Analytic framework for a stochastic binary biological switch

Guilherme C. P. Innocentini
Departamento de Matemática Aplicada, Universidade de São Paulo, São Paulo, Brazil

Sarah Guiziou and Jerome Bonnet
CBS, CNRS UMR 5048 - UM - INSERM U 1054, Montpellier, France

Ovidiu Radulescu
DIMNP, UMR CNRS 5235, University of Montpellier, Montpellier, France

(Received 15 June 2016; revised manuscript received 18 November 2016; published 28 December 2016)

We propose and solve analytically a stochastic model for the dynamics of a binary biological switch, defined as a DNA unit with two mutually exclusive configurations, each one triggering the expression of a different gene. Such a device has the potential to be used as a memory unit for biological computing systems designed to operate in noisy environments. We discuss a recent implementation of this switch in living cells, the recombinase addressable data (RAD) module. In order to understand the behavior of a RAD module we compute the exact time-dependent joint distribution of the two expressed genes starting in one state and evolving to another asymptotic state. We consider two operating regimes of the RAD module, a fast and a slow stochastic switching regime. The fast regime is aggregative and produces unimodal distributions, whereas the slow regime is separative and produces bimodal distributions. Both regimes can serve to prepare pure memory states when all cells are expressing the same gene. The slow regime can also separate mixed states by producing two subpopulations, each one expressing a different gene. Compared to the genetic toggle switch based on positive feedback, the RAD module ensures more rapid memory operations for the same quality of the separation between binary states. Our model provides a simplified phenomenological framework for studying RAD memory devices and our analytic solution can be further used to clarify theoretical concepts in biocomputation and for optimal design in synthetic biology.

DOI: 10.1103/PhysRevE.94.062413

I. INTRODUCTION

Biological machines are a promising new paradigm in computation [1]. By using synthetic biology it is now possible to design memory units and logic gates operating in living cells [2–4]. Contrarily to their silicon counterparts, biological computing units (BCUs) can evolve their computation capacities by proliferation and auto-organization. Furthermore, BCUs can be easily produced in large numbers and could be used for parallel distributed computing in living environments for medical applications such as implants for augmenting capacities or for health monitoring [5].

However, there are prices to pay when replacing the silicon substrate with biological cells. BCUs are submitted to stochastic fluctuations that are ubiquitous in the realm of molecular machines [6,7]. Within cell communities, the result of a computation can vary from one cell to another and therefore this result should be represented as a probability distribution of different states of the system. In order to optimize the design of BCUs, precise calculations of time-dependent population distributions are extremely useful. Here we present exact closed-form solutions for time-dependent probability distributions for a memory device. These solutions facilitate parameter analysis for optimal functioning of BCU devices. Moreover, they reduce drastically the time needed to fit models to data or to compare between various architectures of such devices.

Synthetic biological memory devices have been engineered based on different mechanisms. Some systems use a feedback loop in order to achieve data storage via bistability. For example, positive feedback systems inspired by the phage lambda Cro-CI circuit and using two transcription factors (TFs) mutually repressing each other have been implemented in the bacterium *Escherichia coli* [8,9]. Positive-feedback-based memory devices have also been engineered in the baker's yeast *Sacharomyces cerevisiae* [10]. The theory of genetic toggle switches with positive feedback has been developed elsewhere with emphasis on bimodality and on the stabilizing role of the cooperative binding of TFs to the DNA sequence and/or of the mutual exclusion of competing TFs [11–17]. It has been thus shown that bimodal distributions needed for separating memory states can occur even in the absence of cooperativity and bistability, although they still require positive feedback [16,17]. Recently, genetic memory systems using recombinases from the integrase family, enzymes that catalyze strand exchange between specific DNA sequences [5], have been implemented into living cells. DNA data storage has the advantage that only two discrete states are possible without the need for gene circuits and feedback. The state of the system can be interrogated by direct DNA sequencing but also by measuring the intensity of a fluorescent protein (FP) whose expression is controlled by the invertible DNA sequence (called the DNA register). Fluorescent reporters of two colors, for instance green fluorescent protein (GFP) and red fluorescent protein (RFP), can be used to reveal binary DNA register states of single cells. Although the majority of DNA data storage devices are single-write units, a particular architecture called the recombinase addressable data (RAD) module using two proteins, an integrase and a recombination directionality factor (or excisionase) enables

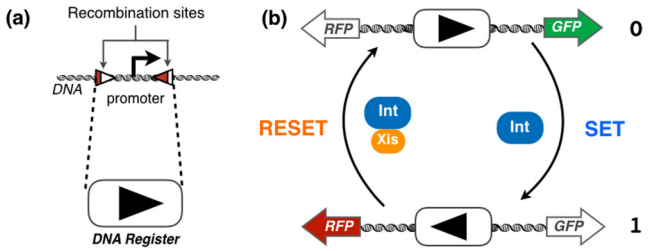

FIG. 1. Schematic architecture and operational principle of the recombinase addressable data (RAD) module. (a) The DNA register is composed of a DNA sequence containing a promoter (driving transcription) flanked by the two recombination sites recognized by the integrase (Int). (b) In state 0, the promoter drives expression of GFP. Integrase catalyzes the SET reaction, inverting the DNA register and enables transition towards state 1, in which the promoter has been inverted and now drives transcription of RFP. Concomitant expression of integrase and excisionase (Xis) catalyze the RESET reaction by enabling transition from state 1 to state 0.

rewritable digital data storage in bacteria [18] (Fig. 1). This mechanism generates bimodal probability distributions in a simpler way than the toggle switch because it does not include positive feedback. Here we introduce a stochastic model for a RAD unit and compute the exact time-dependent distribution probabilities for the two genes. Our model exhibits bimodality without positive feedback which confirms a result from similar models (Cook *et al.* [19], Hornos *et al.* [20], Innocentini and Hornos [21], Radulescu *et al.* [22], Shahrezai and Swain [23], Zeron and Santillan [24]). Although it has peculiarities and is not entirely equivalent to positive feedback toggle switches, this model has the advantage of simplicity and describes the behavior of an existing device.

The design of synthetic biological devices relies on the possibility to predict their performance for various parameter values. For stochastic devices this is usually done by Monte Carlo (Gillespie) simulations and scan of the parameter space, which is utterly time consuming. Therefore, analytic solutions of the time-dependent master equation are particularly useful for this type of application. Such analytic solutions were obtained in the past using the method of generating functions for a number of switch models: promoter on-off with direct protein expression [25], on-off promoter with direct protein expression and negative feedback [20,26], on-off promoter with two-stage expression (mRNA and protein) [21,23,27]. All these models describe the dynamics of a single protein, eventually coupled to its corresponding mRNA. The joint dynamics of two proteins expressed by an exclusive switch has never been studied analytically (the stationary probability distributions are nevertheless available for some toggle switches, as perturbative series [17] or as solutions of Fokker-Planck approximations to the master equation [16]) and can not be derived from the results of previous single-protein models.

II. DESCRIPTION OF THE MODEL

The RAD module synthesized by Bonnet *et al.* [18] is composed of a DNA sequence, which can exist in two configurations inverted one with respect to the other. The transition between the configurations is reversible. Depending on the configuration, single cells produce one of the two reporters, a GFP or RFP (Fig. 1). Although the two internal states of the memory defined by the DNA configurations are mutually exclusive, the expression of the two reporter genes depends on how fast the unit switches between the two configurations. As we will show later in this paper, fast switching between the two configurations allows that both GFP and RFP are expressed by the same cell. In this work, we are interested in the conditions enabling the generation of cell populations residing in a pure state (when all the cells express the same reporter gene) or separated mixed state (when subpopulations express an unique reporter gene but not necessarily the same). To this end, we introduce a mathematical model to describe the stochastic dynamics of the system. Our model is based on two coupled master equations governing the evolution of the joint probability distributions carrying the information about switch status and numbers of GFP (g) and RFP (r) at a given instant of time:

$$\begin{aligned}\frac{d\phi^0_{g,r}(t)}{dt} &= k_g\left[\phi^0_{g-1,r}(t)-\phi^0_{g,r}(t)\right]\\ &+\rho\left[(g+1)\phi^0_{g+1,r}(t)-g\phi^0_{g,r}(t)\right],\\ &+\rho\left[(r+1)\phi^0_{g,r+1}(t)-r\phi^0_{g,r}(t)\right]\\ &-h\phi^0_{g,r}(t)+f\phi^1_{g,r}(t) \qquad (1)\end{aligned}$$

$$\begin{aligned}\frac{d\phi^1_{g,r}(t)}{dt} &= k_r\left[\phi^1_{g,r-1}(t)-\phi^1_{g,r}(t)\right]\\ &+\rho\left[(r+1)\phi^1_{g,r+1}(t)-r\phi^1_{g,r}(t)\right].\\ &+\rho\left[(g+1)\phi^1_{g+1,r}(t)-g\phi^1_{g,r}(t)\right]\\ &+h\phi^0_{g,r}(t)-f\phi^1_{g,r}(t). \qquad (2)\end{aligned}$$

The random variables are g and r. Production of GFP and RFP is controlled by the rates k_g and k_r, respectively. The degradation and dilution rate of both reporters is given by ρ and the switching between the two states is encoded in the rates h (SET) and f (RESET).

As discussed in the introduction, the RAD module switching properties are not based on feedback. More precisely, the SET (h) and RESET (f) rates do not depend on the random variables g and r. In spite of the ostensible simplicity of the model, no analytic solutions for the joint stochastic dynamics of the proteins g and r are available to date. Such solutions are derived in the next section.

III. ANALYTIC SOLUTIONS FOR THE JOINT GENE EXPRESSION DYNAMICS

Introducing the generating functions, $\phi^0(z,y,t)=\sum_{g,r=0}^{\infty}\phi^0_{g,r}(t)z^g y^r$ and $\phi^1(z,y,t)=\sum_{g,r=0}^{\infty}\phi^1_{g,r}(t)z^g y^r$, we transform the master equations in a set of partial differential equations:

$$\begin{aligned}\frac{\partial\phi^0}{\partial t} &= (z-1)\left[k_g\phi^0-\rho\frac{\partial\phi^0}{\partial z}\right]\\ &-(y-1)\rho\frac{\partial\phi^0}{\partial y}-h\phi^0+f\phi^1,\end{aligned}$$

$$\frac{\partial\phi^1}{\partial t} = (y-1)\left[k_r\phi^1 - \rho\frac{\partial\phi^1}{\partial y}\right] - (z-1)\rho\frac{\partial\phi^1}{\partial z} + h\phi^0 - f\phi^1. \quad (3)$$

The desired joint probability distributions are obtained from the generating functions by applying the formulas:

$$\phi^0_{g,r}(t) = \frac{1}{r!}\frac{1}{g!}\frac{\partial^r}{\partial y^r}\frac{\partial^g}{\partial z^g}\phi^0(z,y,t)\Big|_{z=y=0},$$
$$\phi^1_{g,r}(t) = \frac{1}{r!}\frac{1}{g!}\frac{\partial^r}{\partial y^r}\frac{\partial^g}{\partial z^g}\phi^1(z,y,t)\Big|_{z=y=0}. \quad (4)$$

In order to solve the system of PDEs in Eq. (3) we propose a set of transformations, that will lead to an integrable system of ODEs. To do so, we perform a first change of variables:

$$w = \frac{(z-1)}{(y-1)}, \qquad x = (z-1)(y-1). \quad (5)$$

In the new set of variables (w,x) the equations read:

$$\frac{\partial\phi^0}{\partial t} + 2x\rho\frac{\partial\phi^0}{\partial x} = \sqrt{wx}k_g\phi^0 - h\phi^0 + f\phi^1$$
$$\frac{\partial\phi^1}{\partial t} + 2x\rho\frac{\partial\phi^1}{\partial x} = \sqrt{\frac{x}{w}}k_r\phi^1 + h\phi^0 - f\phi^1. \quad (6)$$

Note that this transformation eliminates one of the partial derivatives. To eliminate another partial derivative and obtain a set of ordinary differential equation (ODE) with respect to the remaining variable we propose a second transformation,

$$\nu = \sqrt{\frac{x}{w}} \qquad \mu = \sqrt{\frac{x}{w}}e^{-\rho t}, \quad (7)$$

which leads to the set of ODEs in the variable ν:

$$\nu\rho\frac{\partial\phi^0}{\partial\nu} = \nu w k_g\phi^0 - h\phi^0 + f\phi^1$$
$$\nu\rho\frac{\partial\phi^1}{\partial\nu} = \nu k_r\phi^1 + h\phi^0 - f\phi^1. \quad (8)$$

Now, solving the first equation of the system (8) for $\phi^1(\nu,\mu,w)$, gives

$$\phi^1(\nu,\mu,w) = \frac{1}{f}\left(\nu\rho\frac{\partial\phi^0}{\partial\nu} - \nu w k_g\phi^0 + h\phi^0\right), \quad (9)$$

and by substituting the result in the second equation of the system, we arrive to the second-order ODE with respect to the variable ν for $\phi^0(\nu,\mu,w)$:

$$\nu\rho^2\frac{\partial^2\phi^0}{\partial\nu^2} + \rho[f+h+\rho-\nu(k_r+k_g w)]\frac{\partial\phi^0}{\partial\nu} - [wk_g(\rho+f)+k_r h - \nu w k_g k_r]\phi^0 = 0. \quad (10)$$

This equation has a regular singularity at $\nu = 0$ and a irregular one at infinity. This structure suggest solutions in terms of confluent hypergeometric functions. To make it more clear, let us use the ansatz $\phi^0(\nu,\mu,w) = \exp(\nu w k_g/\rho)\psi(\nu,\mu,w)$, and a last transformation of variables: $\nu = \eta\rho/(k_r - wk_g)$. Putting this all together, we arrive to the equation in the new variable η for $\psi(\eta,\mu,w)$:

$$\eta\frac{\partial^2\psi}{\partial\eta^2} + (b-\eta)\frac{\partial\psi}{\partial\eta} - a\psi = 0, \quad (11)$$

where: $a = h/\rho, b = (h+f+\rho)/\rho$ and $\eta = \nu(k_r - wk_g)/\rho$.

Now, Eq. (11) is in the canonical form of the confluent hypergeometric equation, or Kummer equation, and the general solution is straightforward:

$$\psi(\eta,\mu,w) = F(\mu,w)M(a,b,\eta) + G(\mu,w)\eta^{1-b}M(1+a-b,2-b,\eta). \quad (12)$$

where M stands for Kummer function, $F(\mu,w)$ and $G(\mu,w)$ are arbitrary functions that will be determined by the initial conditions.

To obtain the expressions for the generating functions we have to multiply the solution in Eq. (12) by the exponential factor, $\exp(\nu w k_g/\rho)$, to obtain $\phi^0(\eta,\mu,w)$ using this result and Eq. (9) we can obtain the generating function for $\phi^1(\eta,\mu,w)$, in the variables (η,μ,w):

$$\phi^0(\eta,\mu,w) = e^{\left(\frac{k_g w\eta}{k_r - k_g w}\right)}[F\,M(a,b,\eta) + G\,\eta^{1-b}\,M(1+a-b,2-b,\eta)],$$
$$\phi^1(\eta,\mu,w) = e^{\left(\frac{k_g w\eta}{k_r - k_g w}\right)}\left[F\,\frac{h}{f}\,M(1+a,b,\eta) - G\eta^{1-b}M(2+a-b,2-b,\eta)\right]. \quad (13)$$

The last task remaining is to find $F(\mu,w)$ and $G(\mu,w)$. As said before, these functions are determined by the initial conditions. To accomplish that, one can see that setting $t = 0$ in the second transformation Eq. (7) leads to $\nu = \mu$, which in the variables (η,μ,w) implies in $\eta = \tilde{\eta} = \mu(k_r - wk_g)/\rho$. Specifying the functions $\phi^0(\eta = \tilde{\eta},\mu,w)$ and $\phi^1(\eta = \tilde{\eta},\mu,w)$ that will appear in the left-hand side of Eq. (13) one can find the expressions for $F(\mu,w)$ and $G(\mu,w)$. To do so, let us use vector and matrix notation to express the solutions in Eq. (13) as: $\vec{\phi} = U\vec{F}$, where, $\vec{\phi} = [\phi^0(\eta,\mu,w),\phi^1(\eta,\mu,w)]^{\mathrm{T}}$, $\vec{F} = [F(\mu,w),G(\mu,w)]^{\mathrm{T}}$ and the entries of the matrix U are given by:

$$U_{1,1} = \exp\left(\frac{k_g w\eta}{k_r - k_g w}\right)M(a,b,\eta),$$
$$U_{1,2} = \exp\left(\frac{k_g w\eta}{k_r - k_g w}\right)\eta^{1-b}M(1+a-b,2-b,\eta),$$
$$U_{2,1} = \exp\left(\frac{k_g w\eta}{k_r - k_g w}\right)\frac{h}{f}M(1+a,b,\eta),$$
$$U_{2,2} = -\exp\left(\frac{k_g w\eta}{k_r - k_g w}\right)\eta^{1-b}M(2+a-b,2-b,\eta). \quad (14)$$

Inverting $\vec{\phi} = U\vec{F}$ we obtain the expression $\vec{F} = U^{-1}\vec{\phi}$. Setting $\eta = \tilde{\eta}$ brings us to the position of determining the vector $\vec{F} = (F,G)^{\mathrm{T}}$ through the initial conditions $C_0 = \phi^0(\eta = \tilde{\eta},\mu,w)$ and $C_1 = \phi^1(\eta = \tilde{\eta},\mu,w)$. One final observation, before presenting F and G, concerns the determinant of the matrix U [$\det(U) = U_{1,1}U_{2,2} - U_{1,2}U_{2,1}$] necessary to compute U^{-1}. Inspection of (14) reveals that $\det(U)$ is a product of Kummer functions with an exponential envelope. Due to

the well-known properties of the Kummer functions [28] this determinant assumes a very simple formula:

$$\det(U) = \frac{(b-1)e^{\frac{\eta(k_r - wk_g)}{(k_r - wk_g)}}\eta^{1-b}}{1+a-b},$$

which is used to compute U^{-1}. The explicitly expressions for F and G are

$$\begin{aligned} F &= \exp\left(\frac{-k_r\tilde{\eta}}{k_r - wk_g}\right)\frac{f}{h+f}[M(2+a-b,2-b,\tilde{\eta})C_0 \\ &\quad + M(1+a-b,2-b,\tilde{\eta})C_1], \\ G &= \exp\left(\frac{-k_r\tilde{\eta}}{k_r - wk_g}\right)\tilde{\eta}^{b-1}\left[\frac{h}{h+f}M(a+1,b,\tilde{\eta})C_0\right. \\ &\quad \left. - \frac{f}{h+f}M(a,b,\tilde{\eta})C_1\right]. \end{aligned} \tag{15}$$

The initial conditions are encoded in $C_0 = \phi^0(\eta = \tilde{\eta},\mu,w)$ and $C_1 = \phi^1(\eta = \tilde{\eta},\mu,w)$.

IV. ESSENTIAL BIOLOGICAL PARAMETERS AND TYPES OF DYNAMICS OF THE RAD MODULE

At this point we are in position to exhibit the biological features of our model encoded in the time-dependent joint distributions of three variables: the two-valued DNA register state and the numbers of GFP (g) and RFP (r). However, before doing that, let us rephrase our parameter space in adimensional biological terms. The two numbers $N_g = k_g/\rho$ and $N_r = k_r/\rho$ are the protein production efficiencies. The asymptotic occupancy probabilities are $p_0 = f/(f+h)$ and $p_1 = h/(f+h)$. We call switching flexibility the important parameter $\epsilon = (h+f)/\rho$, representing the sum of the frequencies of the SET and RESET transitions. The two situations $\epsilon < 1$ and $\epsilon > 1$ correspond to slow and fast switching, respectively.

Having the time-dependent solutions at hand we will first illustrate the role of the biological parameters by analyzing their influence on the shape of the asymptotic distributions of the model. The generating function for the steady-state distributions is obtained by performing the limit $t \to \infty$ in Eqs. (13), resulting in:

$$\begin{aligned} \phi^0(\eta,w) &= p_0 \exp\left(-\frac{N_g w\eta}{N_g w - N_r}\right)M(a,b,\eta), \\ \phi^1(\eta,w) &= p_1 \exp\left(-\frac{N_g w\eta}{N_g w - N_r}\right)M(1+a,b,\eta), \end{aligned} \tag{16}$$

where we have: $a = p_1\epsilon$, $b = \epsilon + 1$, and $\eta = \nu(N_r - wN_g)$. Applying formulas (4) we obtain the steady-state joint probability distributions as

$$\begin{aligned} \phi^0_{g,r} &= \frac{f}{f+h}\frac{\Delta^{g+r}e^{-\Delta}}{r!g!}\sum_{s=0}^{g}\binom{g}{s}(-1)^s\frac{(a)_s}{(b)_s}\frac{(a+s)_r}{(b+s)_r}, \\ \phi^1_{g,r} &= \frac{h}{f+h}\frac{\Delta^{g+r}e^{-\Delta}}{r!g!}\sum_{s=0}^{g}\binom{g}{s}(-1)^s\frac{(a+1)_s}{(b)_s}\frac{(a+1+s)_r}{(b+s)_r}, \end{aligned} \tag{17}$$

where we have used $N_g = N_r = \Delta$ and the symbol $(\bullet)_s$ is the Pochhammer's symbol [28].

With Eqs. (17), we have computed the total joint probability distribution in the asymptotic regime of the system for different values of p_0, p_1 and ϵ keeping Δ constant. The steady-state distributions are represented in Figs. 2(a)–2(c). In order to expand the regions of small expression we made a change of variables in the distributions from g,r variables to $\ln(g)$, $\ln(r)$ variables. The analysis of the steady-state joint probability distribution also allows to distinguish between mixed and pure states of the DNA register. This concept is important for the relation between hidden memory states and visible (readable) phenotype. Mixed states are defined by $0 < p_0 < 1$ and $0 < p_1 < 1$, whereas a pure state means that either $p_0 = 1$ or $p_1 = 1$. Also, we define three types of subpopulations regarding the reporters readout: GFP only; RFP only and both GFP and RFP. The corresponding state occupancies are p_g, p_r, p_{rg}, respectively. A well-separated population has low values of p_{rg}. The state occupancies are not necessarily the same as the occupancy probabilities p_0 and p_1 encoding the probability to find the DNA register in state 0 or 1, respectively. A RAD unit in a pure state has a pure phenotype ($p_g = 1$ or $p_r = 1$), in other words all the bacteria express GFP in state 0 or RFP in state 1, as shown in Figs. 2(a) and 2(b). However, the

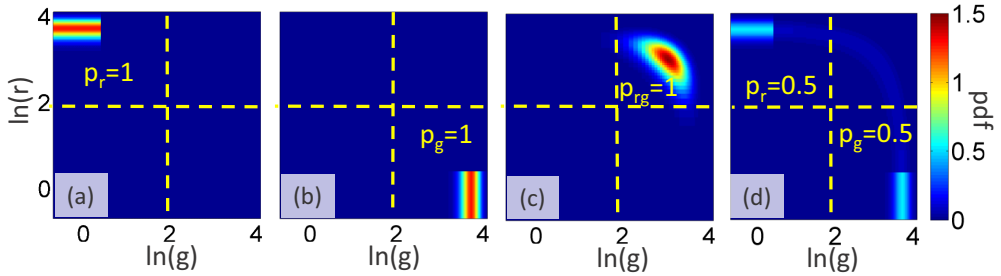

FIG. 2. Computed steady-state distributions of GFP and RFP for (a) pure $p_1 = 1$ red state (all cells express RFP) and $\epsilon = 0.5$ (slow switching); (b) pure $p_0 = 1$ green state (all cells express GFP) $\epsilon = 0.5$ (slow switching); (c) mixed $p_0 = p_1 = 0.5$ unimodal state (all cells express both GFP and RFP) and $\epsilon = 10$ (fast switching); (d) mixed $p_0 = p_1 = 0.5$ bimodal state (half of the cells express only GFP and half only RFP) and $\epsilon = 0.5$ (slow switching). In logarithmic variables, the double integral of the probability distribution is normalized to one. The remaining parameters are: $\Delta = 40$ and $\rho = 1$. The probabilities p_r, p_g, p_{rg} indicate the fractions of cells with $\ln(g) < 2, \ln(r) > 2$; $\ln(g) > 2, \ln(r) < 2$ and $\ln(g) > 2, \ln(r) > 2$, respectively, where the threshold 2 corresponds to the middle of the dynamical expression interval.

readout of a mixed state will depend on the switch flexibility. A fast switch ($\epsilon > 1$) will correspond to unimodal distribution of expressed genes each bacteria expressing both genes in different proportions [see Fig. 2(c)], whereas a slow switch ($\epsilon < 1$) corresponds to bimodal population where some cells are green and others are red [see Fig. 2(d)].

In order to illustrate the dynamical behavior of the RAD module we will first set the initial conditions (C_0 and C_1) of the system. To do so, we use the steady-state solutions as initial conditions but with different values for the occupancy probabilities and switch flexibility ($\tilde{p}_0, \tilde{p}_1, \tilde{\epsilon}$), leading to the initial generating functions:

$$\phi^0(\tilde{\eta}, w) = \tilde{p}_0 \exp\left(-\frac{N_g w \tilde{\eta}}{N_g w - N_r}\right) M(\tilde{a}, \tilde{b}, \tilde{\eta}),$$
$$\phi^1(\tilde{\eta}, w) = \tilde{p}_1 \exp\left(-\frac{N_g w \tilde{\eta}}{N_g w - N_r}\right) M(1+\tilde{a}, \tilde{b}, \tilde{\eta}), \quad (18)$$

At $t = 0$ we abruptly change the occupancy probabilities and switch flexibility from $(\tilde{p}_0, \tilde{p}_1, \tilde{\epsilon})$ to (p_0, p_1, ϵ), keeping them constant for $t \geqslant 0$. To study set and reset dynamics of the RAD unit we consider three distinct experiments and we show the corresponding time-dependent joint probability distributions for each one of these experiments in Fig. 3. The first two experiments correspond to the preparation of a pure state. We start with initial condition corresponding to a pure state $\tilde{p}_0 = 1$ where all cells are expressing only GFP (in the schematic description presented in Fig. 1 this means that a strong excisionase signal is applied together with the integrase), and at time $t = 0$ we change the asymptotic occupancy parameter to the complementary pure state, $p_1 = 1$, driving all the cells to express only RFP in the asymptotic configuration (excisionase is washed out). During the setting we use fast switching ($\epsilon = 10$) for the first experiment Fig. 3(a) and slow switching ($\epsilon = 0.5$) for the second one Fig. 3(b) (corresponding to high and low concentrations of integrase). In the third experiment, Fig. 3(c), we start with an unimodal mixed steady-state configuration, in the fast switch regime ($\epsilon = 10$, corresponding to high integrase and excisionase concentrations) and change the switch flexibility to low values ($\epsilon = 0.1$, lower concentrations). We call this last experiment "developing" because it transforms the initial unimodal mixed state in which single cells express both GFP and RFP (state occupancy $p_{rg} = 1$) into a bimodal, separated mixed state when two subpopulations have pure phenotypes, expressing either RFP or GFP ($p_r = p_g \approx 0.5$ in Fig. 2). More generally, the final state occupancies are given by the DNA configurations probabilities $p_r \approx \tilde{p}_1, p_g \approx \tilde{p}_0$). Therefore, lowering the switching frequencies by lowering the integrase and excisionase concentrations reveals previously hidden information about the DNA register probabilities. The last column of Fig. 3 shows the time dependence of the state occupancies [$p_g(t)$, $p_r(t)$ and $p_{rg}(t)$] corresponding to each one of the three experiments.

These experiments emphasize a clear distinction between slow and fast switches. Slow switches are separative, they

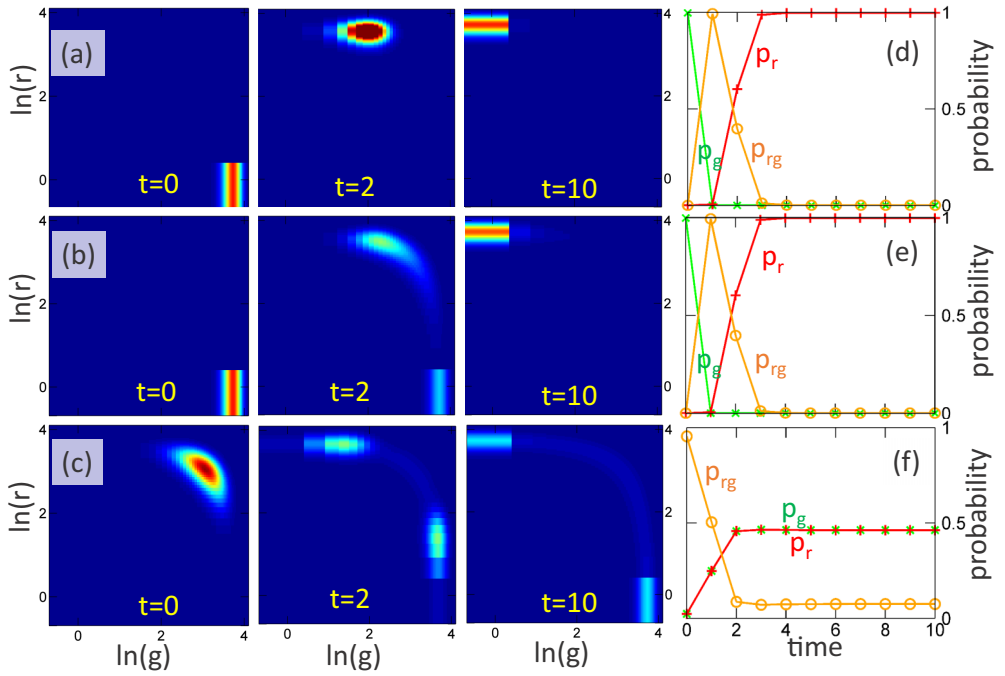

FIG. 3. Time-dependent distributions of a RAD unit (time is measured in units of ρ^{-1}). In (a), (b) we illustrate setting of a pure state. The $t = 0$ state corresponds to a pure distribution with $\tilde{p}_0 = 1$ (green) and switch flexibility $\tilde{\epsilon} = 0.5$. During set operation (for $t > 0$) the unit evolves towards a pure state $p_1 = 1$ (red) but with different switch parameters: $\epsilon = 10$ for (a) (fast switching) and $\epsilon = 0.5$ for (b) (slow switching). In (c) we illustrate separation of a unimodal mixed state. An unimodal mixed state with $\tilde{p}_0 = \tilde{p}_1 = 0.5$ and $\tilde{\epsilon} = 10$ (fast switching) is transformed to a bimodal mixed state by lowering the switch flexibility, $\epsilon = 0.1$ (slow switching). Also, in (c) one can see that for fast switching the phenotype is unimodal: bacteria express both GFP and RFP. The $t = 10$ phenotype, is bimodal: 50% of the cells express RFP and 50% express GFP. The remaining parameters are: $N_g = N_r = 40$ and $\rho = 1$. The rightmost columns (d)–(f) show the time dependence of the state occupancies.

tend to transform unimodal distributions into bimodal ones. They can be used as developers of mixed states, as occupancy probability can be read on bimodal distributions by counting cells in the two subpopulations, whereas it is much more difficult to estimate it from unimodal distributions. This can also be seen from the values of the state occupancies $p_{rg}(t)$ that remain low (indicating separated population) during the dynamics. For slow switches there is a tradeoff between good separation and bandwidth because the time needed to reach a final unimodal, pure state distributions is longer. Fast switches are aggregative, they tend to transform bimodal distributions into unimodal distributions.

V. COMPARISON WITH THE POSITIVE FEEDBACK GENETIC TOGGLE SWITCH

Genetic toggle switches based on positive feedback were extensively studied in the past in relation with their bimodal behavior [9,13–17]. In these models two genes repress each other and generate bimodality. It was shown that the propensity of a toggle switch for bimodality increases when the two genes have overlapping operator sites, which means that simultaneous transcription is not possible when the shared promoter is occupied [14]. An extreme version of mutual exclusion is also present in the RAD module where simultaneous transcription of the two genes is never possible. Contrary to the toggle switch, the gene products from the RAD module do not bind to the shared DNA promoter sequence and there is no positive feedback. The previous sections showed clearly that bimodality is possible in this case as well, which confirms previous results from similar models with no feedback or with negative feedback (Cook *et al.* [19], Hornos *et al.* [20], Innocentini and Hornos [21], Radulescu *et al.* [22], Shahrezai and Swain [23], Zeron and Santillan [24]). In this section we will test the capacity of the RAD module to produce the same quality of separation of memory states as a positive-feedback toggle switch. We will also compare the dynamics of a memory unit during a mixed state developing operation in the two cases. For the comparison we consider the toggle switch model with exclusion, mutual repression, but no cooperativity, discussed in Ref. [17]. This model describes the production of two types of proteins that, like in the RAD model, we denote by r and g. It has three discrete states 0, 1, and 2, corresponding to the free promoter, when a molecule g is bound to the promoter, or when a molecule r is bound to the promoter, respectively. The binding rates are proportional to the numbers of proteins and the binding constant b is the same for r and g. The unbinding rate is u. By binding, the promoter switches production state and starts producing the protein that is bound with even production rates $k_g = k_r = k$. The proteins are degraded with the constant d. The solutions of the time-dependent master equation are not available for this model. Furthermore, the first and second moments of the probability distribution do not satisfy closed form differential equations [17]. For these reasons, we decided to estimate distributions and moments by Monte Carlo simulations.

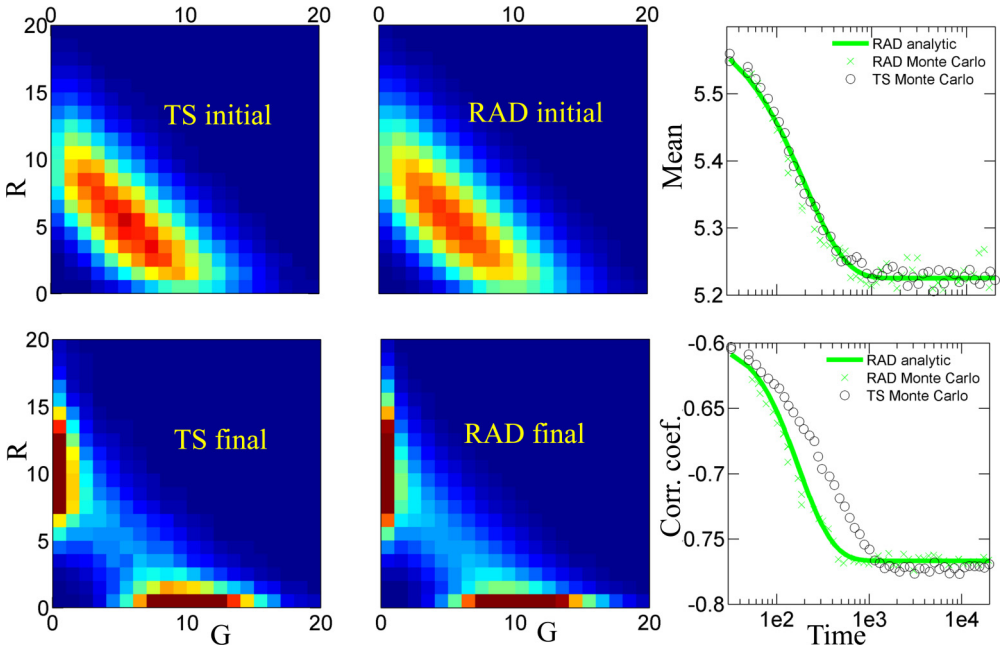

FIG. 4. The RAD model compared to the toggle switch model from Ref. [17] in the symmetric situation when the mean expression is the same for the two competing genes. Parameter fitting of the RAD model enforce that the initial and final bivariate probability distributions and the transient means are the same for the two models. The negative correlation between genes indicating the separation between memory states reaches more rapidly its asymptotic value for the RAD switch. The toggle switch time series were obtained by Monte Carlo simulations. Fitting of the RAD model used analytic expressions, but the final estimates were checked by comparison to Monte Carlo simulations. The parameters of the toggle switch model are $k = 0.05$, $d = 0.005$, $b = 0.1$, $u = 0.2$ for $t = 0$ and $u = 0.05$ for $t > 0$. The fitted parameters of the RAD module are $\rho = 0.005$, $f = h = 0.07$, $k_1 = k_2 = 0.057$ for $t = 0$ and $f = h = 0.0015$, $k_1 = k_2 = 0.053$ for $t > 0$, suggesting that a change of the unbinding parameter in the toggle switch model can be reproduced by a change of the switching rates in the RAD switch.

In a first stage, we simulated the toggle switch long enough to reach the stationary distribution with parameters $k = 0.05$, $d = 0.005$, $b = 0.1$, $u = 0.2$ (measured in s^{-1}). Then the parameter u was changed to $u = 0.05$ and the simulation was continued until the new stationary distribution was reached. These parameter values were used in Ref. [17] to reproduce the behavior of an *E. coli* toggle switch passing from a unimodal to a bimodal behavior. The second part of the simulation reproduces the memory set operation. We used this part to generate time series of the mean protein numbers (equal for r and g for this symmetric choice of production and degradation parameters), of the Pearson correlation coefficient and of the bivariate probability distributions of r and g.

Equation (13) was used to compute analytically the time-dependent first and second moments of the RAD model (see Appendix) and Eq. (17) was used to compute the initial and final stationary bivariate probability distributions. The time-dependent mean and the two stationary distributions were combined to define an objective function quantifying the difference between the time series generated by the simulation of the toggle switch and the RAD model predictions. The minimization of this objective function was used to fit the RAD model parameter to the time series data. Although it is possible to include the time-dependent distributions in the objective function, we found that this choice increases the execution time without improving the fit. Furthermore, as shown in Fig. 4 the transient correlation (resulting from second moments) behaves differently for the RAD model compared with the time series generated with the toggle switch.

The fitting procedure shows that the RAD device can produce the same initial and final distributions as the toggle switch. This implies that the quality of separation of the states in the final bimodal distribution is the same with the two types of switches. The time-dependent mean is the same for the two models. However, an important difference is emphasized by our numerical experiments. The negative correlation of the two genes, which is important for a memory device, evolves more rapidly for the RAD device compared to the toggle switch (Fig. 4). This means that on writing, the RAD device reaches faster than the toggle switch the same quality of the stored information. This argues in favor of the RAD module as a fast and accurate memory device.

VI. CONCLUSION

The analytic time-dependent solutions for the master equations describing the RAD memory unit can be used to quantitatively and qualitatively establish the basis for the design of such biological devices. All the necessary information about the system is encoded in the time-dependent joint probability distributions that are obtained by applying formulas (4) to the analytic solutions of the model. The solutions are expressed in closed form in terms of the well-known Kummer functions and symbolic computation software packages, such as MAPLE, can be used for direct computation of the series expansions leading to the desired joint distributions. Closed expressions for any desired moment of the joint distributions, such as mean values, variance and covariance can be obtained as shown in the Appendix.

Our analysis of the solution of the time-dependent master equation emphasizes several types of states and dynamics of the RAD module. Memory set operations involve preparation of pure states. This happens when one of the transitions, SET or the RESET, is dominating, such that $p_1 \approx 1$ or $p_0 \approx 1$, respectively. In this case, after a large enough time, all the bacteria express the same gene and only one. However, the RAD module has the potential for other type of application involving mixed states. In particular, the developing protocol generates a mixed state by using slow switching dynamics and finite values of the SET and RESET transition rates: $p_0, p_1 \neq 0, 1$. In the resulting population of bacteria, a proportion p_0 expresses only GFP and a proportion $p_1 = 1 - p_0$ expresses only RFP. By this method we obtain a different digital-analog way of storing information. The developing protocol of the RAD module can be used as a biological implementation of a p switch. A p switch [29] is a stochastic element that takes two values, one and zero, with probabilities p and $1 - p$, respectively. In other words, it is a physical implementation of a Bernoulli random variable, which is the basic element in stochastic computing [30–32]. This representation allows for a hybrid, digital-analog, form of computation in which real numbers are represented as random strings of zeros or ones where, the represented real number p is the frequency of occurrence of ones. Multiplication of two numbers, for instance, can be performed by combining random strings in parallel by AND operations, bit by bit [30,32]. In our case, each bit in a random string would be a cell. Although redundant from an information theoretic point of view, this representation was shown to be particularly robust and protected against random faults [30,32]. Theoretical and experimental developments of this application are ongoing and will be presented elsewhere.

Our analytic results apply to the RAD module illustrated in Fig. 1. However, we find that RAD module produces steady-state distributions very similar (practically indistinguishable) to positive feedback toggle switches. In particular, the RAD module is capable of bimodal behavior and good separation of memory states. Differences with respect to toggle switches occur at the level of time-dependent moments. For the same dynamics of the mean, the RAD module perform changes of the gene correlation more rapidly than the toggle switch, allowing more rapid writing. Other dynamical difference concerns the so-called switching time, which is the time that a single cell needs to pass from a state where one gene is highly expressed and the other has low expression to the symmetrically opposite situation. In the RAD model the switching time is given by the parameter ϵ and is relatively independent of the protein copy number (size). For toggle switches this time increases polynomially or exponentially with the size in the absence, or in the presence of cooperativity, respectively [16]. This property should be taken into account for the way memory is operated. Mixed states developing of the RAD module implies that switching time should be increased as much as possible in the long-term information storing state of the memory. This can be performed by gradually decreasing the concentrations of integrase and excisionase. For a toggle switch the same result can be obtained by increasing the affinities of the TFs to the DNA. Biochemistry imposes upper bounds to affinities, which could be a limitation especially for small-size systems.

ACKNOWLEDGMENTS

Work supported by FAPESP, SP, Brazil (G.I., Contract 2012/04723-4) and CNPq, Brazil (G.I., Contract 202238/2014-8). O.R. thanks LABEX Epigenmed for support. S.G. is supported by a PhD Fellowship from the French Ministry of Research. J.B. thanks the INSERM-CNRS (Grant No. RAK14002FSA) Atip-Avenir program (Grant No. R13132FS) and the Bettencourt-Schueller foundation.

APPENDIX: TIME-DEPENDENT MOMENTS

The moments of the distributions can be easily obtained through the analytical solutions for the generating function by using the formula

$$\langle g^p r^q\rangle(\tau) = \left[\left(y\frac{\partial}{\partial y}\right)^q\left(z\frac{\partial}{\partial z}\right)^p[\phi^0(z,y,\tau)+\phi^1(z,y,\tau)]\right]_{z=y=1}. \tag{A1}$$

Due to the well-known properties of the Kummer functions $[M(a,b,x)]$, [28]:

$$M(a,b,x=0)=1;$$
$$\frac{d^j}{dx^j}M(a,b,x)=\frac{(a)_j}{(b)_j}M(a+j,b+j,x); \tag{A2}$$

the expressions for the moments will assume simple and closed forms, as we shall see. So, using the formula in Eq. (A1) and the analytical solutions in Eq. (13) we are in position to obtain any desired moment of the distribution. The mean value for green protein is obtained by setting $p=1$ and $q=0$ in Eq. (A1) and for the mean value of red proteins we set $p=0$ and $q=1$, giving the general structure:

$$\langle g^1\rangle(\tau)=G_1^1e^{-\epsilon\tau}+G_2^1e^{-\tau}+G_3^1,$$
$$\langle r^1\rangle(\tau)=R_1^1e^{-\epsilon\tau}+R_2^1e^{-\tau}+R_3^1, \tag{A3}$$

where the G_j^1 coefficients are

$$G_1^1=\frac{\Delta(p_0-\tilde{p}_0)}{\epsilon-1},$$
$$G_2^1=\frac{\Delta(\tilde{p}_0-\epsilon p_0)}{\epsilon-1}+\tilde{\Delta}\tilde{p}_0, \tag{A4}$$
$$G_3^1=\Delta p_0.$$

The quantities $\tilde{\Delta}$, $\tilde{p}_0$ and $\tilde{\epsilon}$ (which will appear soon) represent the value of these parameters at $\tau=0$, meaning that once these parameters ($\tilde{\Delta}$, $\tilde{p}_0$, and $\tilde{\epsilon}$) are specified the initial configuration of the system is specified as well. All the R_j^1 are obtained by simply changing $p_0=p_1$ and $\tilde{p}_0=\tilde{p}_1$ in Eqs. (A4). The second moments for green and red proteins are obtained by setting $p=2,q=0$, and $p=0,q=2$ in Eq. (A1), respectively. The expressions for the second moments have the general structure:

$$\langle g^2\rangle(\tau)=G_1^2e^{-2\tau}+G_2^2e^{-(\epsilon+1)\tau}+G_3^2e^{-\epsilon\tau}+G_4^2e^{-\tau}+G_5^2,$$
$$\langle r^2\rangle(\tau)=R_1^2e^{-2\tau}+R_2^2e^{-(\epsilon+1)\tau}+R_3^2e^{-\epsilon\tau}+R_4^2e^{-\tau}+R_5^2, \tag{A5}$$

with coefficients G_j^2 given by:

$$G_1^2=\frac{\Delta^2(\epsilon p_0-2\tilde{p}_0)(\epsilon p_0-1)}{(\epsilon-2)(\epsilon-1)}-\frac{2\Delta\tilde{\Delta}[(\tilde{\epsilon}+1)\epsilon p_0-\tilde{\epsilon}\tilde{p}_0-1]}{(\tilde{\epsilon}+1)(\epsilon-1)}+\frac{\tilde{\Delta}^2\tilde{p}_0(\tilde{\epsilon}\tilde{p}_0+1)}{\tilde{\epsilon}+1},$$
$$G_2^2=\frac{2\Delta^2(p_0-1)[(p_0-\tilde{p}_0)\epsilon-\tilde{p}_0]}{(\epsilon-1)(\epsilon+1)}+\frac{2\Delta\tilde{\Delta}\tilde{p}_0[(p_0-\tilde{p}_0)\tilde{\epsilon}+p_0-1]}{(\tilde{\epsilon}+1)(\epsilon-1)},$$
$$G_3^2=\frac{2\Delta^2(\tilde{p}_0-p_0)[(p_0-1)\epsilon+1]}{(\epsilon-2)(\epsilon-1)}-\frac{\Delta(\tilde{p}_0-p_0)}{\epsilon-1},$$
$$G_4^2=\frac{2\Delta^2p_0(\tilde{p}_0-\epsilon p_0)}{\epsilon-1}+\frac{\Delta(\tilde{p}_0-\epsilon p_0)}{\epsilon-1}+\tilde{\Delta}\tilde{p}_0(2\Delta p_0+1),$$
$$G_5^2=\frac{\Delta^2p_0(\epsilon p_0+1)}{\epsilon+1}+\Delta p_0. \tag{A6}$$

Where, as before, the coefficients R_j^2 are obtained by doing $p_0=p_1$ and $\tilde{p}_0=\tilde{p}_1$ in Eqs. (A6).

Now, setting $p=1$ and $q=1$ in Eq. (A1) we obtain an expression for the covariance between the two random variables g and r, representing green and red proteins, respectively. The general structure of the covariance has the form:

$$\langle g^1r^1\rangle(\tau)=T_1^{1,1}e^{-2\tau}+T_2^{1,1}e^{-(\epsilon+1)\tau}+T_3^{1,1}e^{-\epsilon\tau}+T_4^{1,1}e^{-\tau}+T_5^{1,1}. \tag{A7}$$

The coefficients $T_j^{1,1}$ are given by:

$$T_1^{1,1}=-\frac{\Delta^2\epsilon\left[\epsilon p_1^2+(1-2\tilde{p}_1-\epsilon)p_1+\tilde{p}_1\right]}{(\epsilon-2)(\epsilon-1)}+\frac{\tilde{\Delta}^2\tilde{\epsilon}\tilde{p}_0\tilde{p}_1}{\tilde{\epsilon}+1}+\frac{\Delta\tilde{\Delta}[\epsilon(\tilde{\epsilon}+1)((2p_1-1)\tilde{p}_1-p1)+2\tilde{\epsilon}\tilde{p}_0\tilde{p}_1]}{(\tilde{\epsilon}+1)(\epsilon-1)},$$
$$T_2^{1,1}=-\frac{\Delta^2[(p_1-\tilde{p}_1)(2p_1-1)\epsilon+(1-2\tilde{p}_1)p_1+\tilde{p}_1]}{(\epsilon-1)(\epsilon+1)}-\frac{\Delta\tilde{\Delta}[(p_1-\tilde{p}_1)(2\tilde{p}_1-1)\tilde{\epsilon}+(2p_1-1)\tilde{p}_1-p_1]}{(\tilde{\epsilon}+1)(\epsilon-1)},$$
$$T_3^{1,1}=\frac{2\Delta^2(1-\epsilon p_0)(\tilde{p}_1-p_1)}{(\epsilon-2)(\epsilon-1)}-\frac{\Delta(\tilde{p}_1-p_1)}{\epsilon-1},$$
$$T_4^{1,1}=-\frac{\Delta^2(2\epsilon p_1p_0-2\tilde{p}_1p_1+\tilde{p}_1+p_1)}{\epsilon-1}-\Delta\tilde{\Delta}(2\tilde{p}_1p_1-\tilde{p}_1-p_1),$$
$$T_5^{1,1}=\frac{\Delta^2\epsilon p_1p_0}{\epsilon+1}+\Delta p_0. \tag{A8}$$

[1] S. Regot, J. Macia, N. Conde, K. Furukawa, J. Kjellén, T. Peeters, S. Hohmann, E. de Nadal, F. Posas, and R. Solé, Nature (London) **469**, 207 (2011).
[2] J. A. Brophy and C. A. Voigt, Nature Methods **11**, 508 (2014).
[3] J. Bonnet, P. Yin, M. E. Ortiz, P. Subsoontorn, and D. Endy, Science **340**, 599 (2013).
[4] A. A. K. Nielsen, B. S. Der, J. Shin, P. Vaidyanathan, V. Paralanov, E. A. Strychalski, D. Ross, D. Densmore, and C. A. Voigt, Science **352**, 53 (2016).
[5] Y. Benenson, Nature Rev. Genetics **13**, 455 (2012).
[6] C. Bustamante, W. Cheng, and Y. X. Mejia, Cell **144**, 480 (2011).
[7] T. B. Kepler and T. C. Elston, Biophys. J. **81**, 3116 (2001).
[8] Z. Toman, C. Dambly-Chaudière, L. Tenenbaum, and M. Radman, J. Mol. Biol. **186**, 97 (1985).
[9] T. Gardner, C. Cantor, and J. Collins, Nature (London) **403**, 339 (2000).
[10] C. M. Ajo-Franklin, D. A. Drubin, J. A. Eskin, E. P. Gee, D. Landgraf, I. Phillips, and P. A. Silver, Genes Dev. **21**, 2271 (2007).
[11] M. C. Mackey and M. Tyran-Kamińska, J. Math. Biol. **73**, 367 (2016).
[12] D. A. Potoyan and P. G. Wolynes, J. Chem. Phys. **143**, 195101 (2015).
[13] B. Barzel and O. Biham, Phys. Rev. E **78**, 041919 (2008).
[14] P. B. Warren and P. R. ten Wolde, Phys. Rev. Lett. **92**, 128101 (2004).
[15] P. B. Warren and P. R. ten Wolde, J. Phys. Chem. B **109**, 6812 (2005).
[16] T. Biancalani and M. Assaf, Phys. Rev. Lett. **115**, 208101 (2015).
[17] J. Venegas-Ortiz and M. R. Evans, J. Phys. A: Math. Theor. **44**, 355001 (2011).
[18] J. Bonnet, P. Subsoontorn, and D. Endy, Proc. Natl. Acad. Sci. USA **109**, 8884 (2012).
[19] D. L. Cook, A. N. Gerber, and S. J. Tapscott, Proc. Natl. Acad. Sci. USA **95**, 15641 (1998).
[20] J. E. M. Hornos, D. Schultz, G. C. P. Innocentini, J. Wang, A. M. Walczak, J. N. Onuchic, and P. G. Wolynes, Phys. Rev. E **72**, 051907 (2005).
[21] G. Innocentini and J. Hornos, J. Math. Biol. **55**, 413 (2007).
[22] O. Radulescu, A. Muller, and A. Crudu, Technique et Science Informatiques **26**, 443 (2007).
[23] V. Shahrezaei and P. S. Swain, Proc. Natl. Acad. Sci. USA **105**, 17256 (2008).
[24] E. Zeron and M. Santillán, Methods Enzymol. **487**, 147 (2010).
[25] J. Peccoud and B. Ycart, Theor. Population Biol. **48**, 222 (1995).
[26] A. F. Ramos, G. C. P. Innocentini, and J. E. M. Hornos, Phys. Rev. E **83**, 062902 (2011).
[27] G. Innocentini, M. Forger, O. Radulescu, and F. Antoneli, Bull. Math. Biol. **78**, 110 (2016).
[28] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables*, Vol. 55 (Courier Corporation, Mineola, 1964).
[29] D. Wilhelm and J. Bruck, in *2008 IEEE International Symposium on Information Theory* (IEEE, Piscataway, 2008), pp. 1388–1392.
[30] B. Gaines, in *Advances in Information Systems Science* (Springer, Berlin, 1969), pp. 37–172.
[31] W. Poppelbaum, C. Afuso, and J. Esch, in *Proceedings of the November 14-16, 1967, fall joint computer conference* (ACM, New York, 1967), pp. 635–644.
[32] A. Alaghi and J. P. Hayes, ACM Transactions on Embedded Computing Systems (TECS) **12**, 92 (2013).