Sampling Rare Switching Events in Biochemical Networks

Rosalind J. Allen,[1] Patrick B. Warren,[2,1] and Pieter Rein ten Wolde[1,*]

[1]*FOM Institute for Atomic and Molecular Physics, Kruislaan 407, 1098 SJ Amsterdam, The Netherlands*

[2]*Unilever R&D Port Sunlight, Bebington, Wirral CH63 3JW, United Kingdom*

(Received 26 September 2004; published 6 January 2005)

Bistable biochemical switches are widely found in gene regulatory networks and signal transduction pathways. Their switching dynamics are difficult to study, however, because switching events are rare, and the systems are out of equilibrium. We present a simulation method for predicting the rate and mechanism of the flipping of these switches. We apply it to a genetic switch and find that it is highly efficient. The path ensembles for the forward and reverse processes do not coincide. The method is widely applicable to rare events and nonequilibrium processes.

DOI: 10.1103/PhysRevLett.94.018104 PACS numbers: 87.16.Yc, 05.40.–a, 82.39.Rt, 87.16.Ac

Biochemical switches are ubiquitous in living cells. These switches are networks of chemical reactions with more than one steady state; flipping between these states can occur due to stochastic fluctuations. Examples include the lysis-lysogeny switch of phage λ [1] and the lac repressor of *E. coli* [2], as well as networks regulating the cell cycle and developmental fate [3]. Synthetic genetic switches have also been constructed *in vivo* [4,5].

Stochastic simulations [6,7] have an important role to play in revealing the design principles underlying the stability of these switches. However, these systems are often very difficult to simulate in a brute-force manner. This is because they can be extremely stable, showing few or no flips during the simulation. For example, the phage λ switch flips spontaneously less than once in 10^7 bacterial generations [8].

Biochemical networks are described by a set of chemical reactions with given rate constants. These determine the transition probabilities between different states. For many simulations in condensed matter physics, the transition probabilities satisfy detailed balance, leading to an equilibrium steady state with phase space density related to an energy functional. In contrast, biochemical networks usually lack detailed balance [9] and are out of equilibrium. For these systems, the phase space density is not known *a priori*, but is an output of the simulation. Commonly used techniques for the simulation of rare events [10], such as transition path sampling [11], are unsuitable because they rely on knowledge of the phase space density. In this Letter, we present the ''forward flux sampling'' (FFS) technique that allows efficient simulation of rare but important events in biochemical networks. We apply the method to a genetic toggle switch.

The FFS method generates the rare trajectories between two stable steady states A and B in a ratchetlike manner. To this end, it employs a series of nonintersecting interfaces in phase space between regions A and B. A and B are defined in terms of a parameter $\lambda(x)$ (x denotes all coordinates of the phase space), such that the system is in region A if $\lambda(x) \leq \lambda_A$ and in region B if $\lambda(x) \geq \lambda_B$. The parameter λ is also used to define the series of interfaces $\{\lambda_1, \ldots, \lambda_n\}$, such that $\lambda_1 \geq \lambda_A$ and $\lambda_n \leq \lambda_B$ (see Fig. 1) [12]. Defining the history-dependent functions h_A and h_B such that $h_A = 1$ and $h_B = 0$ if the system was more recently in A than in B, and vice versa, the rate constant k_{AB} for transitions from A to B is given by

$$k_{AB} = \frac{\overline{\Phi}_{A,B}}{\overline{h}_A} = \frac{\overline{\Phi}_{A,1}}{\overline{h}_A} P(\lambda_B|\lambda_1). \quad (1)$$

Here, $\Phi_{A,B}$ is the flux of trajectories that cross λ_B, coming from A (i.e., with $h_A = 1$). The overbar denotes a time average. $P(\lambda_{i+1}|\lambda_i)$ is the probability that a trajectory that comes from A and crosses λ_i for the first time will subsequently reach λ_{i+1} before returning to A. Thus the total flux from A to B is the flux from A to λ_1, multiplied by the probability that a trajectory reaching λ_1 from A will arrive in B without returning to A. $P(\lambda_B|\lambda_1)$ can be expressed as the product of the probabilities of reaching each successive interface from the previous one, without returning to A, such that $P(\lambda_B|\lambda_1) = \prod_{i=1}^{n-1} P(\lambda_{i+1}|\lambda_i) \times P(\lambda_B|\lambda_n)$.

In the FFS method, the parameter λ is first chosen, together with values for λ_A, λ_B, and $\{\lambda_1, \ldots, \lambda_n\}$. No assumption is made about the reaction coordinate: the choice of λ affects only the efficiency of the calculation. It is also convenient to define a series of ''subsurfaces'' $\{\lambda_i^{(1)}, \ldots, \lambda_i^{(m_i)}\}$, in between each pair of surfaces λ_i and λ_{i+1}, such that $\lambda_i^{(1)} = \lambda_i$ and $\lambda_i^{(m_i)} = \lambda_{i+1}$.

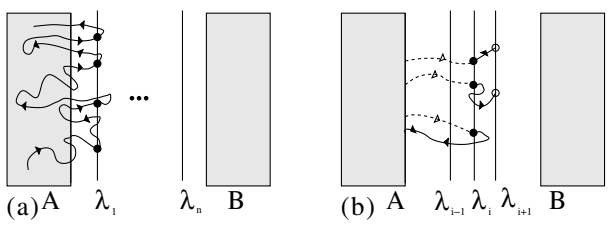

FIG. 1. The first (a) and second (b) stages of the FFS method. The distribution of points at the interfaces depends on the history of the paths, as illustrated by the dashed lines in (b).

In the first stage of the algorithm, a simulation is carried out in region A. After equilibration, $\lambda(t)$ is monitored during a run of length T. Whenever the trajectory crosses the surface λ_1, coming from A, a counter N_f is incremented. If N_f is less than a user-defined number C_1, the phase space coordinates of the system are stored and the run is continued. After time T, one is left with a collection of C_1 points at λ_1, as well as an estimate of the flux $\overline{\Phi}_{A,1}/\overline{h}_A = N_f/T$. This procedure is illustrated schematically in Fig. 1(a): crossings of surface λ_1 that are labeled with a black circle contribute to N_f and to the collection of points at λ_1.

In the second stage of the algorithm [see Fig. 1(b)], M_i trial runs are generated for each surface λ_i. In each trial run, a phase space point from the collection at λ_i is chosen at random and used to start a run, which is continued until the system crosses either λ_{i+1} or λ_A. The maximum λ value, $\lambda_{\max}$, achieved during this run is recorded, and counters N_i^j for all the subsurfaces $\lambda_i^{(j)} \leq \lambda_{\max}$ are incremented by one. After M_i trials, an estimate is obtained for $P(\lambda_i^{(j)}|\lambda_i) = N_i^j/M_i$, for $1 \leq j \leq m_i$ [note that $P(\lambda_i^{(1)}|\lambda_i) = 1$ and $P(\lambda_i^{(m_i)}|\lambda_i) = P(\lambda_{i+1}|\lambda_i)$]. A new collection of C_{i+1} points at the surface λ_{i+1} has also been generated: these are the end points of those trials which make it to λ_{i+1}. These points are then used for trial runs to the next interface, and so on. Eventually λ_B is reached, and one is left with a series of histograms $P(\lambda_{i+1}^{(j)}|\lambda_i)$, for $1 < j \leq m_i$ and $1 \leq i \leq n$. These are connected by fitting to a polynomial to give a curve $P(\lambda|\lambda_1)$, the value of which at $\lambda = \lambda_B$ is $P(\lambda_B|\lambda_1)$. The rate constant k_{AB} is obtained on multiplying $P(\lambda_B|\lambda_1)$ by the flux $\overline{\Phi}_{A,1}/\overline{h}_A$ calculated previously [see Eq. (1)].

The interfaces allow FFS to cross the "barrier" efficiently even for systems with highly stochastic dynamics, in contrast to other schemes [13]. FFS is not the first path sampling method to use a series of interfaces in phase space [12,14]. However, to our knowledge, it is the first such method which does not require prior knowledge of the phase space density. In the method of van Erp *et al.* [12], this is required to propagate paths backwards in time. FFS also differs from other approaches [14], in that it does not assume that the distribution of phase space points at the interfaces $\{\lambda_1, \ldots, \lambda_n\}$ is equal to the stationary distribution of states: each point at interface i lies on a path which originates in region A [see Fig. 1(b)]. We show below that this is essential for the genetic switch.

		k_f	k_b
$2A \rightleftharpoons A_2$	$2B \rightleftharpoons B_2$	$5k$	$5k$
$O + A_2 \rightleftharpoons OA_2$	$O + B_2 \rightleftharpoons OB_2$	$5k$	k
$O \rightarrow O + A$	$O \rightarrow O + B$	k	
$OA_2 \rightarrow OA_2 + A$	$OB_2 \rightarrow OB_2 + B$	k	
$A \rightarrow \varnothing$	$B \rightarrow \varnothing$	$0.25k$	

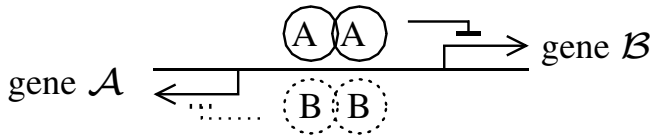

FIG. 2. Reaction scheme for the genetic switch. Forward and backward rate constants k_f and k_b are identical for A and B.

We have applied FFS to a genetic toggle switch [15,16]. This could be regarded as a model for a switch recently constructed *in vivo* [4], or a very minimal representation of the phage λ switch [1]. The switch consists of two proteins A and B and their corresponding genes $\mathcal{A}$ and $\mathcal{B}$ (see Fig. 2). A and B form homodimers A_2 and B_2 which can bind to the DNA (here labeled O). When A_2 is bound, gene $\mathcal{B}$ is not transcribed, while B_2, when bound, correspondingly blocks transcription of gene $\mathcal{A}$: thus A and B mutually repress one another's production. We consider here the "exclusive" version of the switch, in which only one dimer can bind to the DNA at any time (for example, the operator regions of genes $\mathcal{A}$ and $\mathcal{B}$ might be overlapping [15]). The scheme is symmetric on interchanging $\mathcal{A}$ with $\mathcal{B}$ and A with B (although the FFS method is equally applicable to asymmetric switches, such as that of phage λ). All rate constants are relative to that for protein production (k), so that the unit of time is k^{-1}. The volume of the system is taken to be unity. Simulations were carried out using the Gillespie algorithm [6], which is a kinetic Monte Carlo scheme [17] that propagates numbers of molecules in time according to the chemical master equation.

Figure 3 shows the results of a "brute-force" simulation of the switch. The difference Δ in the total copy numbers of the two proteins, $\Delta = N_B - N_A$, is plotted as a function of time in Fig. 3(a), where $N_A = n_A + 2n_{A_2} + 2n_{OA_2}$, $N_B = n_B + 2n_{B_2} + 2n_{OB_2}$, and n_α denotes the number of molecules of species α. The probability distribution $P(\Delta)$, shown in Fig. 3(b), shows that there are two steady states. We define phase space region A by $\Delta \leq -25$, and B by $\Delta \geq 25$. The system flips stochastically between states with $h_A = 1$ ($h_B = 0$) and with $h_B = 1$ ($h_A = 0$). The rate constant k_{AB} can be obtained by fitting the integral

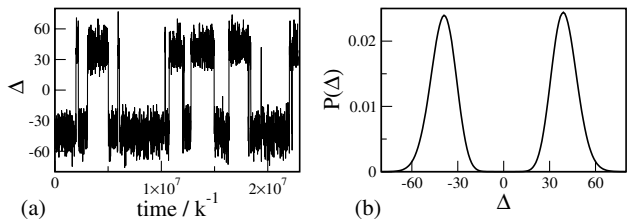

FIG. 3. (a) Δ as a function of time for a typical run. (b) Probability $P(\Delta)$, calculated over a total time of $5 \times 10^9 k^{-1}$.

TABLE I. FFS and brute-force results for $f \equiv \overline{\Phi}_{A,1}/\overline{h}_A$, $P(\lambda_B|\lambda_1)$, and k_{AB} [see Eq. (1)]. For all runs, $\lambda_A = -\lambda_B$. The top three rows show FFS results, averaged over ten runs, for various values of λ_A and numbers n of interfaces. The fourth row shows f, $P(\lambda_B|\lambda_1)$, and $k_{AB} = f \times P(\lambda_B|\lambda_1)$, evaluated using brute-force simulations of total length $9 \times 10^9 k^{-1}$. The number in the bottom right-hand corner is the brute-force result obtained by fitting $F(t)$ as described in the text.

λ_B	n	$f/k \times 10^{-2}$	$P(\lambda_B\|\lambda_1) \times 10^{-5}$	$k_{AB}/k \times 10^{-7}$
30	16	2.98 ± 0.01	3.2 ± 0.1	9.5 ± 0.3
25	11	1.211 ± 0.007	7.8 ± 0.3	9.5 ± 0.3
20	10	0.282 ± 0.002	33.7 ± 0.8	9.6 ± 0.2
25		1.2112 ± 0.004	7.70 ± 0.09	9.3 ± 0.1
				9.4 ± 0.2

$F(t) = \int_0^t dt' p(t')$ of the distribution $p(t)$ of times between flips to the Poisson function $F(t) = 1 - \exp[-k_{AB}t]$. The result is $k_{AB} = (9.4 \pm 0.2) \times 10^{-7} k$ (total simulation time $9 \times 10^9 k^{-1}$; 8808 flips).

We next recalculated k_{AB} using FFS, taking $\lambda = \Delta$. Several calculations were carried out using different values of λ_A, λ_B, and $\{\lambda_1, \ldots, \lambda_n\}$. In all cases, we set $\lambda_1 = \lambda_A = -\lambda_B$, and for each surface λ_i, $C_i = 10^4$ points were stored and $M_i = 10^5$ shooting trials were made. All results were averaged over ten independent FFS runs, leading to rate constants with error bars similar to those of the brute-force result.

Table I shows good agreement between the FFS and brute-force results. The FFS result for k_{AB} does not depend on λ_A, on n, or on the position of the interfaces. These parameters affect only the efficiency of the method, which will be discussed in detail in future work.

We have also calculated the flipping rate for a general version of the switch [15], in which A_2 and B_2 dimers can bind to the DNA simultaneously (again, $k_f = 5k$; $k_b = k$). The result of the FFS calculation [$k_{AB} = (4.11 \pm 0.07) \times 10^{-5} k$] is again in good agreement with the brute-force result [$k_{AB} = (4.21 \pm 0.05) \times 10^{-5} k$].

Returning to the exclusive switch, the CPU time for the FFS calculation was 40–90 times less than for the brute-force simulation. The CPU time for FFS increases much more slowly as the rate k_{AB} decreases than for the brute-force approach. As an example of a very rare event, we have calculated the flipping rate for a switch in which the rate constant for protein degradation is reduced to $0.175k$. Using FFS, we obtain $k_{AB} = (1.92 \pm 0.09) \times 10^{-9} k$, 500 times slower than the switch considered above. This result would have been extremely difficult to obtain using brute-force simulation.

Figure 4(a) shows five of the transition paths obtained for the switch of Fig. 2, plotted as a function of N_A and N_B. To obtain these paths, we begin with the collection of partial trajectories that reach λ_B from λ_n and trace these back via the intervening surfaces to λ_A. The transition path ensemble (TPE) is seen to be rather broad in the N_A, N_B plane. Figure 4(b) shows the difference in the number of *free* protein molecules $\tilde{\Delta} = (n_B + 2n_{B_2}) - (n_A + 2n_{A_2})$, and the occupancy of the operator sites, as functions of time, for a particular transition path. Also plotted is the committor, $P_B(x)$. This is the probability that a new trajectory fired from point x will reach region B before region A [11], here estimated by firing 100 test trajectories. It is clear that the transition process is quite diffusive: $\tilde{\Delta}$ and P_B increase rather smoothly, and the operator state changes many times during the transition process.

The transition state ensemble (TSE) is defined as the collection of points along paths in the TPE for which $P_B = 0.5$. We have extracted these points from a sample of 1000 paths. Figure 4(c) shows the probability distribution $p(\tilde{\Delta})$ for the TSE, separated into the components due to the three operator states O, OA_2, and OB_2 (color coded). The operator state and the number of free molecules are corre-

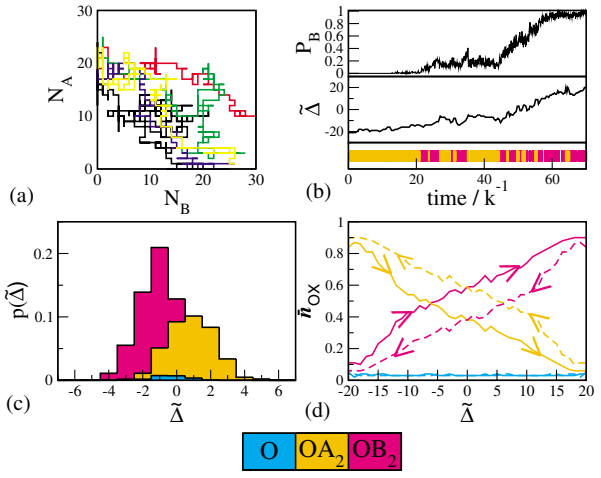

FIG. 4 (color). (a) N_A and N_B, plotted for five transition paths. (b) P_B (top), $\tilde{\Delta} = (n_B + 2n_{B_2}) - (n_A + 2n_{A_2})$ (middle), and operator occupancy (bottom, color coded) are plotted versus time for a typical path. (c) Probability $p(\tilde{\Delta})$ for the TSE with $P_B = 0.5$. $p(\tilde{\Delta})$ is split into color-coded contributions from the three operator states. The area of each histogram gives the probability $\overline{n}_{OX}$ of finding the operator in a particular state (the three areas thus sum to unity) (d) $\overline{n}_{OX}$ as a function of $\tilde{\Delta}$. Solid lines correspond to transitions from A to B, and dotted lines to those from B to A, obtained by exchanging $\mathcal{A}$ with $\mathcal{B}$ and A with B (i.e., $\tilde{\Delta} \to -\tilde{\Delta}$ and $OA_2 \leftrightarrow OB_2$).

lated: the histograms for OA_2 are shifted to higher values of $\tilde{\Delta}$ than those for OB_2. This shows that when A_2 is bound to the DNA, on average, a larger excess of free B molecules (larger $\tilde{\Delta}$) is required to obtain $P_B = 0.5$ than when B_2 is bound. Thus the reaction coordinate depends both on $\tilde{\Delta}$ and the operator state.

A key point to note from Fig. 4(c) is that, although the switch is symmetric on interchanging $\mathcal{A}$ with $\mathcal{B}$ and A with B, the TPE for the transition from A to B does not coincide with that from B to A. For $P_B = 0.5$, operator states OA_2 and OB_2 are not equally populated (the DNA is mostly in state OB_2). Moreover, the equivalent plots for $P_B = 0.2$ and for $P_B = 0.8$ do not map onto one another on reversing the roles of $\mathcal{A}$ and $\mathcal{B}$ and A and B (data not shown). This asymmetry is further illustrated in Fig. 4(d), where the probability $\overline{n}_{\mathrm{OX}}$ of finding the operator in a particular state is plotted as a function of $\tilde{\Delta}$ for paths from A to B (solid lines) and from B to A (dotted lines). Clearly, the forward and backward paths do not coincide. Rare event problems in condensed matter physics often have dynamics that obey detailed balance and microscopic reversibility. In that case, the ensembles of forward and reverse transition paths must be the same. In contrast, our system does not obey detailed balance [9], and so the forward and reverse path ensembles need not be the same.

In this Letter, we have presented a rare event simulation of a biochemical network. To this end, we have developed a scheme, forward flux sampling, that makes it possible to generate transition paths and calculate rates in systems where the phase space density is not known *a priori*. In FFS, the barrier separating the stable states A and B is traversed in a ratchetlike manner, making the method highly suitable for very rare events. We have used FFS to analyze the rate and mechanism of the flipping of a genetic switch: we find that, although the switch is symmetric on exchanging $\mathcal{A}$ with $\mathcal{B}$ and A with B, the TPE for the A to B transition differs from that for B to A. This implies that the distribution of transition paths does not follow the steady state phase space density [18] (which must be symmetric). Hence, even when the dynamics is highly stochastic, one should not *a priori* assume equilibration during the transition process. Finally, the FFS method is not limited to biochemical networks: it could be used with any stochastic dynamical scheme to study rare events and nonequilibrium transitions.

We thank Peter Bolhuis for very helpful discussions and Daan Frenkel, Rutger Hermsen, and Harald Tepper for their careful reading of the manuscript. This work is part of the research program of the "Stichting voor Fundamenteel Onderzoek der Materie (FOM)," which is financially supported by the "Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO)." R. J. A. was partly funded by the European Union Marie Curie Program.

*Electronic address: tenwolde@amolf.nl

[1] M. Ptashne, *A Genetic Switch: Gene Control and Phage λ* (Blackwell, Oxford, 1992).
[2] E. M. Ozbudak, M. Thattai, H. N. Lim, B. I. Shraiman, and A. van Oudenaarden, Nature (London) **427**, 737 (2004).
[3] J. R. Pomerening, E. D. Sontag, and J. E. Ferrell, Jr., Nat. Cell Biol. **5**, 346 (2003); W. Xiong and J. E. Ferrell, Jr., Nature (London) **426**, 460 (2003).
[4] T. S. Gardner, C. R. Cantor, and J. J. Collins, Nature (London) **403**, 339 (2000).
[5] F. J. Isaacs, J. Hasty, C. R. Cantor, and J. J. Collins, Proc. Natl. Acad. Sci. U.S.A. **100**, 7714 (2003); A. Becskei, B. Séraphin, and L. Serrano, EMBO J. **20**, 2528 (2001).
[6] D. T. Gillespie, J. Phys. Chem. **81**, 2340 (1977).
[7] C. J. Morton-Firth and D. Bray, J. Theor. Biol. **192**, 117 (1998); J. S. van Zon and P. R. ten Wolde, q-bio.MN/0404002.
[8] E. Aurell, S. Brown, J. Johanson, and K. Sneppen, Phys. Rev. E **65**, 051914 (2002).
[9] P. B. Warren and P. R. ten Wolde (to be published); M. Mukamel, in *Soft and Fragile Matter*, edited by M. E. Cates and M. R. Evans (IOP Publishing, London, 2000).
[10] D. Frenkel and B. Smit, *Understanding Molecular Simulation. From Algorithms to Applications* (Academic Press, Boston, 2002), 2nd ed.
[11] C. Dellago, P. G. Bolhuis, F. S. Csajka, and D. Chandler, J. Chem. Phys. **108**, 1964 (1998); C. Dellago, P. G. Bolhuis, and P. L. Geissler, Adv. Chem. Phys. **123**, 1 (2002).
[12] T. S. van Erp, D. Moroni, and P. G. Bolhuis, J. Chem. Phys. **118**, 7762 (2003).
[13] G. E. Crooks and D. Chandler, Phys. Rev. E **64**, 026109 (2001).
[14] D. Moroni, P. G. Bolhuis, and T. S. van Erp, J. Chem. Phys. **120**, 4055 (2004); A. K. Faradjian and R. Elber, J. Chem. Phys. **120**, 10 880 (2004).
[15] P. B. Warren and P. R. ten Wolde, Phys. Rev. Lett. **92**, 128101 (2004).
[16] T. B. Kepler and T. C. Elston, Biophys. J. **81**, 3116 (2001); J. L. Cherry and F. R. Adler, J. Theor. Biol. **203**, 117 (2000).
[17] A. B. Bortz, M. H. Kalos, and J. L. Lebowitz, J. Comput. Phys. **17**, 10 (1975).
[18] P. R. ten Wolde and D. Chandler, Proc. Natl. Acad. Sci. U.S.A. **99**, 6539 (2002).