Journal of Statistical Physics, Vol. 90, Nos. 5/6, 1998

Nonequilibrium Measurements of Free Energy Differences for Microscopically Reversible Markovian Systems

Gavin E. Crooks[1]

Received October 20, 1997; final December 16, 1997

An equality has recently been shown relating the free energy difference between two equilibrium ensembles of a system and an ensemble average of the work required to switch between these two configurations. In the present paper it is shown that this result can be derived under the assumption that the system's dynamics is Markovian and microscopically reversible.

KEY WORDS: Nonequilibrium statistical mechanics; free energy; work; thermodynamic integration; thermodynamic perturbation.

1. INTRODUCTION

Consider a classical system in contact with a constant temperature heat bath where some degree of freedom of the system can be controlled. Manipulation of this degree of freedom results in an expenditure of some amount of work, a change in the energy and free energy of the system, and an exchange of heat with the bath. Let λ be a parameter specifying the current value of the controllable degree of freedom. We consider a process where λ is switched between an initial and final value over some finite length of time. It has recently been shown(1, 2) that the free energy difference, ΔF, between the equilibrium ensembles corresponding to the initial and final values of λ can be related to an average of the amount of work, W, expended in this switching process. Specifically

$$\overline{e^{-\beta W}} = e^{-\beta \Delta F} \tag{1}$$

---

[1] Department of Chemistry, University of California, Berkeley, California 94720; e-mail: gavinc@garnet.berkeley.edu.

0022-4715/98/0300-1481$15.00/0

where $\beta = 1/k_B T$, k_B is the Boltzmann constant, and T is the temperature of the heat bath. The overbar indicates an ensemble average over all possible paths through phase space, given an equilibrium initial state, and the value of the control parameter at all times during the switching process. This can be approximated by an average over many measurements of the work required for the same switching process. Note that this is a non-equilibrium measurements. The system is not assumed to be in equilibrium during the switching process.

This equality is a generalization(1, 2) of several relations that are commonly used to calculated free energy differences in computer simulations. In the limit of an infinitely long switching process Eq. (1) becomes equivalent to thermodynamic integration.(3, 4) This assumes that the system is always in equilibrium, that the switching process is reversible, and therefore that no energy is dissipated. It follows that $\Delta F = W$. The finite time of actual simulations leads to systematic overestimation of the free energy difference.

In the limit of infinitely fast switching Eq. (1) is equivalent to

$$\langle e^{-\beta W} \rangle_0 = e^{-\beta \Delta F} \tag{2}$$

The angled brackets indicate an equilibrium average with λ fixed in its initial value. W is the work required to instantaneously change the control parameter to the new value while keeping the system configuration fixed. Thermodynamic perturbation(3–5) uses this relation to evaluate free energy differences. To reduce statistical error the control parameter is changed in a series of small steps. At each step the system is allowed to equilibrate before evaluating the work required to switch to the next step. Again this leads to systematic errors due to the finite simulation time available to relax the system to equilibrium at each step.

Both thermodynamic integration and perturbation suffer from systematic errors because the simulated systems are never truly in equilibrium. Calculations based on Eq. (1) only require that the initial ensemble is in equilibrium, which is computationally feasible. Although this results in a reduction of systematic errors it increases the statistical errors,(1, 2) which may limit its applicability.

Previously this relation has been derived for a Hamiltonian system weakly coupled to a heat bath,(1) and based on a master equation approach.(2) In this paper it will be shown that this equality directly follows if we assume that the dynamics of the system are Markovian, and microscopically reversible. The Markovian(6) condition ensures that the system is memory less. The property of microscopic reversibility(7, 8) (principle of detailed balance) ensures that the system is time reversible and that the equilibrium probability distributions are correctly given by the canonical ensemble. These two conditions are both met by many computer simulations.

2. NOTATION AND ASSUMPTIONS

The system of interest is held at a constant temperature, T. The state of the system is specified by two parameters. The internal state of the system at time t is labeled by i_t. In addition the energy of each state is affected by an externally controlled parameter, λ_t. The energy of the system at time t is denoted by $E(i_t, \lambda_t)$.

To give an example, we could consider a classical gas confined in a cylinder. Then i_t is a vector specifying the instantaneous momenta and positions of all the particles. The parameter, λ_t can control the volume of the gas, manipulated by means of a piston. The time evolution of the system could be governed by Langevin dynamics, where a frictional and a stochastic force are added to the deterministic classical mechanical equations of motion.

As an alternative example consider the Ising model of a ferromagnet. The parameter, i_t is an index specifying the current position (up or down) of all the magnetic spins. Work is done on the system by applying an external magnetic field, whose value is specified by λ_t. The time evolution of this system is then governed by single spin flip Metropolis Monte Carlo.(9)

For notational convenience we will assume discrete time and a discrete phase space. Both of these conditions are necessarily true for any simulation on a digital computer. The results can be readily generalized to continuous time and continuous phase space.

In a canonical ensemble the equilibrium probability of a state, A, given a fixed value of the control parameter, λ, is

$$\mathrm{P}(A \mid \lambda) = \frac{e^{-\beta E(A,\lambda)}}{\sum_i e^{-\beta E(i,\lambda)}} = \exp[+\beta F(\beta, \lambda) - \beta E(A, \lambda)] \tag{3}$$

The sum is over all states of the system and $F(\beta, \lambda) = -\beta^{-1} \ln \sum_i e^{-\beta E(i,\lambda)}$ is the Helmholtz free energy of the system.

We wish to consider the evolution of this system through time, as the control parameter is moved through a fixed sequence, $\{\lambda_0, \lambda_1, ..., \lambda_\tau\}$. A particular path through phase space can be written as

$$i_0 \xrightarrow{\lambda_1} i_1 \xrightarrow{\lambda_2} i_2 \xrightarrow{\lambda_3} \cdots \xrightarrow{\lambda_\tau} i_\tau \tag{4}$$

At time $t = 0$ the system is in state i_0 and the control parameter is λ_0. The time evolution of the system is considered to occur in two substeps. First the control parameter is moved to a new value, λ_1. This takes an amount of work, $E(i_0, \lambda_1) - E(i_0, \lambda_0)$. Then the state of the system evolves, at constant λ_1, to state i_1. During this evolution the system exchanges a quantity $E(i_1, \lambda_1) - E(i_0, \lambda_1)$ of heat with the reservoir. This evolution through phase space is repeated for τ time steps.

The total work performed on the system, W, the total heat exchanged with the reservoir, Q, and the total change in energy, ΔE, are given by

$$W = \sum_{t=0}^{\tau-1} [E(i_t, \lambda_{t+1}) - E(i_t, \lambda_t)] \tag{5}$$

$$Q = \sum_{t=1}^{\tau} [E(i_t, \lambda_t) - E(i_{t-1}, \lambda_t)], \qquad \text{and} \tag{6}$$

$$\Delta E = Q + W = [E(i_\tau, \lambda_\tau) - E(i_0, \lambda_0)] \tag{7}$$

The reversible work, $W_r = \Delta F = F(\beta, \lambda_1) - F(\beta, \lambda_0)$ is the free energy difference between two equilibrium ensembles. The dissipative work, $W_d = W - W_r$ is defined as the difference between the reversible work and the actual work. If an amount of work is expended in changing the free energy of the system then the change in entropy of the universe is βW_d, in units of Boltzmann's constant. Note that the work and the dissipative work depend on the path followed through phase space, but the reversible work depends only on the initial and final ensembles.

It will prove useful to consider reversing the direction of time. The reverse time path through phase space, corresponding to the forward time path specified in Eq. (4), can be written as $(i_0 \xleftarrow{\lambda_1} i_1 \xleftarrow{\lambda_2} i_2 \xleftarrow{\lambda_3} \cdots \xleftarrow{\lambda_\tau} i_\tau)$. The sequence in which states are visited is reversed, as is the order in which λ is changed. Note that the forward path begins with a change in λ, whereas the reverse path begins with a change in the internal sate of the system. The work, heat, change in energy and change in free energy are defined above in the forward time direction. For the reversed time direction these quantities would be the negative of the forward time value. Where necessary the dependence on the time direction and path can be made explicit. Elsewhere it is to be understood that these quantities are defined in the forward time direction.

If it is assumed that the evolution of the system is Markovian then the probability of making a transition between two states, $\mathrm{P}(i_t \xrightarrow{\lambda} i_{t+1})$, depends only on the state of the system at time t, and not on the previous history of the system. Thus the probability of following a path through phase space, $\mathrm{P}(i_0 \xrightarrow{\lambda_1} i_1 \xrightarrow{\lambda_2} i_2 \xrightarrow{\lambda_3} \cdots \xrightarrow{\lambda_\tau} i_\tau)$, (given the initial state,

i_0, and the control parameter at all times, $\{\lambda_1, \lambda_2,..., \lambda_\tau\}$) can be split into its constituent single time step parts.

$$\begin{aligned}
&\mathrm{P}(i_0 \xrightarrow{\lambda_1} i_1 \xrightarrow{\lambda_2} i_2 \cdots \xrightarrow{\lambda_\tau} i_\tau) \\
&\quad = \mathrm{P}(i_0 \xrightarrow{\lambda_1} i_1)\, \mathrm{P}(i_1 \xrightarrow{\lambda_2} i_2) \cdots \mathrm{P}(i_{\tau-1} \xrightarrow{\lambda_\tau} i_\tau)
\end{aligned}$$

The single time steps are assumed to be microscopically reversible,(7, 8) and therefore obey the detailed balance condition for all fixed values of the external control parameter, λ.

$$\frac{\mathrm{P}(A \xrightarrow{\lambda} B)}{\mathrm{P}(A \xleftarrow{\lambda} B)} = \frac{\mathrm{P}(B \mid \lambda)}{\mathrm{P}(A \mid \lambda)} = \frac{e^{-\beta E(B, \lambda)}}{e^{-\beta E(A, \lambda)}} \tag{8}$$

We will now prove an analogous detailed balance condition for a multiple time step process, where we perform an arbitrary amount of work on the system. Given that the process is Markovian, and that individual steps obey detailed balance, then the ratio of probabilities of a forward path, versus the corresponding time reversed path is

$$\begin{aligned}
&\frac{\mathrm{P}(i_0 \xrightarrow{\lambda_1} i_1 \xrightarrow{\lambda_2} i_2 \cdots \xrightarrow{\lambda_\tau} i_\tau)}{\mathrm{P}(i_0 \xleftarrow{\lambda_1} i_1 \xleftarrow{\lambda_2} i_2 \cdots \xleftarrow{\lambda_\tau} i_\tau)} \\
&\quad = \frac{\mathrm{P}(i_0 \xrightarrow{\lambda_1} i_1)\, \mathrm{P}(i_1 \xrightarrow{\lambda_2} i_2) \cdots \mathrm{P}(i_{\tau-1} \xrightarrow{\lambda_\tau} i_\tau)}{\mathrm{P}(i_0 \xleftarrow{\lambda_1} i_1)\, \mathrm{P}(i_1 \xleftarrow{\lambda_2} i_2) \cdots \mathrm{P}(i_{\tau-1} \xleftarrow{\lambda_\tau} i_\tau)} \\
&\quad = \frac{e^{-\beta E(i_1, \lambda_1)} e^{-\beta E(i_2, \lambda_2)} \cdots e^{-\beta E(i_\tau, \lambda_\tau)}}{e^{-\beta E(i_0, \lambda_1)} e^{-\beta E(i_1, \lambda_2)} \cdots e^{-\beta E(i_{\tau-1}, \lambda_\tau)}} \\
&\quad = e^{-\beta Q}
\end{aligned} \tag{9}$$

Here Q is the energy exchanged with the heat bath as the system moves along the forward path (Eq. (7)), and $-\beta Q$ is the corresponding change in entropy (in units of Boltzmann's constant) of the bath.

Essentially this equation is a statement that detailed balance continues to hold for Markovian microscopically reversible systems irrespective of how much work is actually performed on the system.

If we also specify that both the forward and reverse paths start from equilibrium distributions, then we find that

$$\begin{aligned}
\frac{\mathrm{P}(i_0 \mid \lambda_0)\, \mathrm{P}(i_0 \xrightarrow{\lambda_1} i_1 \cdots \xrightarrow{\lambda_\tau} i_\tau)}{\mathrm{P}(i_\tau \mid \lambda_\tau)\, \mathrm{P}(i_0 \xleftarrow{\lambda_1} i_1 \cdots \xleftarrow{\lambda_\tau} i_\tau)} &= e^{+\beta \Delta E - \beta \Delta F} e^{-\beta Q} \\
&= e^{+\beta W_d}
\end{aligned} \tag{10}$$

3. DERIVATION

These consequences of detailed balance and the Markov condition will allow a simple proof that $\overline{e^{-\beta W}} = e^{-\beta \Delta F}$ (Eq. (1)). The overbar indicates an average over all paths thorough phase space, given a fixed sequence of the external control parameter, and that the system starts from a canonical equilibrium distribution.

$$\overline{e^{-\beta W}} = \sum_{i_0, i_1, \ldots, i_\tau} \mathrm{P}(i_0 \mid \lambda_0)\, \mathrm{P}(i_0 \xrightarrow{\lambda_1} i_1 \cdots \xrightarrow{\lambda_\tau} i_\tau)\, e^{-\beta W} \tag{11}$$

This average over the forward time path can be changed to an average over the reversed time path using Eq. (10). Therefore

$$\begin{aligned} \overline{e^{-\beta W}} &= \sum_{i_0, i_1, \ldots, i_\tau} \mathrm{P}(i_\tau \mid \lambda_\tau)\, \mathrm{P}(i_0 \xleftarrow{\lambda_1} i_1 \cdots \xleftarrow{\lambda_\tau} i_\tau)\, e^{+\beta W_d - \beta W} \\ &= e^{-\beta \Delta F} \end{aligned} \tag{12}$$

The last step follows because the reversible work $(\Delta F = W_r = W - W_d)$ is path independent, and because probabilities are normalized.

This result readily generalizes to other isothermal ensembles. For example consider a classical gas, held at constant temperature and pressure, p. Work is performed by gradually inserting an additional particle into the system, via a slow growth process.(10) In this isothermal-isobaric ensemble the change in Gibbs free energy, $\Delta G = G(\beta, p, \lambda_\tau) - G(\beta, p, \lambda_0)$, is equal to the excess chemical potential of the inserted particle at that temperature and pressure.

The detailed balance condition is now

$$\frac{\mathrm{P}(A \xrightarrow{\lambda} B)}{\mathrm{P}(A \xleftarrow{\lambda} B)} = \frac{e^{-\beta E(B, \lambda) - \beta p V(B, \lambda)}}{e^{-\beta E(A, \lambda) - \beta p V(B, \lambda)}} = e^{-\beta Q - \beta p \Delta V} \tag{13}$$

where V is the volume, and p is the pressure. The change in volume, ΔV can be considered a baric equivalent of the thermal heat.(11) This condition continues to hold for arbitrary paths through phase space, irrespective of the amount of work performed. The ratio of equilibrium probabilities in this ensemble is $\mathrm{P}(i_\tau \mid \lambda_\tau)/\mathrm{P}(i_0 \mid \lambda_0) = \exp(\beta \Delta E + \beta p \Delta V - \beta \Delta G)$. Thus Eq. (10) is unmodified, and holds both in the canonical, and in the isothermal-isobaric ensembles. It immediately follows that $\overline{e^{-\beta W}} = e^{-\beta \Delta G}$.

4. SUMMARY

The free energy differences of isothermal systems can be directly related to a nonequilibrium exponential average of the work require to switch between the ensembles (Eq. (1)). This can be easily proved (Eqs. (11) and (12)), based on the assumptions that the system in question is Markovian and microscopically reversible (Eqs. (8)–(10)). Thermodynamic integration and perturbation appear as limiting cases of this relation. Explicit use of Eq. (1) to evaluate free energy differences should result in lower systematic error, but its applicability may be limited by an increase in statistical error.

ACKNOWLEDGMENTS

I am indebted to C. Jarzynski for his stimulating correspondence, and D. Chandler for his advice and guidance. This work was supported by NSF grant number 441230-22420.

REFERENCES

1. C. Jarzynski, Nonequilibrium equality for free energy differences, *Phys. Rev. Lett.* **78**(14):2690–2693 (1997).
2. C. Jarzynski, Equilibrium free energy differences from nonequilibrium measurements: A master equation approach, *Phys. Rev. E.* **56**(5):5018–5035 (1997).
3. D. L. Beveridge and F. M. DiCapua, Free energy via molecular simulation: Applications to chemical and biomolecular systems, *Annu. Rev. Biophys. Biophys. Chem.* **18**:431–92 (1989).
4. T. P. Straatsma and J. A. McCammon, Computational alchemy, *Annu. Rev. Phys. Chem.* **43**:407–435 (1992).
5. R. W. Zwanzig, High-temperature equation of state by a perturbation method: I. nonpolar gases, *J. Chem. Phys.* **22**:1420–1426 (1954).
6. G .R. Grimmett and D. R. Stirzaker, *Probability and Random Processes* (Clarendon Press, Oxford, 2nd edition, 1992).
7. S. R. de Groot and P. Mazur, *Nonequilibrium Thermodynamics* (North-Holland, Amsterdam, 1962).
8. D. Chandler, *Introduction to Modern Statistical Mechanics* (Oxford University Press, New York, 1987), pp. 165.
9. N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, Equation of state calculations by fast computing machines, *J. Chem. Phys.* **21**:1087–1092 (1953).
10. K. K. Mon and R. B. Griffiths, Chemical potential by gradual insertion of a particle in monte carlo simulation, *Phys. Rev. A.* **31**:956–959 (1985).
11. E. T. Jaynes, Information theory and statistical mechanics, *Phys. Rev.* **106**:620–630(1957).

Communicated by J. L. Lebowitz