{i+2}\end{aligned} \tag{5}$$

where $i = 2, \ldots, n$. At steady state

$$X_{n+2} = \left(\prod_{j=1}^{n} \frac{a_j}{b_j}\right) X_1 X_2^n, \tag{6}$$

which is a simple monomial where the recurrent molecule appears at power n. From a stochastic point of view, the same cascade is described by a master equation where each complex formation node brings a second derivative as in (3). To ensure a closed system of equations at second-order moments, we split each complex formation node [23]. The resulting master equation contains a set of λ_i parameters,

$$\begin{aligned}\partial_t F = {} & a_1\lambda_1(z_3 - 1)\frac{F_{12}}{F_1} z_1 \partial_{z_1} F\\ & + a_1(1-\lambda_1)(z_3 - 1)\frac{F_{12}}{F_2} z_2 \partial_{z_2} F\\ & + b_1\left(z_3^{-1} - 1\right) z_3 \partial_{z_3} F\\ & + \sum_{i=2}^{n} a_i \lambda_i (z_{i+2} - 1)\frac{F_{2,i+1}}{F_{i+1}} z_{i+1}\partial_{z_{i+1}} F\end{aligned}$$

$$+ a_i(1-\lambda_i)(z_{i+2}-1)\frac{F_{2,i+1}}{F_2} z_2 \partial_{z_2} F$$
$$+ b_i\left(z_{i+2}^{-1}-1\right) z_{i+2}\partial_{z_{i+2}} F, \tag{7}$$

where the index $i = 1,\ldots,n$, counts the complex formation nodes. During all subsequent derivations, we will use the rapid steady-state assumption (RSSA), which assumes that all intermediary processes are much faster than the time evolution of the input molecules, 1 and 2. The assumption works as follows: take two small time intervals $(t,t+\Delta t)$ and $(t+\Delta t, t+2\Delta t)$ such that during each time interval the input functions of 1 and 2 can be considered constants. When the inputs, 1 and 2, change from the constant values at $(t,t+\Delta t)$ to another constant values at $(t+\Delta t, t+2\Delta t)$, the intermediate molecules inside the cascade, 3,4, ..., adjust very quickly around the time $t+\Delta t$ to another equilibrium state. In this way, the equilibrium value of the output, C, changes from one time interval to the other reflecting the changes in molecules 1 and 2. To implement the RSSA we take $a_i, b_i \to \infty$ so that a_i/b_i tends to a constant Γ_i. Under these conditions, the RSSA equilibrium state is characterized by $dF_i(t)/dt = 0, dF_{i,j}(t)/dt = 0$ with $i,j \neq 1,2$. The transfer function of the cascade is obtained via elimination of all intermediary factorial moments to find F_C and F_{CC} in terms of the input $F_1, F_{11}, F_2, F_{22}, F_{12}$.

For the cascade of Fig. 4(a), the transfer function for the mean, F_C, is given by (8). It factorizes into n factors, the first being F_{12}. The presence of F_{12} is not surprising since the start of the cascade is the collision between 1 and 2, which drives the formation of the complex, 3, through the average value of the product $q_1 q_2$. The other $n-1$ factors are similar to each other, except that each factor accumulates an additional λ as the complexes are formed down the cascade. For the deterministic case, F_{12} becomes $F_1 F_2$ and $F_{22} = F_2(F_2-1)$ and so we arrive at $F_{n+2} = (\prod_{j=1}^{n}\Gamma_j) F_1 F_2^n$ for (8). This result shows that the length of the cascade is the Hill coefficient, which is generated by the molecule that repeatedly binds to the intermediary complexes.

The second-order moment of the output, F_{CC}, appears as a more complex product of factors (9). It can be factored into n simple factors and a polynomial of order $3n$, $\mathcal{P}_{3n}$, which depends on all input variables and parameters Γ_j. Although the polynomial is not elementary, it is easily handled by any symbolic computation program.

The polynomial, $\mathcal{P}_{3n}$, from (9) offers the possibility of constructing Hill functions that depend on many parameters. This may be useful for experimental data that contains enough information to allow for the estimation of many parameters. However, most of the time the aim is to use Hill functions with as few parameters as possible. This can be achieved if the cascade has a localized bottleneck at the point $n-1$ in the cascade. Namely, set the $(n+1)$th molecule's autodegradation to be the slowest by setting $b_{n-1} = \eta b_{n-2}$ and then take the limit $\eta \to 0$. The bottleneck factorizes the polynomial, $\mathcal{P}_{3n}$, which causes the cascade output, F_{CC}, to take the form seen in (10) with F_C^2 and an additional factor. This factor is similar to the factors of (9), but also depends on the input $F_{11}+F_1$. The overall power of F_2 in F_{CC} is n^2, which is consistent for a second-order moment. For the case of $n=1$, F_{CC} is different than (9) and is instead $\mathcal{P}_3 F_{12}/(F_1 F_2)$.

$$F_C = \Gamma_n F_{12}\left[\prod_{j=1}^{n-1}\Gamma_j\left((1-\Lambda_j)\frac{F_2+F_{22}}{F_2}+\Lambda_j\frac{F_{12}}{F_1}\right)\right] \tag{8}$$

$$F_{CC} = \frac{F_{12}}{F_1^n F_2^n}\left[\prod_{j=1}^{n-2}\left((1-\Lambda_j)\frac{F_2+F_{22}}{F_2}+\Lambda_j\frac{F_{12}}{F_1}\right)\right]\left((1-\Lambda_{n-1})\frac{F_2+F_{22}}{F_2}+\Lambda_{n-1}\frac{F_{12}}{F_1}\right)^2\mathcal{P}_{3n} \tag{9}$$

$$F_{CC} = (F_C)^2\left((1-\Lambda_n)^2\frac{F_2+F_{22}}{F_2^2}+\Lambda_n^2\frac{F_1+F_{11}}{F_1^2}+2\Lambda_n(1-\Lambda_n)\frac{F_{12}}{F_1F_2}\right) \tag{10}$$

$$\Lambda_i = \lambda_1\lambda_2\ldots\lambda_{i-1}\lambda_i, \quad a_i = \Gamma_i b_i.$$

The above cascade incorporates the input signals from two molecules, which may be a requirement for some applications. However, in many applications of the Hill function the goal is to have one input molecule control a downstream network. In that case molecules 1 and 2 need to be the same. This is the case of Fig. 4(b) for which the master equation is (11)

$$\partial_t F = \sum_{j=1}^{n} a_j\lambda_j(z_{j+1}-1)\frac{F_{1j}}{F_1}z_1\frac{\partial F}{\partial z_1}$$
$$+ a_j(1-\lambda_j)(z_{j+1}-1)\frac{F_{1j}}{F_j}z_j\frac{\partial F}{\partial z_j}$$
$$+ b_j\left(z_{j+1}^{-1}-1\right)z_{j+1}\frac{\partial F}{\partial z_{j+1}}. \tag{11}$$

The output, F_C and F_{CC}, which now depends only on F_1 and F_{11}, is given by (12) and (13), where $\mathcal{P}'_{3n} \neq \mathcal{P}_{3n}$. With the rate-determining bottleneck step placed in the same location as before, (13) simplifies to (14),

$$F_C = \left(\prod_{j=1}^{n}\Gamma_j\right)F_{11}\left(\frac{F_1+F_{11}}{F_1}\right)^{n-1}, \tag{12}$$

$$F_{CC} = \frac{F_{11}}{F_1^{2n}}\left(\frac{F_1+F_{11}}{F_1}\right)^n\mathcal{P}'_{3n}, \tag{13}$$

$$F_{CC} = (F_C)^2\frac{F_1+F_{11}}{F_1^2}. \tag{14}$$

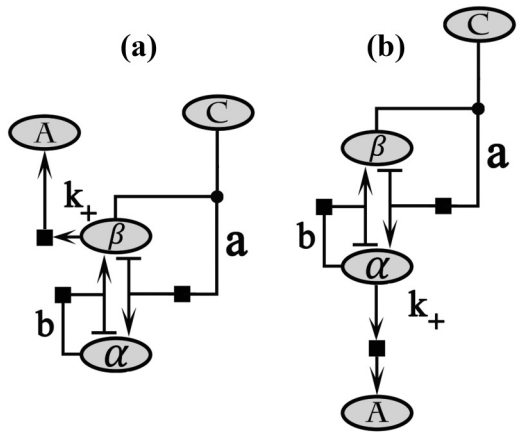

FIG. 5. Molecule C is given by the cascade. Its levels stay unchanged while it interacts with molecule β to produce molecule α. In (a), we see molecule β drives molecule A. Thus, C acts as a repressor for A. If we move A to be driven by α, then C works as an activator for A. Notice that creation of A is not accompanied by an annihilation of α or β, as is commonly seen in the Michaelis-Menten process. Adding an additional action line to include the missing annihilation would only create, in the mathematical formulas, sums of the form $k_+ + b$. Since $b \to \infty$ and k_+ is finite, the effect of an additional action line is irrelevant. Similar approaches are taken for the deterministic case in Refs. [11,28–30].

A dimer cascade of length n generates a Hill coefficient of $n+1$. This is because the dimer cascade starts with a complex formation from two identical molecules and so for $n=1$ the Hill coefficient is already 2.

III. RATIONAL TRANSFER FUNCTION

The RTF biocircuits from Figs. 5(a) and 5(b) have C as the input. The output of the RTF can be chosen to be either α or β depending on which Hill function we want to obtain, an activator or a repressor, respectively. Because we aim to find a closed analytical expression for the stochastic Hill function, we will employ the RSSA for the RTF just as we did for the cascade. This implies that around the time $t+\Delta t$, when the input 1 and 2 changes, the RTF's internal processes run fast and the outputs, β or α, quickly adjust to their RSSA equilibrium state. The net effect is that the outputs, α or β, will follow the slow time evolution of 1 and 2. In other words, the difference in the time scales for the fast internal processes versus the slow input processes allows the Hill functions be used with time-dependent arguments of 1 and 2. Before solving the master equation we present the traditional deterministic solution for Fig. 5(a). The deterministic time evolution of the concentrations are

$$\frac{dX_\alpha}{dt} = aX_\beta X_C - bX_\alpha \tag{15}$$

$$\frac{dX_\beta}{dt} = -aX_\beta X_C + bX_\alpha \tag{16}$$

$$\frac{dA}{dt} = k_+ X_\beta. \tag{17}$$

From here we find that the system possesses a conserved quantity $X_\alpha + X_\beta = Q$. In addition, for rapid equilibrium, $aX_\beta X_C = bX_\alpha$. From these two equations

$$X_\beta = \frac{Q}{1+\frac{a}{b}X_C}, \tag{18}$$

which represents a rational function in terms of X_C. From here $\frac{dA}{dt}$ is readily obtained in terms of X_C. This result shows that RTF needs to possess conserved quantities and complex formation processes.

Turning to the stochastic case, the master equation for the RTF biocircuit, after splitting the complex formation node, is (19).

$$\begin{aligned}\partial_t F = {} & \lambda a\left(z_\beta^{-1}z_\alpha - 1\right)\frac{F_{\beta C}}{F_\beta} z_\beta \frac{\partial F}{\partial z_\beta} \\ & + (1-\lambda)a\left(z_\beta^{-1}z_\alpha - 1\right)\frac{F_{\beta C}}{F_C} z_C \frac{\partial F}{\partial z_C} \\ & + b\left(z_\beta z_\alpha^{-1} - 1\right) z_\alpha \frac{\partial F}{\partial z_\alpha} + \text{driven molecule term.}\end{aligned} \tag{19}$$

The format of the driven molecule term from (19) depends on whether we choose to use the RTF as an activator or a repressor. For an activator, then molecule A is attached to α as seen in Fig. 5(b) and the term is $k_+(z_A-1)z_\alpha \frac{\partial F}{\partial z_\alpha}$. However, if we choose to use the RTF as a repressor, molecule β drives and the term becomes $k_+(z_A-1)z_\beta \frac{\partial F}{\partial z_\beta}$, Fig. 5(a). As before, both a and b tend towards infinity while their ratio tends towards a finite constant. The transition probability for the creation of A, k_+, is a finite constant and is thus a slower process.

The steady-state condition, $dF_\alpha(t)/dt = 0$, and the same conditions for $F_\beta, F_{\alpha\alpha}, F_{\alpha\beta}, F_{\beta\beta}$ are not enough to eliminate the intermediate variables and find the rational transfer function. The topology of the biocircuit needs to be chosen so that one or more conserved quantities can be extracted from it. The specific structure of the biocircuit from Figs. 5(a) and 5(b) enforces α and β to change their states simultaneously, but in opposite directions. This generates the basic conserved quantity $q_\alpha + q_\beta = Q$. Multiplying it with constants or with q_α, q_β and taking the average value generates more conserved quantities between factorial moments. The steady-state equations, together with the conserved quantities, supply the rational transfer function.

First we find the Hill function for the case of repression as presented in Fig. 5(a), where molecule A is driven by molecule β. From (19), we find that

$$\frac{dF_A}{dt} = k_+ F_\beta, \tag{20}$$

thus we seek F_β in terms of only the input, C. The steady-state condition, $dF_\alpha/dt = 0$, together with the average of the conserved quantity, $F_\beta + F_\alpha = Q$, gives

$$\frac{dF_A}{dt} = k_+\left(Q - \frac{a}{b}F_{\beta C}\right). \tag{21}$$

The dependence on the second-order moment, $F_{\beta C}$, is expected because of the complex formation processes between β and C. The steady-state equations give the following relation between the second-order moments:

$$F_{\alpha C} = \frac{aF_{\beta C}[\lambda F_C F_{\beta C} + (1-\lambda)F_\beta(F_C + F_{CC})]}{bF_\beta F_C}. \tag{22}$$

This relation, together with the average of the conserved quantity, $q_\alpha q_C + q_\beta q_C = Q q_C$, delivers an implicit expression for $F_{\beta C}$,

$$\lambda Q \frac{b + aF_C}{bQ - aF_{\beta C}} \left(F_{\beta C} - Q \frac{bF_C}{b + aF_C} \right) + (1 - \lambda)$$
$$\frac{(a+b)F_C + aF_{CC}}{bF_C} \left(F_{\beta C} - Q \frac{bF_C^2}{(a+b)F_C + aF_{CC}} \right) = 0. \tag{23}$$

Solving this second-order equation for $F_{\beta C}$ gives the rational transfer function of the RTF biocircuit of Fig. 5(a),

$$H_{\text{repressor RTF}} = Q \left(H_1 + \sqrt{H_1^2 + H_2} \right) \tag{24}$$

$$H_1 = \frac{1}{2} \left(1 - (1 - \lambda)^{-1} \frac{F_C^2 + \xi \lambda F_C}{(1 + \xi)F_C + F_{CC}} \right) \tag{25}$$

$$H_2 = (1 - \lambda)^{-1} \frac{\xi \lambda F_C}{(1 + \xi)F_C + F_{CC}}, \tag{26}$$

where the constant ξ is $\xi = b/a$ as $a, b \to \infty$

The λ dependence reveals that the RTF actually produces a family of transfer functions. This family interpolates between the $\lambda = 1$ and $\lambda = 0$ extremes. For $\lambda = 1$ (23) gives

$$F_{\beta C} = Q \frac{bF_C}{b + aF_C}, \tag{27}$$

$$\frac{dF_A}{dt} = k_+ Q \frac{\xi}{\xi + F_C}, \tag{28}$$

which is simply the deterministic result for a Hill function that represents repression. Taking the other limit, $\lambda = 0$, yields

$$F_{\beta C} = Q \frac{bF_C^2}{(a+b)F_C + aF_{CC}}, \tag{29}$$

$$\frac{dF_A}{dt} = k_+ Q \frac{\xi F_C + F_{CC} + F_C - F_C^2}{\xi F_C + F_C + F_{CC}}, \tag{30}$$

where the flow of stochastic information from the cascade has clearly been preserved through F_{CC}.

Interestingly, if we take the limit $b \to 0$ in (29) we get

$$F_A = k_+ Q \frac{F_{CC} + F_C - F_C^2}{F_C + F_{CC}} = k_+ Q \frac{\langle q_C^2 \rangle - \langle q_C \rangle^2}{\langle q_C^2 \rangle}, \tag{31}$$

which shows that molecule A is completely controlled by the standard deviation of molecule C. Finally, the Hill function is obtained by the composition of the polynomial transfer function of Sec. II and the rational transfer function from (24).

To better understand the transfer function for the RTF we study its dynamics as a separate biocircuit, not connected to the cascade. Instead of a cascade, the dynamical evolution of C is driven by two autoregulation loops, Fig. 6 (inset). We take the parameters $a = 1$, $b = 0.1$, $k_p = 9.9 \times 10^{-3}$, and $k_n = 1.0 \times 10^{-2}$ and the initial condition $q_C = 1$ at $t = 0$. The reason we chose the initial condition and parameters as such was to ensure that the RTF is placed in the region of small molecular numbers where fluctuations are high relative to the mean value. The RSSA is accommodated with a and b being much larger than k_p and k_n. The first autoregulation loop, a negative autoregulation, is stronger than the second, a positive autoregulation, causing the levels of C to deplete with time as these are the only actions on C.

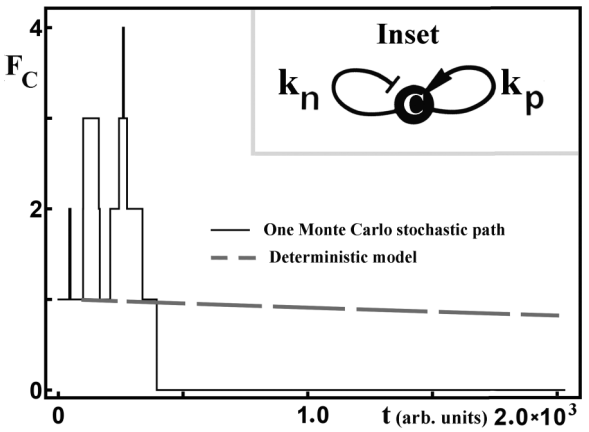

FIG. 6. The majority of the paths for molecule C terminate quickly when the molecule level falls to zero. The deterministic model does not capture this quick degradation. The negative and positive autoregulation loops on C are $k_n = 1.0 \times 10^{-2}$ and $k_p = 9.9 \times 10^{-3}$, respectively. The transition probabilities in the RTF are $a = 1.0$ and $b = 0.1$. Initial values are $q_C = 1$, $q_\alpha = 1$, and $q_\beta = 19$.

Looking at the case of the repressor, we expect the driven molecule to reach high levels when C is low. The LC-method stochastic formula, (24), predicts that this output will reach high levels in $t = 8.0 \times 10^2$. This prediction was confirmed by Monte Carlo simulations of the RTF, Fig. 7. However, the deterministic model incorrectly predicts the time of repression; it does not repress until a time two orders of magnitude later at $t = 8.0 \times 10^4$. The behavior of the deterministic model is due to the exponential nature of its prediction, whereas the stochastic simulation shows an abrupt decrease that takes place through jumps.

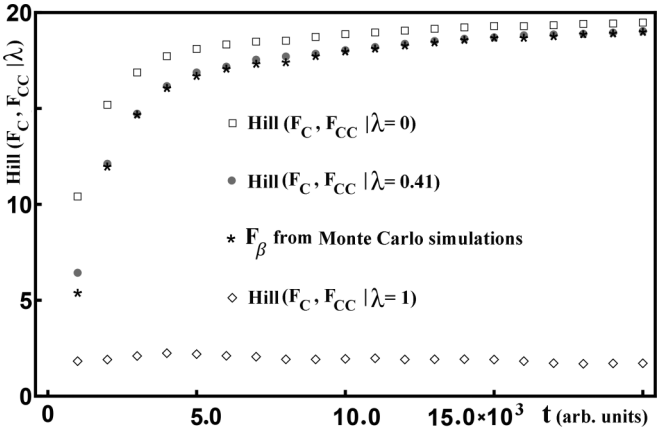

FIG. 7. The deterministic Hill function ($\lambda = 1$) does not match the Monte Carlo simulations, which are closer to the case of $\lambda = 0$. The minimum average distance between the LC-method stochastic Hill function and the Monte Carlo simulations corresponds to $\lambda = 0.41$.

The simulations show that the majority of molecule C paths terminate abruptly when the levels of C fall to zero, Fig. 6. Once this occurs, the RTF is driven only by molecule α and eventually ends when $\alpha = 0$ and $\beta = 20$. This final state where β reaches high values when C is small occurs two orders of magnitude more quickly in the Monte Carlo simulation than the deterministic model. Thus, the RTF repressor is described much better by the LC-method stochastic model.

So far we have concentrated exclusively on the repressor RTF. For the case of the activator, the molecule A is being driven by α as shown in Fig. 5(b). From the conserved quantity $F_\alpha + F_\beta = Q$ we get the RTF activator transfer function

$$H_{\text{activation RTF}} = Q\big(1 - H_1 - \sqrt{H_1^2 + H_2}\big), \tag{32}$$

where H_1 and H_2 are the same as in (25) and (26), respectively.

The deterministic case appears again for $\lambda = 1$,

$$\frac{dF_A}{dt} = k_+ \frac{F_C}{\xi + F_C}, \tag{33}$$

whereas for $\lambda = 0$, the Hill function is

$$\frac{dF_A}{dt} = k_+ Q \frac{F_C^2}{(1+\xi)F_C + F_{CC}}, \tag{34}$$

where the stochastic information is preserved through F_{CC}.

IV. APPLICATION TO THE INCOHERENT FEED-FORWARD LOOP

As an example, we examine an incoherent feed-forward loop (I1-FFL). In an I1-FFL, an activator transcription factor, X, regulates a gene, W, and a repressor of the gene, Y, Fig. 8. Extensive experimentation on I1-FFL has been done with *E. coli* and yeast [11,31–33]. The motif leads to distinct pulses that can affect cell fate decisions and can accelerate responses of genes in the network [11,31–33].

The biocircuit of Fig. 8 is represented mathematically by (35),

$$\begin{aligned}\partial_t F = {} & g(t)(z_X - 1)F + b_1\big(z_X^{-1} - 1\big)z_X \frac{\partial F}{\partial z_X} + H_a(F_X, F_{XX}, \xi_1)(z_Y - 1)z_X \frac{\partial F}{\partial z_X} + b_2\big(z_Y^{-1} - 1\big)\frac{\partial F}{\partial z_Y} \\ & + H_a(F_X, F_{XX}, \xi_2)H_r(F_Y, F_{YY}, \xi_3)(z_W - 1)\left(\lambda\, z_X \frac{F_{XY}}{F_X}\frac{\partial F}{\partial z_X} + (1-\lambda)z_Y \frac{F_{XY}}{F_Y}\frac{\partial F}{\partial z_Y}\right) + b_3\big(z_W^{-1} - 1\big)z_W \frac{\partial F}{\partial z_W},\end{aligned} \tag{35}$$

where $g(t)$ represents a pulse from an upstream network and H_a and H_r are stochastic Hill functions based on a dimer cascade with $n = 1$. As mentioned in Sec. II, a dimer RTF with $n = 1$ generates a Hill coefficient of 2 because the dimer cascade starts with a complex formation from two identical molecules. For this case,

$$H_a = \frac{F_C^2}{(1+\xi)F_C + F_{CC}}, \tag{36}$$

$$H_r = \frac{\xi F_C + \big(F_{CC} + F_C - F_C^2\big)}{(1+\xi)F_C + F_{CC}}, \tag{37}$$

$$F_C = \Gamma F_{VV}, \tag{38}$$

$$F_{CC} = \Gamma^2 \left(\frac{F_{VV}}{F_V}\right)^2 (F_V + F_{VV}), \tag{39}$$

where $V = X$ and $V = Y$ for H_a and H_r, respectively. For the coupling of X into Y the transition probability in (35) is $H_a(F_X, F_{XX}, \xi_1)$. Molecule W is driven by both X and Y and the transition probability is $H_a(F_X, F_{XX}, \xi_2)H_r(F_Y, F_{YY}, \xi_3)$.

We look to test the difference between the stochastic model of (35) and the same stochastic model for which the Hill functions are the traditional ones,

$$H_a = \frac{F_C}{\xi + F_C}, \tag{40}$$

$$H_r = \frac{\xi}{\xi + F_C}, \tag{41}$$

$$F_C = \Gamma F_V^2, \tag{42}$$

where $V = X$ and $V = Y$ for H_a and H_r, respectively.

In fact, the Hill functions represented by (40) and (41) are the deterministic functions that are commonly used. Instead of referring to them as deterministic we use the word "traditional" here to avoid any confusion because both models that we compare are stochastic models based on (35).

The entire biocircuit is activated by a rectangular pulse, $g(t)$, sent through molecule X as shown in Fig. 8. The unit of time for the biocircuit's dynamics is set by the degradation of molecule X, $b_1 = 1$ (arbitrary time units). The degradation of both Y and W is chosen to be 0.25. Concentration units are not used because the stochastic process delivers the probability

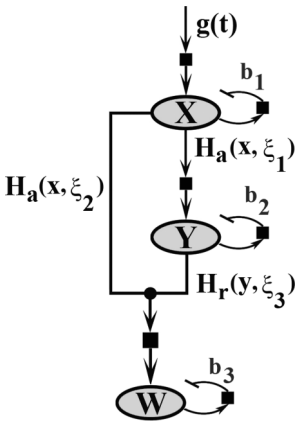

FIG. 8. The IFFL biocircuit. X activates W whereas Y represses it. The activation and repression are describe by Hill functions with three different thresholds, $\xi_i, i = 1,2,3$. $g(t)$ is a deterministic input signal. The degradation coefficients for X, Y, and W are b_1, b_2, and b_3, respectively.

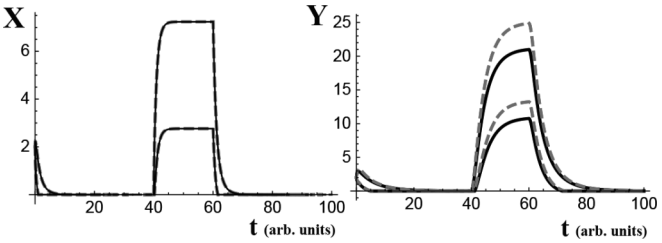

FIG. 9. The continuous line represents the dynamics produced by the stochastic Hill function, whereas the traditional Hill function case is plotted as a dotted line. The top lines and the bottom lines depict $F_V + \sigma_V$ and $F_V - \sigma_V$, respectively, with $V = X$ or $V = Y$.

for a number of molecules q_X,q_Y, and q_W to be present in the system at a give time. The initial, $t = 0$, probability distribution for X is $\delta(q_X - 2)$ where δ is the Kronecker-delta function. The same probability distributions were chosen for Y and Z. Similar to the Monte Carlo simulation from Sec. III, the biocircuit is in the region of low molecular numbers where stochastic fluctuation becomes relevant.

The rectangular pulse, $g(t)$, has a height of 5, a duration of 20, and a midpoint located at time $t = 50$. The response of the molecule X to this input pulse, Fig. 9 (X), is very close to a rectangular pulse. To represent the stochastic dynamics of X, the upper level curves in Fig. 9 (X) depict $F_X + \sigma_X$ whereas the lower level curves represent $F_X - \sigma_X$. The continuous lines were generated by the stochastic model, (35), with the stochastic Hill functions from (36) to (39). The dotted lines were generated by the same stochastic model (35), but with the traditional Hill functions from (40) to (42). For the dynamics of X, the continuous and the dotted lines match one another because the input rectangular pulse, $g(t)$, is independent on any Hill function. However, this is not the case for Y and W.

Molecule X sends its stochastic pulse to Y, Fig. 9 (Y), which is under the influence of a Hill function. This is where we start to see a difference in the stochastic and traditional Hill cases. The difference is striking for the output of the I1-FFL. The W pulse is quite distinct, depending on which Hill function is used, stochastic or traditional, Fig. 10. The traditional Hill functions delivers an output pulse, which is three times smaller than the stochastic Hill function. In addition, the bands, $F_W \pm \sigma_W$, do not even overlap for the numerical parameters in use.

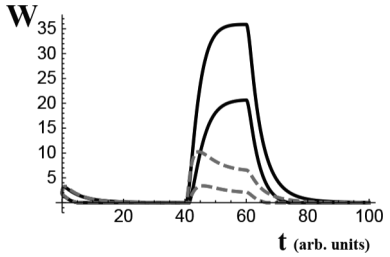

FIG. 10. The parameters for the stochastic Hill functions from (43) are $\xi_1 = 5$,$\xi_2 = 5$,$\xi_3 = 20$ for the RTF thresholds, $\Gamma = 5$ for the cascade strength, and $\lambda = 0.5$. The specific chosen numerical parameters are not crucial for the overall conclusion that there is a distinction between the use of stochastic versus traditional Hill functions.

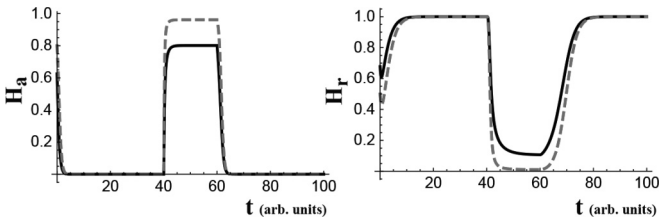

FIG. 11. The effect of $H_r(F_Y, F_{YY}, \xi_3)$ on W is more powerful than the effect on $H_a(F_X, F_{XX}, \xi_2)$. The output of I1-FFL is small for the traditional Hill function because $H_r(F_Y, F_{YY}, \xi_3)$ is almost zero during the duration of the pulse (dotted line).

This striking difference can be traced back to the repressor connection between Y and W. As both X and Y act on molecule W, Fig. 8, the term responsible for the production of W in dF_W/dt is

$$H_a(F_X, F_{XX}, \xi_2) H_r(F_Y, F_{YY}, \xi_3) F_{XY}. \tag{43}$$

The traditional Hill function repressor H_r has a strong effect on the response of W. To observe this effect, we notice that out of the three factors of (43), H_a and F_{XY} are larger for the traditional Hill function, Fig. 11 (H_a) and Fig. 12 (F_{XY}). Thus, they are not responsible for the effect. The traditional Hill repressor H_r, Fig. 11 (H_r) dotted line, hits a low level, which is close to zero. This heavily influences the production of W, Fig. 12 ($H_a H_r F_{XY}$), keeping its value low. The repressor part is what allows the stochastic Hill function to grow larger than the traditional Hill case.

V. CONCLUSION

We have shown that by using a cascade and a rational transfer function biocircuit, one can construct a stochastic Hill function with a Hill coefficient related to the number of collisions in the cascade. The Hill function is either a composition of the RTF repressor, (24), or the activator, (32), and either a two-input cascade, (8)–(10), or a dimer cascade (12)–(13). Since these stochastic Hill functions are in a closed analytical form, they can be used to model threshold events in stochastic genetic regulatory networks. Moreover, because we worked with a system of differential equations for the moments instead of the joint probability distribution of the network, the stochastic Hill function is useful for large genetic networks. The stochastic Hill functions are also time dependent since the input molecules 1 and 2 in Fig. 3 are allowed to vary in time. This was achieved through the LC method that was specifically designed to handle time-dependent stochastic

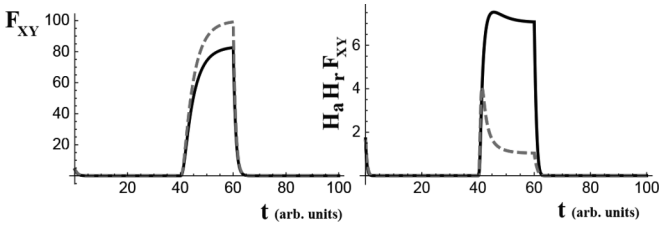

FIG. 12. F_{XY} together with H_a and H_r generates the production of W, (43).

processes. The approximation involves bimolecular reactions responsible for complex formation. There are few exact result that involve bimolecular reactions. One such result is presented in Ref. [34] where the number of molecules for the complex can be either 0 or 1. The master equation is elegantly solved at steady state and appears as a sum over Kummer functions. In the most general case the normalization of the probability distribution must be numerically computed. Reference [34] shows that the complex formation process is far from trivial and thus needs careful attention since it is ubiquitous in genetic regulatory networks. Through the LC method, we were able to obtain analytical solutions for the stochastic Hill functions in an as simple as possible format that can be applied for large genetic regulatory networks.

[1] A. V. Hill, J. Physiol. **40**, 4 (1910).
[2] J. L. Cherry, J. Theor. Biol. **203**, 117 (2000).
[3] Y. Shindo *et al.*, Nat. Commun. **7** 10485 (2016).
[4] H. Kim and E. Gelenbe, IEEE/ACM Trans. Comput. Biol. Bioinform. **9**, 973 (2012).
[5] T. S. Gardner, C. R. Cantor, and J. Collins, Nature (London) **403**, 339 (2000).
[6] Y. Xu, Y. Li, H. Zhang, X. Li, and J. Kurths, Sci. Rep. **6**, 31505 (2016).
[7] M. Malleshaiah, V. Shahrezaei, P. Swain, and S. Michnick, Nature (London) **465**, 101 (2010).
[8] H. T. Yang, C. P. Hsu, and M. J. Hwang, J. Biochem. **142**, 135 (2007).
[9] W. Fu *et al.*, Nat. Immunol. **13**, 972 (2012).
[10] M. Santillan, Math. Model. Nat. Phenom. **3**, 85 (2008).
[11] U. Alon, *An Introduction to Systems Biology: Design Principles of Biological Circuits* (CRC, Boca Raton, 2007).
[12] J. E. Ferrell, Jr., Trends Biomed. Sci. **23**, 461 (1998).
[13] C. Merkwirth, V. Jovaisaite, J. Auwerx, and A. Dillin, Cell **165**, 1209 (2016).
[14] J. A. Pospisilik, Cell **164**, 353 (2016).
[15] K. T. Tam, P. K. Chan, W. Zhang, P. P. Law, Z. Tian, G. C. Fung Chan, S. Philipsen, R. Festenstein, and K. C. Tan-Un, Nucleic Acids Res. **45**, 115 (2017).
[16] S. I. Rubinow, *Introduction to Mathematical Biology* (Dover Publications, Mineola, 2003).
[17] D. Gonze, W. Abou-Jaoude, D. Ouattara, and J. Halloy, in *Methods in Enzymology*, edited by M. Johnson and L. Brand (Elsevier, Amsterdam, 2011), Chap. 7, pp. 171–215.
[18] A. D. Keller, J. Theor. Biol. **172**, 169 (1995).
[19] F. M. Rossi, A. M. Kringstein, A. Spicher, O. M. Guicherit, and H. M. Blau, Mol. Cell **6**, 723 (2000).
[20] A. Narang, J. Theor. Biol. **247**, 695 (2007).
[21] M. B. Elowitz and S. Leibler, Nature (London) **403**, 335 (2000).
[22] D. Schnoerr, G. Sanguinetti, and R. Grima, J. Phys. A **50**, 093001 (2017).
[23] C. Ferwerda and O. Lipan, Phys. Rev. E **94**, 052404 (2016).
[24] P. Thomas, A. V. Straube, and R. Grima, BMC Syst. Biol. **6**, 39 (2012).
[25] I. Segel, *Enzyme Kinetics: Behavior and Analysis of Rapid Equilibrium and Steady State Enzyme Systems* (Wiley, New York, 1975).
[26] N. Kampen, *Stochastic Processes in Physics and Chemistry* (North Holland, Amsterdam, 2007).
[27] J. Gunawardena, Proc. Natl. Acad. Sci. **102**, 17 (2005).
[28] M. Johnson and L. Brand, *Methods in Enzymology: Computer Methods, Part C* (Elsevier, Amsterdam, 2011).
[29] B. Ingalls, *Mathematical Modeling in Systems Biology* (MIT Press, Cambridge, 2013).
[30] I. Segel, *Enzyme Kinetics: Behavior and Analysis of Rapid Equilibrium and Steady State Enzyme Systems* (Wiley, New York, 1993).
[31] S. Kaplan, A. Bren, E. Dekel, and U. Alon, Mol. Syst. Biol. **4**, 1 (2008).
[32] R. Entus, B. Aufderheide, and H. M. Sauro, Syst. Synth. Biol. **3**, 119 (2007).
[33] S. Mangan, S. Itzkovitz, A. Zalslaver, and U. Alon, J. Mol. Biol. **356**, 1073 (2006).
[34] R. Grima, D. Schmidt, and T. Newman, J. Chem. Phys. **137**, 035104 (2012).