Fluctuation-Preserving Coarse Graining for Biochemical Systems

Bernhard Altaner and Jürgen Vollmer
Max-Planck-Institut für Dynamik und Selbstorganisation (MPIDS), 37077 Göttingen, Germany
Fakultät für Physik, Universität Göttingen, 37077 Göttingen, Germany
(Received 1 February 2012; published 29 May 2012)

Finite stochastic Markov models play a major role in modeling biological systems. Such models are a coarse-grained description of the underlying microscopic dynamics and can be considered mesoscopic. The level of coarse-graining is to a certain extent arbitrary since it depends on the resolution of accommodating measurements. Here we present a systematic way to simplify such stochastic descriptions which preserves both the meso-micro and the meso-macro connections. The former is achieved by demanding locality, the latter by considering cycles on the network of states. Our method preserves fluctuations of observables much better than naïve approaches.

DOI: 10.1103/PhysRevLett.108.228101 PACS numbers: 87.10.Mn, 05.40.−a, 05.70.Ln, 87.18.Tt

In recent years nonequilibrium fluctuations have become the center interest of stochastic thermodynamics [1,2]. Rare events in situations far from equilibrium can now be described universally by fluctuation theorems [3–5]. Especially for stochastic modeling of biophysical processes, which had started in the 1960s with Hill's cycle kinetics [6,7], the attention has shifted from average behavior to the importance of fluctuations, cf., e.g., Ref. [8].

Although Hill's methods were designed for biological problems, they have also lead to general insights in statistical physics [9] and mathematics [10,11]. It was understood that in nonequilibrium situations currents are driven by nontrivial forces called affinities. Affinities have a direct thermodynamic interpretation when they are applied to cycles on the network of states rather than to the states themselves [5,9]. This hints at possible redundancy in the description; indeed, Hill has asked how and when a network reduction would be possible [7]. In the context of statistical physics such a reduction is referred to as coarse graining (CG). It was recently shown for a special CG procedure that the ability to capture fluctuations depends on the preservation of the cycle topology of the network [12].

In this Letter we present a new paradigm for coarse graining of stochastic dynamics that preserves the nonequilibrium steady-state fluctuations of physical currents. Though we focus on biological situations the method can be universally applied to any finite model of stochastic thermodynamics. Our method is based on two requirements: (i) the preservation of the topological and algebraic structure of the cycles of the network and (ii) locality. Additionally, (iii) the variation of the system's entropy along single trajectories [3] is considered to close the equations. To illustrate our method we consider the molecular motor kinesin that performs directed motion along intracellular filaments called microtubuli [2,13–15]. It has two heads (active sites) where adenosine triphosphate (ATP) is catalytically split into adenosine diphosphate (ADP) and inorganic phosphate (P). During the reaction, the molecule undergoes a conformational change that couples the two active sites and induces a mechanical transition. This allows the motor to ''walk'' in a ''hand-over-hand'' motion [13].

The catalytic cycle of a single head [Fig. 1(a)] is an example of general enzymatic activity (Fig. 2). This mesoscopic, stochastic description with its fluctuations has its origins in a microscopic, deterministically chaotic dynamical system. Here, we investigate how a stochastic description can be further simplified while preserving its fluctuations.

Stochastic formalism.—We consider a Markov process on a finite number of mesoscopic states $i \in [1, \ldots, N]$. We call them mesoscopic because for physical systems they amount to a partition of the underlying microscopic phase space. Transitions between the states i and j occur with time-independent rate constants $w^i_j \geq 0$. For simplicity, we assume that there is only one mechanism by which the transition between two states can happen (although a

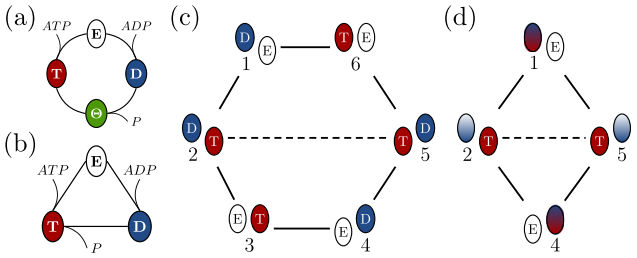

FIG. 1 (color online). (a) The catalytic cycle at kinesin's active site. In a four-stage process ATP binds (state T) to the empty molecule (state E) and is split into Θ = ADP + P. Then first P and later ADP (state D) is released. Since the release of P happens immediately after the splitting, often a three-stage process (b) is assumed where the state Θ is absorbed into its neighbor states. (c) 6-state model of kinesin [2]. The dashed line is the mechanical transition that allows the motor to move. (d) Coarse-grained description with states 3 and 6 reduced.

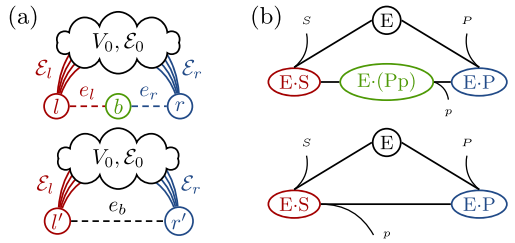

FIG. 2 (color online). (a) Illustration of the coarse-graining procedure that leaves cycle topology constant: reduction of a bridge b will absorb the two dashed edges of the original graph (top) into one edge in the coarse-grained graph (bottom). This also leads to a change of rates along the edges $\mathcal{E}_l$ and $\mathcal{E}_r$, whereas edges $\mathcal{E}_0$ connecting only unchanged vertices V_0 remain unchanged. (b) Enzyme catalysis. An enzyme E binds a substrate S to form a complex ES. The substrate is split to form products P and p where the latter is always released first. The dynamics can be modeled with four (top) and three (bottom) states.

generalization is possible [1,5]). Because of the reversibility of microscopic physical laws we demand dynamical reversibility, i.e., $w_j^i > 0 \Leftrightarrow w_i^j > 0$. One can visualize the system as a graph $G = (V, \mathcal{E})$ with the mesoscopic states as vertices V and edges $\mathcal{E}$ where $w_j^i > 0$. At time t the system will be in state i with a probability $p_i(t)$. The flux from state i to state j is

$$\phi_j^i := p_i w_j^i, \qquad i \neq j. \tag{1}$$

Assuming connectedness of the network, a unique invariant distribution p_i exists [16]. In the steady state the net influx to each state equals the net outflux (Kirchhoff's law for currents $I_j^i := \phi_j^i - \phi_i^j$),

$$\sum_j I_i^j = \sum_j [\phi_i^j - \phi_j^i] = \sum_j [p_j w_i^j - p_i w_j^i] = 0 \ \forall i. \tag{2}$$

The steady-state probability distribution can be calculated explicitly as a polynomial in the rates w_j^i using a graph-theoretic matrix-tree method [7]. Henceforth, all variables are time independent steady-state quantities unless stated otherwise. Equation (2) can be used to decompose the steady-state fluxes using different sets of cycles on G [5,7,9]. A cycle α of length s_α is an ordered set of vertices which form a self-avoiding closed path, where we identify cycles differing only by a cyclical permutation of vertices. Hereafter, when referring to cycles we mean nontrivial cycles with $s_\alpha \geq 3$. Central quantities for this work are the edge affinities $A_j^i = \log(\phi_j^i/\phi_i^j)$. Along a cycle $\alpha = (i_1, i_2, \ldots, i_{s_\alpha})$ they add up to cycle affinities

$$A_\alpha = \sum_{k=1}^{s_\alpha} \log\left[\frac{\phi_{i_k}^{i_{k-1}}}{\phi_{i_{k-1}}^{i_k}}\right] = \sum_{k=1}^{s_\alpha} \log\left[\frac{w_{i_k}^{i_{k-1}}}{w_{i_{k-1}}^{i_k}}\right]. \tag{3}$$

In physical models they take only values that reflect the macroscopic thermodynamic affinities [7,9].

Coarse graining.—We suggest a coarse-graining procedure based on natural requirements: (i) (a) cycle topology: The number and mutual connections of cycles are preserved. This determines possible targets for the reduction. (i) (b) cycle affinities: The algebraic values of the affinity of any cycle is preserved. This yields the connection to the macroscopic level, i.e., thermodynamics. (ii) locality: Fluxes, probabilities and observables may only change locally. This yields the connection to the microscopic level, i.e., the microscopic phase space. (iii) trajectories: The system's entropy variation along trajectories is preserved. This is a natural choice and closes the equations.

To demonstrate our method we address cycles that contain bridge states, e.g., states 3 and 6 in Fig. 1. Bridges are connected to exactly two neighbor states that are themselves not connected to each other as shown in Fig. 2(a). We use the index b for the target bridge state, and l or r for the left or right neighbor. Without loss of generality we assume that there is a positive net current $I = I_b^l = I_r^b > 0$ flowing through the bridge from the left to the right neighbor state. The other states, which must not be influenced by the procedure [cf., (ii)] are summarized in the set $V_0 \subset V$. In the CG procedure we absorb the bridge into its neighbors leading to new states l' and r' and adjust the transition rates for the sets of edges $\mathcal{E}_l$ and $\mathcal{E}_r$ connecting l and r to the rest of the network. This has to be done in accordance with requirement (i) (b) yielding $A_\alpha = A'_\alpha$ for any cycle in the network. Demanding the conservation of fluxes along any edge $e \in \mathcal{E}/\{e_l, e_r\}$ not belonging to the bridge, Eq. (3) yields

$$\phi_{r'}^{l'}/\phi_{l'}^{r'} \overset{!}{=} (\phi_b^l \phi_r^b)/(\phi_b^r \phi_l^b). \tag{4a}$$

Any trajectory passing through the two edges (l, b) and (b, r) in the original model will be a trajectory through (l', r') in the coarse-grained model. The change of a trajectory's entropy (starting from a steady-state ensemble) is the difference of the logarithms of the invariant distribution [3]. With that, (iii) leads to

$$p_{l'}/p_{r'} \overset{!}{=} p_l/p_r. \tag{4b}$$

A priori, other closures of the form $p_{l'}/p_{r'} = c$ are also possible but lack the advantage of the stochastic thermodynamic interpretation.

Together with the steady-state balance condition [Eq. (2)] and the locality assumption, Eqs. (4) uniquely determine all rate constants of the coarse-grained model. They can be found to be

$$w_{n'}^i = w_n^i \qquad \text{for } i \in V_0, \qquad n \in \{l, r\}, \tag{5a}$$

$$w_i^{n'} = w_i^n/f \qquad \text{for } i \in V_0, \qquad n \in \{l, r\}, \tag{5b}$$

$$w_{r'}^{l'} = (I + m)/(f p_l), \tag{5c}$$

$$w_{l'}^{r'} = m/(f p_r), \tag{5d}$$

where

$$f = (p_l + p_r + p_b)/(p_l + p_r), \tag{6a}$$
$$m = \phi_b^r \phi_l^b/(I + \phi_b^r + \phi_l^b). \tag{6b}$$

One may argue that the method might not be practical, because one has to solve the original model to compute the rates for the simpler model. However, to derive the CG model one only needs the steady-state probability distribution of the original model which is accessible numerically to arbitrary precision by different means [16]. The much more difficult task of determining the full fluctuation spectra of any observable (see below) can then be done in the simplified model.

Single cycle: simple catalysis model.—The easiest reducible topology is a cycle consisting of four states, e.g., the enzyme catalysis presented in Figs. 1(a) and 2(b) where a transient intermediate state is identified as the bridge. A naïve approximation for the new rates would be $w_{r'}^{l'} = w_b^l w_r^b \langle \tau_b \rangle$ and $w_{l'}^{r'} = w_b^r w_l^b \langle \tau_b \rangle$ where $\langle \tau_b \rangle^{-1} = w_l^b + w_r^b$ is the time constant for decay out of the bridge state. Hill [7] derived this result for three linearly connected states with the center one being transient. This choice is also the basis of the method proposed in Ref. [12], where its shortcomings have already been discussed. It fulfills

$$w_{r'}^{l'}/w_{l'}^{r'} = (w_b^l w_r^b)/(w_b^r w_l^b). \tag{7}$$

In Ref. [15] this relation is interpreted as a condition on the local energy landscape and was used to reduce the enzymatic reaction of kinesin's active site [Fig. 1(a)]. If all other rates are unchanged, Eq. (7) preserves the affinity [Eq. (3)] of the cycle. However, in general, it leads to a nonlocal redistribution of steady-state probabilities. Hence, it does not comply with our method. This yields another motivation of Eq. (4b). One can easily check that this condition on the ratio of the new probabilities is the only one that leads to rates that fulfill Eq. (7).

Fluctuations of physical observables.—Since the CG procedure changes the mesoscopic state space, coarse-grained physical observables need to be defined. We consider physical currents, which are modeled by anti-symmetric matrices $O \in \mathbb{R}^{N\times N}$ that assign a value $O_j^i = -O_i^j$ to each transition $i \to j$. The observable $\tilde{O}$ for the case where the bridge state has been eliminated has entries

$$\tilde{O}_{r'}^{l'} = O_b^l + O_r^b + (d_l - d_r), \tag{8a}$$
$$\tilde{O}_i^{n'} = O_i^n + d_n \quad \text{for } i \in V_0, \quad n \in \{l, r\}, \tag{8b}$$
$$\tilde{O}_j^i = O_j^i \quad \text{for } i, j \in V_0. \tag{8c}$$

The constants d_l and d_r depend on microscopic dynamics and the chosen partitioning of phase space. As we do not know these details, d_l and d_r act as gauges that do not change the macroscopic observations; therefore we choose $d_l \equiv d_r \equiv 0$ for simplicity.

A special observable with a differently prescribed gauge is the quantity

$$B_j^i = \log(w_j^i/w_i^j). \tag{9}$$

It is determined solely by the mesoscopic transition rates and therefore takes a special role. Seifert identifies it with the heat dissipated in the medium for transition $i \to j$ [3]. Hill calls it the basic free energy difference between two mesoscopic states [7]. One finds

$$\tilde{B}_{r'}^{l'} = B_b^l + B_r^b, \tag{10a}$$
$$\tilde{B}_i^{n'} = B_i^n - \log f \quad \text{for } i \in V_0, \quad n \in \{l, r\}, \tag{10b}$$
$$\tilde{B}_j^i = B_j^i \quad \text{for } i, j \in V_0. \tag{10c}$$

Equation (10a) is the logarithm of Eq. (7). Equation (10b) states, that along the edges $\mathcal{E}_n$, $n \in \{l, r\}$, there is an additional contribution $-\log f$ to $\tilde{B}_i^n$, which is the same for both neighbors due to the closure [Eq. (4b)]. Equation (10c) expresses locality and is independent of the closure. We note that observables which are defined on the states rather than on the transitions (and therefore are no current observables) can also be consistently transformed [17].

Multiple cycles: kinesin's network of states.—To investigate the steady-state fluctuations of the observables we consider stochastic trajectories $\omega = (\omega_0, \omega_1, \ldots, \omega_{N_\omega})$ featuring N_ω jumps in a prescribed time $\tau_\omega = \tau$. The time-averaged mean of current observable O along trajectory ω,

$$j_\tau^O(\omega) = \frac{1}{\tau} \sum_{i=1}^{N_\omega} O_{\omega_i}^{\omega_{i-1}}, \tag{11}$$

is a bounded random variable with the distribution function f_τ^O. For $\tau \to \infty$ it converges weakly and fulfills a large-deviation principle, i.e.,

$$f_\tau^O(s) = \exp[-\tau I_O(s) + o(\tau)], \tag{12}$$

where $o(\tau)$ stands for a term sublinear in τ. Further, by the Gärtner-Ellis theorem [18], the large-deviation function $I_O(s)$ is the unique Legendre transform of the scaled cumulant generating function (SCGF)

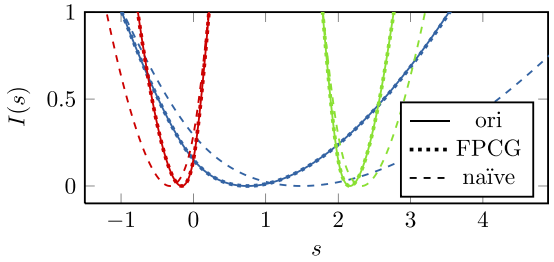

FIG. 3 (color online). Large-deviation function (LDF) $I(s)$ for the entropy production (blue, center), the product (red, left) and the substrate (green, shifted to the right by $s_0 = 2$) association for the enzyme model [Fig. 2(b)]. All transition rates are unity but the release rate for the first product p which has the value 100. The LDF obtained from the fluctuation preserving coarse-graining method (FPCG) overlaps almost perfectly with the original model (ori) while the naïve choice strongly changes fluctuations.

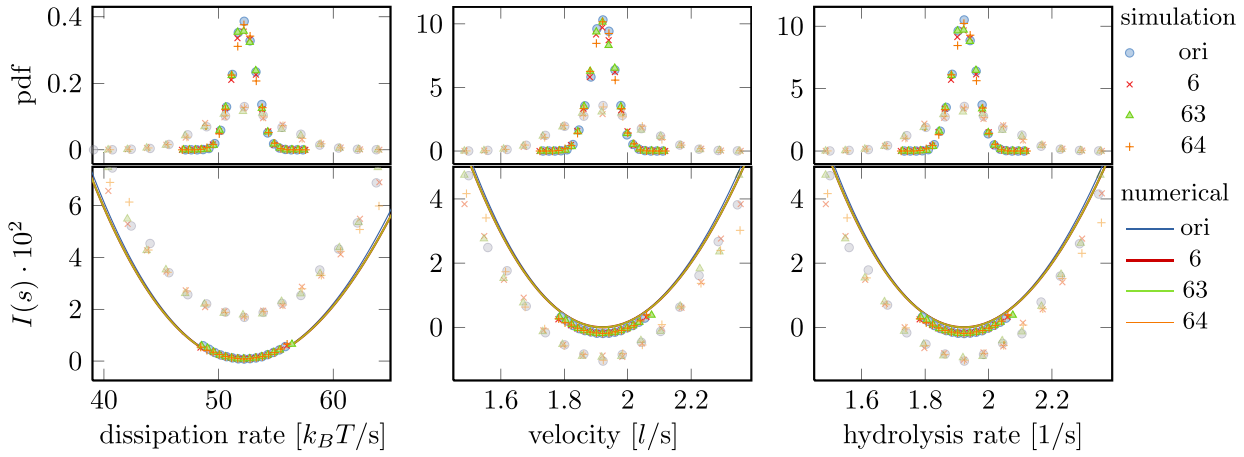

FIG. 4 (color online). Simulation and numerical results for dissipation rate, moving velocity and hydrolysis rate of the kinesin model. Data are shown for the original 6-state model (ori), a 5-state model with state 6 reduced (6) and two four-state models with states 6,3 or 6,4 reduced (63 and 64, respectively). The rate constants for the original model are taken from Ref. [2] describing the data in Ref. [14] for chemical concentrations $c_{\mathrm{ADP}} = c_{\mathrm{P}} = c_{\mathrm{ATP}} = 1\ \mu\mathrm{M}$ and stepping size $l \approx 8$ nm. The top row shows the sampled pdf for $\tau \approx 1200$ s (opaque symbols) and $\tau \approx 120$ s (transparent symbols). The bins with the width of half an empirical standard deviations are centered around the empirical mean. For the simulation we sampled $N = 5000$ trajectories. The bottom row shows convergence of rescaled data [cf., Eq. (12)] to the numerically obtained rate function $I(s)$ (solid lines).

$$\zeta(\lambda) = \lim_{\tau\to\infty} \frac{1}{\tau} \log \mathbb{E}\left[\exp(\lambda \tau j_\tau^O)\right], \tag{13}$$

where $\mathbb{E}[\cdot]$ denotes the expectation value on the space of trajectories running for time τ. The SCGF can be calculated [18] as the dominant eigenvalue of the tilted transition matrix $W_O(\lambda)$ with entries

$$(W_O)^i_j = w^i_j \exp(\lambda O^i_j). \tag{14}$$

To obtain numerical data for the rate function $I_O(s)$ of an observable O we first calculate $W_O(\lambda)$, determine its largest eigenvalue $\zeta(\lambda)$ and find its Legendre transform with respect to λ. For the last step, the algorithm described in Ref. [19] is used.

Figure 3 shows such numerical results for different physical currents of the enzyme model [Fig. 2(b)]. Unlike the naïve choice for the rates (dashed lines), our CG method (dotted lines) preserves steady-state averages and fluctuations of the original model (solid lines) to a very high degree. Bounds of the deviation can be obtained from inequalities for Perron-Frobenius eigenvalues [17].

Our CG mechanism also captures fluctuations of observables for finite times and in models with multiple cycles. Figure 1(c) shows kinesin's network of states [2]. Using our method [Eqs. (5)] we reduced the bridge states appearing in the diagram. The result of a successive reduction of states 6 and 3 is shown in Fig. 1(d). Additionally, we analyzed models with only state 6 reduced and both states 6 and 4 reduced. Figure 4 shows the results of simulations, and the convergence to the large-deviation rate function $I(s)$ for the total dissipation rate (entropy production in the medium), the steady-state velocity and the hydrolysis rate of the kinesin model. The agreement between the original and the reduced models is extremely good already for finite times. The rate function $I(s)$ for the different models agree extremely well to the level of being indistinguishable in the vicinity of the average value. Only in the far tails can one observe that the result is not exact.

Discussion.—In this Letter we presented a new method to simplify stochastic dynamics on finite state spaces. A coarse-graining method that preserves the connection with both the underlying microscopic dynamics and the macroscopic thermodynamics was constructed. Here we considered bridge states, but the same ideas apply to tree-like subgraphs. Two biochemical examples where considered: a generic single-cycle model for enzymatic catalysis and a well-established multicycle model for the molecular motor kinesin. Reduction using the new paradigm preserves fluctuations of current observables in great detail. Future work will focus on coarse graining that includes changes of the cycle topology.

The authors are indebted to L. Rondoni, A. Puglisi, S. Herminghaus, and H. Touchette for many illuminating discussions.

[1] M. Esposito and C. Van den Broeck, Phys. Rev. E **82**, 011143 (2010).
[2] S. Liepelt and R. Lipowsky, Phys. Rev. Lett. **98**, 258102 (2007).
[3] U. Seifert, Phys. Rev. Lett. **95**, 040602 (2005).
[4] U. Seifert, Europhys. Lett. **70**, 36 (2005).
[5] A. Faggionato and D. Di Pietro, J. Stat. Phys. **143**, 11 (2011).
[6] T. Hill, J. Theor. Biol. **10**, 442 (1966).
[7] T. Hill, *Free Energy Transduction in Biology* (Academic Press, New York, 1977).

[8] O. Hallatschek, Proc. Natl. Acad. Sci. U.S.A. **108**, 1783 (2010).
[9] J. Schnakenberg, Rev. Mod. Phys. **48**, 571 (1976).
[10] S. Kalpazidou, *Cycle Representations of Markov Processes* (Springer Verlag, Berlin, 2006).
[11] D. Jiang, M. Qian, and M.-P. Qian, *Mathematical Theory of Nonequilibrium Steady States* (Springer, New York, 2004).
[12] A. Puglisi, S. Pigolotti, L. Rondoni, and A. Vulpiani, J. Stat. Mech. (2010) P05015.
[13] A. Yildiz, M. Tomishige, R. D. Vale, and P. R. Selvin, Science **303**, 676 (2004).
[14] N. Carter and R. Cross, Nature (London) **435**, 308 (2005).
[15] R. Lipowsky and S. Liepelt, J. Stat. Phys. **130**, 39 (2008).
[16] W. Feller, *An Introduction to Probability Theory and Its Applications* (Wiley, New York, 1968), 3rd ed., Vol. 1.
[17] B. Altaner and J. Vollmer (unpublished).
[18] H. Touchette, Phys. Rep. **478**, 1 (2009).
[19] Y. Lucet, Numerical Algorithms **16**, 171 (1997).