Equilibrium time correlation functions from irreversible transformations in trajectory space

Phillip L. Geissler*
Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Christoph Dellago
Institute for Experimental Physics, University of Vienna,Boltzmanngasse 5, 1090 Vienna, Austria
(Dated: November 24, 2003)

We present a new identity for the statistical mechanics of trajectories, showing that a distribution of irreversible transformations between ensembles of trajectories is sufficient to determine equilibrium time correlation functions. This general and exact result extends to the dynamical realm recently derived connections between thermodynamic free energies and statistics out of equilibrium. In the context of chemical kinetics, we use the identity to compute reaction rate constants through appropriate averaging of an effective work to switch from non-reactive to reactive trajectory ensembles.

The most interesting dynamical features of a complex system are often much slower than the basic microscopic motions that advance them. For instance, the characteristic time scale of protein folding (~ 1 s for a typical protein) greatly exceeds that of torsional rotations along the backbone (~ 100 ps). Other examples of processes exhibiting widely different time scales include nucleation of first order phase transitions, chemical reaction in solution, dynamics near a glass transition, and transport in and on solids. Such a separation of important time scales seriously hinders the study of dynamical mechanisms. Experiments cannot usually resolve the molecular fluctuations that drive condensed phase dynamics. Computer simulations, on the other hand, can feasibly generate trajectories only a few orders of magnitude longer than the duration of microscopic fluctuations, making long-time behavior elusive.

One common source of disparate time scales is the presence of dynamical bottlenecks between basins of attraction in phase space. Fig. 1 depicts some microstates in such a basin, A, (e.g., a protein's unfolded state) and trajectories emanating from each. If free energy barriers separating A from adjacent basins are much larger than typical thermal excitations, then only a very small fraction of trajectories will escape A to end in some other metastable state, B (e.g., the protein's native state) within microscopic correlation times. Many computational schemes have been devised to explore the important but infrequently visited transition state region, but most presume knowledge of mechanism or introduce ad hoc dynamical rules [1–4]. More systematic strategies simply focus on the rare short trajectories that exhibit transitions of interest [5, 6].

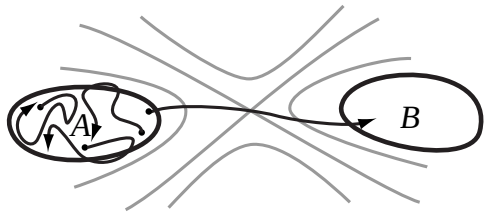

FIG. 1: Several trajectories originating in a basin of attraction A. Free energy contours (in gray) indicate a bottleneck between A and a second basin B. The vast majority of short trajectories do not surmount this free energy barrier and thus do not connect the two basins

We make these ideas quantitative by defining a partition function for the restricted ensemble of microstates whose trajectories of length t end in B,

$$Z_{AB}(t) = \int dx_0 \rho(x_0) h_A[x_0] h_B[x_t(x_0)]. \qquad (1)$$

For the moment we imagine that dynamics are deterministic, so that each phase space point x_0 uniquely specifies a trajectory, represented here as an ordered sequence of microstates separated by a time step Δt , $x(t) = \{x_0, x_{\Delta t}, x_{2\Delta t}, ..., x_t\}$. In Eq. (1) $\rho(x)$ is an equilibrium probability density, for example the canonical distribution function $\rho(x) \propto e^{-E(x)/k_{\rm B}T}$. The functions h_A and h_B project onto A and B respectively, i.e., $h_i[x]$ is unity for x in state i and otherwise vanishes. We now compare $Z_{AB}(t)$ with the partition function for the entire basin, $Z_A = \int dx_0 \rho(x_0) h_A[x_0]$. The logarithm of their ratio determines a free energy difference between the two ensembles, $\Delta F(t) \equiv -k_{\rm B}T \ln [Z_{AB}(t)/Z_A]$, that depends parametrically on time. Although free energy is typically conceived as a static quantity, $\Delta F(t)$ is directly related to an equilibrium time correlation function,

$$C(t) \equiv \frac{\langle h_A[x_0] h_B[x_t(x_0)] \rangle}{\langle h_A \rangle} = \exp\left[-\Delta F(t)/k_{\rm B}T\right], \qquad (2)$$

whose approach to equilibrium describes the phenomeno-

*Current address: Department of Chemistry, University of California at Berkeley, Berkeley, CA 94720.

logical kinetics of transitions from A to B [7]. (We use angled brackets to denote an average over the equilibrium distribution $\rho(x)$.) The unusual partition function $Z_{AB}(t)$ thus connects the thermodynamic concept of free energy to intrinsically dynamical quantities such as rate constants.

In order to clarify this connection in principle, and to make it useful in practice, we along with others have developed a statistical mechanics of trajectories, in close analogy to the familiar statistical mechanics of individual microstates [5, 8, 9]. Ensembles of trajectories, weighted by their frequency of occurrence, form the basis for this approach. The probability density of observing a particular pathway $x(t)$ at equilibrium depends in general on the distribution of initial conditions $\rho(x_0)$, and on transition probabilities for each step in time:

$$P_{\rm equ}[x(t)] = \rho(x_0) \prod_{i=0}^{t/\Delta t-1} p(x_{i\Delta t} \to x_{(i+1)\Delta t}). \quad (3)$$

Integrating the path distribution $P_{\rm equ}[x(t)]$ over all possible pathways consistent with relevant constraints generates the analog of a partition function, $\mathcal{Z}(t)$. For the ensemble of trajectories whose initial points lie in A, $\mathcal{Z}_A(t) = \int \mathcal{D}x(t) P_{\rm equ}[x(t)] h_A(x_0)$. (The notation $\int \mathcal{D}x(t)$ represents integration over all phase space points comprising a trajectory of length t.) Further constraining trajectories to end in B at time t gives $\mathcal{Z}_{AB}(t) = \int \mathcal{D}x(t) P_{\rm equ}[x(t)] h_A(x_0) h_B(x_t)$. For deterministic dynamics transition probabilities are Dirac delta functions, and $\mathcal{Z}_{AB}(t)$ reduces to the function $Z_{AB}(t)$ in Eq. (1). For stochastic dynamics (e.g., Langevin or Monte Carlo), however, transition probabilities may have finite width.

Within this perspective thermodynamic quantities related to partition functions have analogous meaning for trajectories. For example, $-\ln[\mathcal{Z}_{AB}(t)/\mathcal{Z}_A(t)]$ is effectively the "free energy" difference between the ensemble of all trajectories originating in A and the ensemble including only reactive trajectories that originate in A and end in B. As a result of this connection, many standard techniques of statistical mechanics have analogous utility. Importance sampling can be used to focus on rare trajectories of interest, such as those executing transitions between basins of attraction. This transition path sampling technique, in effect a biased random walk through the space of all dynamical paths, has been applied to several very slow but important processes in complex systems [10–13]. Other analogies to equilibrium thermodynamics have been revealed and exploited, primarily toward the end of determining reaction rates and mechanisms [14–16].

In this Letter we show that connections between the statistical mechanics of trajectories and that of microstates exist even *out* of equilibrium. We do so by deriving a dynamical version of the nonequilibrium identity recently presented by Jarzynski [17]. That identity relates differences in equilibrium free energy, or reversible work $W_{\rm rev}$, to repeated irreversible transformations between the corresponding ensembles. This exact relationship amends Clausius's inequality, $\overline{W} \geq W_{\rm rev}$, profoundly. A corollary to the first and second laws of thermodynamics, the standard inequality establishes the average (denoted by an overbar) of expended work W over many realizations of an irreversible process as an upper bound on $W_{\rm rev}$. Jarzynski's result, on the other hand,

$$\overline{\exp\left[-W/k_{\rm B}T\right]} = \exp\left(-W_{\rm rev}/k_{\rm B}T\right), \quad (4)$$

shows how to properly average over nonequilibrium realizations so that $W_{\rm rev}$ is obtained exactly [17]. The mechanical work W appearing in (4) is the total change in energy $E(x)$ accumulated along non-equilibrium transformations due to the action of an external field driving the system out of equilibrium [17, 18].

We extend the concept of mechanical work to trajectory space by defining a path "energy," $\mathcal{E}[x(t)] = -\ln\{P[x(t)]\mathcal{Z}(t)\}$, corresponding to a particular path distribution $P[x(t)]$. In this case, the "work" is the change in path "energy" $\mathcal{E}[x(t)]$ accumulated when changing one path distribution to another at finite rate. If the change is done in N sudden steps, then the total "work" supplied is

$$\mathcal{W} = \sum_{i=0}^{N-1} \left(\mathcal{E}^{(i+1)}[x^{(i)}(t)] - \mathcal{E}^{(i)}[x^{(i)}(t)] \right). \quad (5)$$

By averaging the exponential of this effective work over many realizations of the transformation, we will show that a result identical in form to Jarzynski's applies to ensembles of trajectories.

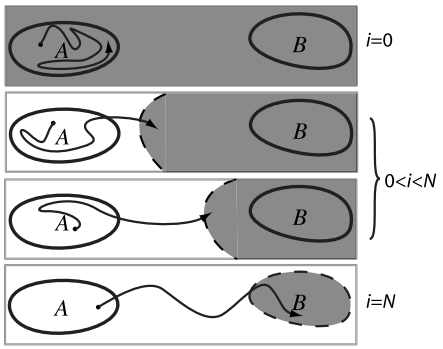

FIG. 2: Schematic depiction of one possible sequence of trajectories of the model system described in the main text. Dark regions are areas where the biasing potential $U(x_t, \lambda^{(i)})$ is not much greater than unity. At step $i = 0$, this region encompasses all of phase space, since the trajectory endpoint is unbiased. As i increases toward N, however, the region shrinks to include only state B.

A specific set of rules for sampling dynamical paths constitutes a "dynamics" in trajectory space. We emphasize that a system's natural dynamics, governing the time

evolution within individual trajectories, are not modified by this set of evolution rules, which determines a sequence of trajectories, such as that depicted in Fig. 2. Viewing the exploration of trajectory space as a random walk, we represent a particular set of rules using transition matrices, $\{\mathbf{M}^{(0)}, \mathbf{M}^{(1)}, \ldots, \mathbf{M}^{(N)}\}$, that conserve the distributions $\{P^{(0)}, P^{(1)}, \ldots, P^{(N)}\}$:

$$\int \mathcal{D}x(t) M^{(i)}[x(t) \to x'(t)] P^{(i)}[x(t)] = P^{(i)}[x'(t)]. \quad (6)$$

Matrix elements $M^{(i)}[x(t) \to x'(t)]$ express the probability that a walker situated at path $x(t)$ at step $i-1$ will reside at path $x'(t)$ at step i. In transforming between two trajectory ensembles, path distributions conserved by the initial and final steps should clearly describe the initial and final ensembles, respectively. Beyond this requirement, we will see, the choice of $\mathbf{M}^{(i)}$ at intermediate steps is arbitrary. Crooks has shown that the exact form of phase space dynamics is analogously arbitrary for Jarzynski's identity [18].

Given a set of N transition matrices, the probability of observing a particular sequence of trajectories, $\mathcal{X}(N,t) = \{x^{(0)}(t), x^{(1)}(t), \ldots, x^{(N)}(t)\}$, is:

$$\mathcal{P}[\mathcal{X}(N,t)] = P^{(0)}[x^{(0)}(t)] \prod_{i=1}^{N} M^{(i)}[x^{(i-1)}(t) \to x^{(i)}(t)]. \quad (7)$$

This weight determines averages of sequence-dependent quantities, $\mathcal{Y}[\mathcal{X}(N,t); \{\mathbf{M}^{(i)}\}]$, over many realizations of the random walk:

$$\overline{\mathcal{Y}} = \sum_{\mathcal{X}(N,t)} \mathcal{P}[\mathcal{X}(N,t)] \mathcal{Y}[\mathcal{X}(N,t); \{\mathbf{M}^{(i)}\}]. \quad (8)$$

The sum in Eq. (8) runs over all possible path sequences of length N. In general such an average depends on the choice of $\mathbf{M}^{(i)}$ and is therefore sensitive to the way in which ensembles are transformed.

Combining Eqs. (5), (6), (7), and 8, we obtain for the case $\mathcal{Y} = e^{-\mathcal{W}}$:

$$\overline{e^{-\mathcal{W}}} = \frac{\mathcal{Z}^{(N)}(t)}{\mathcal{Z}^{(0)}(t)} \int \mathcal{D}x^{(N)}(t) \int \mathcal{D}x^{(N-1)}(t) \frac{P^{(N)}[x^{(N-1)}(t)]}{P^{(N-1)}[x^{(N-1)}(t)]} M^{(N)}[x^{(N-1)}(t) \to x^{(N)}(t)] \cdots$$
$$\times \int \mathcal{D}x^{(1)}(t) \frac{P^{(2)}[x^{(1)}(t)]}{P^{(1)}[x^{(1)}(t)]} M^{(2)}[x^{(1)}(t) \to x^{(2)}(t)] \int \mathcal{D}x^{(0)}(t) P^{(1)}[x^{(0)}(t)] M^{(1)}[x^{(0)}(t) \to x^{(1)}(t)]. \quad (9)$$

Since $P^{(1)}[x^{(0)}(t)]$ is the stationary distribution corresponding to $\mathbf{M}^{(1)}$, the integral over $x^{(0)}(t)$ simply produces a factor $P^{(1)}[x^{(1)}(t)]$, which is cancelled by the denominator in the second line of Eq. (9). The integral over $x^{(1)}(t)$ then yields a factor $P^{(2)}[x^{(2)}(t)]$, which is similarly cancelled. The last of the N integrations on the right hand side of Eq. (9) is $\int \mathcal{D}x^{(N)}(t) P^{(N)}[x^{(N)}(t)] = 1$, leaving

$$\overline{e^{-\mathcal{W}}} = \frac{\mathcal{Z}^{(N)}(t)}{\mathcal{Z}^{(0)}(t)}. \quad (10)$$

The right hand side of Eq. (10), a ratio of trajectory ensemble partition functions, is the exponential of work required to *reversibly* switch between initial and final path ensembles. The left hand side is an average over *irreversible* transformations in a path sequence of finite length. Equation (10) is thus the dynamical analog of Jarzynski's identity. Remarkably, as in the case of Eq. (4), this result is independent of the procedure used to switch ensembles. By choosing $P^{(0)}[x(t)] = P_{\rm equ}[x(t)]$ and $P^{(N)}[x(t)] = P_{\rm equ}[x(t)]w[x(t)]/Z(t)$, one can in principle calculate the equilibrium average of any dynamical function $Z^{(N)}(t) = \int \mathcal{D}x(t) P_{\rm equ}[x(t)] w[x(t)] = \langle w[x(t)] \rangle$ in this way, including time correlation functions of an arbitrary number of observables [19].

We illustrate the practical utility of Eq. (10) by using it to calculate the kinetics of a model reaction, namely the isomerization of a diatomic solute in dense solvent [14, 20]. This two-dimensional model consists of a diatomic molecule immersed in a bath of purely repulsive soft particles. (See Ref. [14] for a detailed description of the model.) The diatomic molecule can reside in two states, A (for which $r < r_A$) and B (for which $r > r_B$), differing in intramolecular distance r and separated by a potential energy barrier from each other. If the barrier is high compared to $k_{\rm B}T$, transitions between the two states are rare. The relevant correlation function describing the kinetics of such isomerizations, $C(t)$ in Eq. (2), is obtained from Eq. (10) by selecting as initial and final path distributions $P^{(0)} = P_{\rm equ} h_A[x_0]/\mathcal{Z}_A(t)$ and $P^{(N)} = P_{\rm equ} h_A[x_0] h_B[x_t]/\mathcal{Z}_{AB}(t)$, respectively. To convert an ensemble of trajectories unconstrained at time t into an ensemble of reactive trajectories switching from isomer A to isomer B, we make the arbitrary choice for steps $0 < i < N$

$$P^{(i)}[x(t)] \equiv \frac{P_{\rm equ}[x(t)] h_A(x_0) \exp[-U(x_t, \lambda^{(i)})]}{Z^{(i)}(t)}, \quad (11)$$

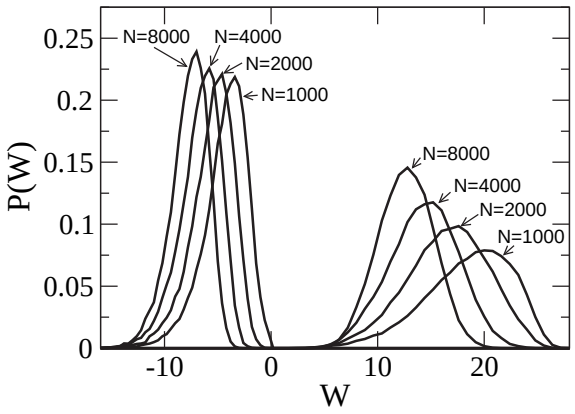

FIG. 3: Distributions of effective work $\mathcal{W}$ to transform an ensemble of trajectories beginning in isomer A to an ensemble switching from isomer A to isomer B. The distributions peaked at positive values of $\mathcal{W}$ correspond to path "compressions" in which the biasing potential was gradually applied (raising the drawbridge). The distributions peaked at negative value of $\mathcal{W}$, on the other hand, result from simulations in which the biasing potential was gradually released (lowering the drawbridge).

where $Z^{(i)}(t)$ is determined by normalization and

$$U(x,\lambda) \equiv \lambda \times [r_B - r(x)] \times (1 - h_B[x]) \qquad (12)$$

is an effective potential whose exponential interpolates smoothly between 1 (for $\lambda = 0$) and $h_B[x]$ (for $\lambda = \infty$). This "potential" $U(x,\lambda)$ acts as a drawbridge (anchored at $r(x) = r_B$) that, when lifted, pulls trajectory endpoints into B. Here, we let $\lambda^{(i)}$ vary linearly with i, i.e., $\lambda^{(i)} = \lambda_{\max} i/(N-1)$ for $i < N$ and $\lambda^{(N)} = \infty$. In all our simulations $\lambda_{\max} = 100$, and $r_A = 1.30\sigma$ and $r_B = 1.45\sigma$, where σ is the particle diameter.

To evaluate $C(t)$ we must finally determine rules for evolution in trajectory space that are consistent with the chosen bias functions. For this purpose we adapt the methods of transition path sampling, a Metropolis Monte Carlo sampling of trajectories [9, 21]. Specifically, we construct trial moves by the shooting and shifting algorithms described in Ref. [9], and accept them at step i with probability

$$P^{(i)}_{\rm acc}[x(t) \to x'(t)] = \min\left[1, \frac{P_{\rm equ}[x'(t)]w^{(i)}[x'(t)]}{P_{\rm equ}[x(t)]w^{(i)}[x(t)]}\right]. \qquad (13)$$

The corresponding transition matrices $\mathbf{M}^{(i)}$ conserve path distributions $P^{(i)}[x(t)]$ by construction.

Results for $C(t)$ computed in this way are given in Fig. 3. We have focused on a single value of t but have considered several values of N in the range $10^3 - 10^4$. Our numerical results bear out the validity of Eq. (10). Although the mean "work," $\overline{\mathcal{W}}$, is very different for different values of N, the estimates of $\Delta F(t) = -\ln(\overline{e^{-\mathcal{W}}})$ are indeed identical within statistical error of a few percent.

The computational effort expended in these calculations is comparable to that of calculating $C(t)$ (to similar accuracy) with the methods of transition path sampling alone (i.e., implementing a quasi-reversible change of trajectory ensemble). In light of applications of Jarzynski's identity [22–24], this observation is not surprising. On the one hand, these nonequilibrium relations provide computational savings by allowing fast switching between ensembles. On the other hand, the averages in Eqs. (4) and (10) can be slow to converge. Although the distributions of $\mathcal{W}$ plotted in Fig. 3 are not especially broad, the corresponding distributions of $e^{-\mathcal{W}}$ have substantial weight in the wings. As a result, a small fraction of path sequences can dominate the average. Significant effort is thus spent exploring regions of sequence space with little quantitative importance. One possible remedy for this disadvantage is to introduce a bias that preferentially guides sampling toward the most important path sequences. Sun has formulated an analogous approach for Jarzynski's identity [25], in effect employing transition path sampling to focus on the regions of trajectory space that dominate the average in Eq. (4). Applied to our result (Eq. (10)) such an approach may seem abstract, amounting to an importance sampling for sequences of trajectories that are themselves generated by the methods of transition path sampling. But it could well make possible important dynamical calculations (e.g., rates of folding for atomistic models of protein) that are currently impracticable.

PLG was an MIT Science Fellow during most of this work.

[1] J. B. Anderson, *J. Chem. Phys.* **58**, 4684 (1973); C. H. Bennett, in *Algorithms for Chemical Computations*, ed. Christoffersen R. E., pp. 63–97. Washington, D. C.: Amer. Chem. Soc. (1977); D. Chandler, *J. Chem. Phys.* **68**, 2959 (1978).
[2] R. Elber and M. Karplus, *Chem. Phys. Lett.* **139**, 375 (1987); R. Elber, J. Meller, and R. Olender, *J. Phys. Chem. B* **103**, 899 (1999).
[3] G. Henkelman, G. Johannesson, and H. Jónsson, in "Progress on Theoretical Chemistry and Physics", ed. S. D. Schwartz, Kluwar Academic Publishers (2000).
[4] J.P.K. Doye and D.J. Wales, *Z. Phys. D* **40**, 194 (1997).
[5] C. Dellago, P. G. Bolhuis, F. S. Csajka, and D. Chandler, *J. Chem. Phys.* **108**, 1964 (1998).
[6] W. E, W. Ren, and E. Vanden-Eijnden, *Phys. Rev. B* **66**, 052301 (2002).
[7] D. Chandler, "Introduction to Modern Statistical Mechanics", Oxford University Press (1987).
[8] P. G. Bolhuis, D. Chandler, C. Dellago, and P. L. Geissler, *Ann. Rev. Phys. Chem.* **53** 291 (2002).
[9] C. Dellago, P. G. Bolhuis, and P. L. Geissler, *Adv. Chem. Phys.* **123** 1 (2002).
[10] P. L. Geissler, C. Dellago, D. Chandler, J. Hutter, and M. Parrinello, *Science* **291** 2121 (2001).
[11] P. G. Bolhuis, C. Dellago, D. Chandler, *Proc. Natl. Acad.*

Sci. USA **97**, 5877 (2000).
[12] B. Ensing and E. J. Baerends, *J. Phys. Chem.* **106**, 7902 (2002).
[13] J. Rodriguez, G. Moreina, and D. Laria, *Chem. Phys. Lett.* **356**, 147 (2002).
[14] C. Dellago, P. G. Bolhuis and D. Chandler, *J. Chem. Phys.* **110**, 6617 (1999).
[15] T.J.H. Vlugt and B. Smit, *Phys. Chem. Comm.* **2**, 11 (2001).
[16] C. Dellago and P. L. Geissler, Proceedings of "The Monte Carlo Method in the Physical Sciences: Celebrating the 50th anniversary of the Metropolis algorithm", to be published (2003).
[17] C. Jarzynski, *Phys. Rev. Lett.* **78**, 2690 (1997).
[18] G. Crooks, *J. Stat. Phys.* **90**, 1481 (1998).
[19] For $P^{(N)}[x(t)]$ to be a meaningful probability density, we must have $w[x(t)] > 0$ for all $x(t)$. A generating function, $w = \exp(\phi A[x(t)])$ could be used to compute the average of a path functional $A[x(t)]$ with indefinite sign.
[20] J. E. Straub, M. Borkovec, and B. J. Berne, *J. Chem. Phys.* **89**, 4833 (1988).
[21] N. Metropolis, A. W. Metropolis, M. N. Rosenbluth, A. H. Teller, and E. Teller, *J. Chem. Phys.* **21**, 1087 (1953).
[22] G. Hummer and A. Szabo, *Proc. Natl. Acad. Sci. USA* **98**, 3658 (2001); J. Liphardt, S. Dumont, S. B. Smith, I. Tinoco, and C. Bustamante, *Science* **296**, 1832 (2002).
[23] D. A. Hendrix and C. Jarzynski, *J. Chem. Phys.* **114**, 5974 (2001).
[24] G. Hummer, *J. Chem. Phys.* **114**, 7330 (2001).
[25] Sean X. Sun, *J. Chem. Phys.* **118**, 5769 (2003)