Nonequilibrium Equality for Free Energy Differences

C. Jarzynski*
Institute for Nuclear Theory, University of Washington, Seattle, Washington 98195
(Received 7 June 1996)

An expression is derived for the equilibrium free energy difference between two configurations of a system, in terms of an ensemble of *finite-time* measurements of the work performed in parametrically switching from one configuration to the other. Two well-known identities emerge as limiting cases of this result. [S0031-9007(97)02845-7]

PACS numbers: 05.70.Ln, 87.10.+e, 82.20.Wt

Consider a finite classical system in contact with a heat reservoir. A central concept in thermodynamics is that of the *work* performed on such a system, when some external parameters of the system are made to change with time. (These parameters may represent, for instance, the strength of an external field, or the volume of space within which the system is confined, or, more abstractly, some particle-particle interactions which are turned on or off during the course of a molecular dynamics simulation.) When the parameters are changed *infinitely slowly* along some path γ from an initial point A to a final point B in parameter space, then the total work W performed on the system is equal to the Helmholtz free energy difference ΔF between the initial and final configurations [1]: $W = \Delta F \equiv F^B - F^A$. [Here F^A (F^B) refers to the equilibrium free energy of the system, with the parameters held fixed at A (B).] By contrast, when the parameters are switched along γ at a *finite* rate, then W will depend on the microscopic initial conditions of the system and reservoir, and will, on average, exceed ΔF:

$$\overline{W} \geq \Delta F\,. \tag{1}$$

The overbar denotes an average over an *ensemble* of measurements of W, where each measurement is made after first allowing the system and reservoir to equilibrate at temperature T, with the parameters fixed at A. (The path γ from A to B, and the rate at which the parameters are switched along this path, remain unchanged from one measurement to the next.) Note that the right side of Eq. (1) still refers to the *equilibrium* free energy difference $F^B - F^A$. The difference $\overline{W} - \Delta F$ is just the dissipated work, $W_{\rm diss}$, associated with the increase of entropy during an irreversible process.

Equation (1) is an inequality. By contrast, the new result derived in this paper is the following *equality*:

$$\overline{\exp(-\beta W)} = \exp(-\beta \Delta F)\,, \tag{2a}$$

or, equivalently,

$$\Delta F = -\beta^{-1} \ln \overline{\exp(-\beta W)}\,, \tag{2b}$$

where $\beta \equiv 1/k_B T$. This result, which is independent of both the path γ from A to B, and the rate at which the parameters are switched along the path, is surprising: It says that we can extract equilibrium information (ΔF) from the ensemble of *nonequilibrium* (finite-time) measurements described above.

Before proceeding with the proof of Eq. (2), we establish notation and then relate Eq. (2) to two well-known equilibrium identities for ΔF. Since we have fixed our attention on a particular path γ in parameter space, it will be convenient to henceforth view the system as parametrized by a single quantity λ, which increases from 0 to 1 as we travel from A to B along γ. Let $\mathbf{z} \equiv (\mathbf{q}, \mathbf{p})$ denote a point in the phase space of the system, and let $H_\lambda(\mathbf{z})$ denote the Hamiltonian for the system, parametrized by the value of λ. Next, let Z_λ denote the partition function, let $\langle \cdots \rangle_\lambda$ denote a canonical average, and let $F_\lambda = -\beta^{-1} \ln Z_\lambda$ denote the free energy, all with respect to the Hamiltonian H_λ and the temperature T. We are interested in the following scenario, which we will refer to as "the switching process": The system evolves, in contact with a heat reservoir, as the value of λ is switched from 0 to 1, over a total switching time t_s. Without loss of generality, assume a constant switching rate, $\dot{\lambda} = t_s^{-1}$. For a given realization of the switching process, the evolution of the system is described by a (effectively stochastic) trajectory $\mathbf{z}(t)$, and the work performed on the system is the time integral of $\dot{\lambda}\partial H_\lambda/\partial\lambda$ along this trajectory:

$$W = \int_0^{t_s} dt\, \dot{\lambda}\, \frac{\partial H_\lambda}{\partial \lambda}\, (\mathbf{z}(t))\,. \tag{3}$$

Now imagine an *ensemble* of realizations of the switching process (with γ and t_s fixed), with initial conditions for the system and reservoir generated from a canonical ensemble at temperature T. Then W may be computed for each trajectory $\mathbf{z}(t)$ in the ensemble, and the overbars appearing in Eqs. (1) and (2) indicate an average over the distribution of values of W thus obtained.

In the limiting cases of infinitely slow and infinitely fast switching of the external parameters, we know explicitly the ensemble distribution of values of W, and thus can readily check the validity of our central result. In the slow limit ($t_s \to \infty$), the system is in quasistatic equilibrium with the reservoir throughout the switching process, hence $W = \int_0^1 d\lambda \langle \partial H_\lambda/\partial\lambda \rangle_\lambda$ for every trajectory in the

