Fluctuation-preserving coarse graining for biochemical systems

Bernhard Altaner and Jürgen Vollmer
Max-Planck-Institut für Dynamik und Selbstorganisation (MPIDS), 37077 Göttingen, Germany and
Fakultät für Physik, Universität Göttingen, 37077 Göttingen, Germany
(Dated: June 5, 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 extend 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 connection. 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.

In recent years non-equilibrium fluctuations have become the center interest of stochastic thermodynamics [1, 2]. Rare events in situations far from equilibrium can now be universally described by fluctuation theorems [3, 4]. Intensive stochastic modelling of biophysical processes has started in the 1960s with Hill’s cycle kinetics [3–5]. Intensive stochastic modelling of biophysical processes has started in the 1960s with Hill’s cycle kinetics [3–5]. Intensive stochastic modelling of biophysical processes has started in the 1960s with Hill’s cycle kinetics [3–5].

Although Hill’s methods were designed for biological problems, they have lead to general insights in statistical physics [6] and mathematics [7]. It was understood that in non-equilibrium situations currents driven by non-trivial forces which are usually called affinities. Assigning these affinities to cycles on the network of states rather then to the states themselves, they have a direct thermodynamic interpretation [8, 9]. This hints at possible redundancy in the description and already Hill asked how and when a network reduction would be possible. In statistical physics, such reduction are often summarized under the term of coarse-graining (CG) methods. It was recently shown for a special CG procedure that the ability to capture fluctuations depends on the preservation of cycle topology of the network [10, 11].

In this Letter we present a new paradigm for coarse-graining of stochastic dynamics which preserves the non-equilibrium 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 algebraic and topological structure of the cycles of the network and (ii) locality. Further, (iii) the variation of the system’s entropy along single trajectories is considered to close the equations. To illustrate our method we consider the molecular motor kinesin which is able to perform directed motion along intracellular filaments called microtubuli [2, 12]. It has two heads (active sites) where adenosin triphosphate (ATP) is catalytically split into adenosin 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” mechanism [13].

The catalytic cycle of a single head (Fig. 1a) is an example of a general enzymatic activity (Fig. 2). This mesoscopic, stochastic description with its fluctuations has its origins in a microscopic, deterministically chaotic dynamics. 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 \( i \in [1..N] \). 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 \( w_{ij} \geq 0 \). For simplicity, we assume that there is only one mechanism by which the transition between two states can happen (although a generalization is possible [14]). Because of the reversibility of microscopic physical laws we demand dynamical reversibility, i.e., \( w_{ij} > 0 \iff w_{ji} > 0 \). One can visualize the system as a graph \( G = (V, E) \) with the meso-
Cycle topology: The number and mutual connections of cycles are preserved. This determines possible targets for the reduction.

(ii) 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, like 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. 1. (a) Enzyme catalysis. An enzyme $E$ binds a substrate $P$ and three (bottom) states.

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 $E_l$ and $E_r$ whereas edges $E_0$ connecting only unchanged vertices $V_0$ remain unchanged. (b) Enzyme catalysis. An enzyme $E$ binds a substrate $P$ and forms a complex $E\cdot P$. The substrate $P$ binds a substrate $P$ and forms a complex $E\cdot P$. The substrate $P$ binds a substrate $P$ and forms a complex $E\cdot P$. The substrate $P$ binds a substrate $P$ and forms a complex $E\cdot P$.

Coarse graining We suggest a coarse-graining procedure based on natural requirements:

(a) Cycle topology: The number and mutual connections of cycles are preserved. This determines possible targets for the reduction.

(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, like 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. 1. (a) Enzyme catalysis. An enzyme $E$ binds a substrate $P$ and forms a complex $E\cdot P$. The substrate $P$ binds a substrate $P$ and forms a complex $E\cdot P$. The substrate $P$ binds a substrate $P$ and forms a complex $E\cdot P$. The substrate $P$ binds a substrate $P$ and forms a complex $E\cdot P$. The substrate $P$ binds a substrate $P$ and forms a complex $E\cdot P$.
The basis of the method proposed in Ref. [12], where its
with the center one being transient. This choice is also
[7] derived this result for three linearly connected states
as reducible topology is a cycle consisting of four states,
e.g. This overlap almost perfectly with the original model ("ori") while
the "naïve" choice strongly changes fluctuations.

One may argue that the method might not be practical,
because one has to solve the original model to compute
the correct rates for the simpler model. However, the
situation is better than this: One only needs the steady
state of the original model, which is accessible numerically
to arbitrary precision by different means. In particular, if
the rates of the original model are subject to inaccuracies
either by measurement or by modeling, the numerical
error (if there is any) to the steady-state probabilities is
related to the differences in the rates compared to the original
model, which is accessible numerically. Hence, it does not
change the macroscopic observations. Hence, in general,
we choose $d_l \equiv d_r \equiv 0$ for simplicity.

A special observable, where the gauge is prescribed
differently, is the quantity

$$ B_j^l = \log \left( \frac{w_{ij}^l}{w_{ij}^r} \right). \tag{9} $$

It is determined solely by the mesoscopic transition rates
and therefore takes a special role. Hill calls it the basic
free energy difference between two mesoscopic states [7].
Recently, Seifert made this point more clear by stating the
assumptions, under which it can be identified with the
heat dissipated in the medium for transition $i \rightarrow j$ [3]
[17]. In an electric analogy it would be the electromotive
force $\tau_I$.

One finds

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

Eq. (10a) is the logarithm of Eq. (7). Eq. (10b) states,
that along the edges $E_n, n \in \{l, r\}$, there is an additional
contribution $-\log f$ to $\tilde{B}_i^r$, which is the same for both
neighbors due to the closure [13]. Eq. (10c) expresses locality and is independent of the closure. We note, that
non-current observables defined on the states rather than
the transitions can also consistently be transformed [19].

To investigate the steady-state fluctuations of the
observables we consider stochastic trajectories $\omega = (\omega_0, \omega_1, \ldots, \omega_{N})$ featuring $N$ jumps in a prescribed time
$\tau_\omega = \tau$. The time-averaged mean of current observable
$O$ along trajectory $\omega$,

$$ f^O_\omega(\omega) = \frac{1}{\tau} \sum_{i=1}^{N} O_{\omega_i}^{\omega_{i+1}}, \tag{11} $$

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

$$ f^O_\omega(s) = \exp \left(-\tau I_O(s) + o(\tau)\right), \tag{12} $$

where $o(\tau)$ stands for a term sublinear in $\tau$. Further, by
the Gärtnert-Ells theorem [20], the large-deviation function
$I_O(s)$ is the unique Legendre transform of the scaled

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure3.png}
\caption{(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). 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.}
\end{figure}
cumulant generating function (SCGF)
\[
\zeta(\lambda) = \lim_{\tau \to \infty} \frac{1}{\tau} \log E[\exp(\lambda \tau f_O^\tau)]
\]  
(13)
where \( E[\cdot] \) denotes the expectation value on the space of trajectories running for time \( \tau \). The SCGF can be calculated \[20\] as the dominant eigenvalue of the tilted transition matrix \( W_O(\lambda) \) with entries
\[
(W_O)_{ij}^\tau = w_{ij}^\tau \exp(\lambda O_{ij}^\tau).
\]  
(14)
To obtain numerical data for the rate function \( I_O(s) \) of an observable \( O \) one follows the numerical scheme: Calculate \( W_O(\lambda) \), determine its largest eigenvalue \( \zeta(\lambda) \) and find its Legendre transform with respect to \( \lambda \). For the last step, the algorithm described in Ref. \[21\] was used.

FIG. 4. (color online) Simulation and numerical results for dissipation rate, moving velocity and hydrolysis rate of the kinesin model. Data is shown for the original 6-state model ("ori"), a 5-state model with state 6 reduced ("6") and two 4-state models with state 6 or 6, 4 reduced ("63" and "64"). The rate constants for the original model are taken from Ref. \[2\] describing the data in Ref. \[14\] for chemical concentrations \( c_{ADP} = c_P = c_{ATP} = 1\mu\text{M} \) and stepping size \( l \approx 8\text{nm} \). The top row shows the sampled pdf for \( \tau \approx 1200s \) (opaque symbols) and \( \tau \approx 120s \) (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 rate function \( I(s) \) (solid lines).

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 multi-cycle model for the molecular motor kinesin. The 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 \textbf{82}, 011143 (2010).
2. S. Liepert and R. Lipowsky, Phys. Rev. Lett. \textbf{98}, 258102 (2007).
3. U. Seifert, Phys. Rev. Lett. \textbf{95}, 40602 (2005).
4. U. Seifert, Europhys. Lett. \textbf{70}, 36 (2005).
5. A. Faggionato and D. Di Pietro, J. Stat. Phys. \textbf{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, P.N.A.S. 108, 1783 (2010).
[9] J. Schnakenberg, Rev. Mod. Phys. 48, 571 (1976).
[10] S. Kalpazidou, Cycle representations of Markov processes, Vol. 28 (Springer, Berlin, 2006).
[11] D. Jiang, M. Qian, and M.-P. Qian, Mathematical theory of nonequilibrium steady states (Springer, Berlin Heidelberg, 2004).
[12] A. Puglisi, S. Pigolotti, L. Rondoni, and A. Vulpiani, J. Stat. Mech. 2010, P05015 (2010).
[13] A. Yildiz, M. Tomishige, R. D. Vale, and P. R. Selvin, Science 303, 676 (2004).
[14] N. Carter and R. Cross, Nature 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, 3rd ed., Vol. 1 (Wiley, 1968).
[17] U. Seifert, The European Physical Journal E: Soft Matter and Biological Physics 34, 1 (2011).
[18] B. Altaner, S. Grosskinsky, S. Herminghaus, L. Katthän, M. Timme, and J. Vollmer, Physical Review E 85, 041133 (2012).
[19] B. Altaner and J. Vollmer, (unpublished).
[20] H. Touchette, Physics Reports 478, 1 (2009).
[21] Y. Lucet, Numerical Algorithms 16, 171 (1997).