8.5: Jarzynski's Equality and Nonequilibrium Methods

In this section, the relationship between work and free energy will be explored in greater detail. We have already introduced the inequality in Equation ???, which states that if an amount of work $W_{\mathcal{AB}}$ is performed on a system, taking from state $\mathcal{A}$ to state $\mathcal{B}$, then $W_{\mathcal{AB}} \geq A_{\mathcal{AB}}$. Here, equality holds only if the work is performed reversibly. The work referred to here is thermodynamic quantity and, as such, must be regarded as an ensemble average. In statistical mechanics, we can also introduce the mechanical or microscopic work $\mathcal{W}_{\mathcal{AB}}(\mathrm{x})$ performed on one member of the ensemble to drive it from state $\mathcal{A}$ to state $\mathcal{B}$. Then, $W_{\mathcal{AB}}$ is simply an ensemble average of $\mathcal{W}_{\mathcal{AB}}$. However, we need to be somewhat careful about how we define this ensemble average because the work is defined along a particular path or trajectory which takes the system from state $\mathcal{A}$ to state $\mathcal{B}$, and equilibrium averages do not refer not to paths but to microstates. This distinction is emphasized by the fact that the work could be carried out irreversibly, such that the system is driven out of equilibrium. Thus, the proper definition of the ensemble average follows along the lines already discussed in the context of the free-energy perturbation approach, namely, averaging over the canonical distribution for the state $\mathcal{A}$. In this case, since we will be discussing actual paths $\mathrm{x_t}$, we let the initial condition x_0 be the phase space vector for the system in the (initial) state $\mathcal{A}$. Recall that $x_t = x_t(x_0)$ is a unique function of the initial conditions. Then

$$W_{\mathcal{AB}} = \langle \mathcal{W}_{\mathcal{AB}}(\mathrm{x}_0)\rangle_{\mathcal{A}} \tag{8.5.1}$$

$$= \frac{C_N}{Q_{\mathcal{A}}(N,V,T)} \int d\mathrm{x}_0 \mathrm{e}^{-\beta \mathrm{H}_{\mathcal{A}}(\mathrm{x}_0)} \mathcal{W}_{\mathcal{AB}}(\mathrm{x}_0) \tag{8.5.2}$$

and the Clausius inequality can be stated as $\langle \mathcal{W}_{\mathcal{AB}}(\mathrm{x}_0)\rangle_{\mathcal{A}} \geq A_{\mathcal{AB}}$.

From such an inequality, it would seem that using the work as a method for calculating the free energy is of limited utility, since the work necessarily must be performed reversibly, otherwise one obtains only upper bound on the free energy. It turns out, however, that irreversible work can be used to calculate free energy differences by virtue of a connection between the two quantities first discovered in 1997 by C. Jarzynski that as come to be known as the **Jarzynski equality**. This equality states that if, instead of averaging $\mathcal{W}_{\mathcal{AB}}(\mathrm{x}_0)$ over the initial canonical distribution (that of state $\mathcal{A}$), an average of $\exp[-\beta \mathcal{W}_{\mathcal{AB}}(\mathrm{x}_0)]$ is performed over the same distribution, the result is $\exp[-\beta A_{\mathcal{AB}}]$, i.e.

$$e^{-\beta A_{\mathcal{AB}}} = \langle e^{-\beta \mathcal{W}_{\mathcal{AB}}(\mathrm{x}_0)}\rangle_{\mathcal{A}} \tag{8.5.3}$$

$$= \frac{C_N}{Q_{\mathcal{A}}(N,V,T)} \int d\mathrm{x}_0 \mathrm{e}^{-\beta \mathrm{H}_{\mathcal{A}}(\mathrm{x}_0)} \mathrm{e}^{-\beta \mathcal{W}_{\mathcal{AB}}(\mathrm{x}_0)} \tag{8.5.4}$$

This remarkable result not only provides a foundation for the development of nonequilibrium free-energy methods but also has profound implications for thermodynamics, in general.

The Jarzynski equality be proved using different strategies. Here, however, we will present a proof that is most relevant for the finite-sized systems and techniques employed in molecular dynamics calculations. Consider a time-dependent Hamiltonian of the form

$$H(\mathbf{p}, \mathbf{r}, t) = \sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m_i} + U(\mathbf{r}_1, \ldots, \mathbf{r}_N, t) \tag{8.5.5}$$

For time-dependent Hamiltonian's, the usual conservation law $dH/dt = 0$ no longer holds, which can be seen by computing

$$\frac{dH}{dt} = \nabla_{\mathrm{x}_t} H \dot{\mathrm{x}}_t + \frac{\partial H}{\partial t} \tag{8.5.6}$$

where the phase space vector $\mathrm{x} = (\mathbf{p}_1, \ldots, \mathbf{p}_N, \mathbf{r}_1, \ldots, \mathbf{r}_N) \equiv (\mathbf{p}, \mathbf{r})$ has been introduced. Integrating both sides over time from $t = 0$ to a final time $t = \tau$, we find

$$\int_0^\tau dt \, \frac{dH}{dt} = \int_0^\tau dt \, \nabla_{\mathrm{x}_t} H \mathrm{x}_t + \int_0^\tau dt \, \frac{\partial H}{\partial t} \tag{8.5.7}$$

Equation 8.5.7 can be regarded as a microscopic version of the first law of thermodynamics, in which the first and second terms represent the heat absorbed by the system and the work done on the system over the trajectory, respectively. Note that the work is actually a function of the initial phase-space vector x_0, which can be seen by writing this term explicitly as

$$W_\tau(\mathrm{x}_0) = \int_0^\tau dt\ \frac{\partial}{\partial t} H(\mathrm{x}_t(\mathrm{x}_0), t) \tag{8.5.8}$$

where the fact that the work depends explicitly on τ in Equation 8.5.8 is indicated by the subscript. In the present discussion, we will consider that each initial condition, selected from a canonical distribution in x_0, evolves according to Hamilton's equations in isolation. In this case, the heat term $\nabla_{\mathrm{x}_t} H \cdot \mathrm{x}_t = 0$, and we have the usual addition to Hamilton's equations $dH/dt = \partial H/\partial t$.

With the above condition, we can write the microscopic work as

$$\mathcal{W}_{\mathcal{AB}} = \int_0^\tau \frac{d}{dt} H(\mathrm{x}_t(\mathrm{x}_0), \mathrm{t})\mathrm{dt} = \mathrm{H}(\mathrm{x}_\tau(\mathrm{x}_0), \tau) - \mathrm{H}(\mathrm{x}_0, 0) \tag{8.5.9}$$

The last term $H(\mathrm{x}_0, 0)$ is also $H_{\mathcal{A}}(\mathrm{x}_0)$. Thus, the ensemble average of the exponential of the work becomes

$$\langle e^{-\beta \mathcal{W}_{\mathcal{AB}}} \rangle_{\mathcal{A}} = \frac{C_N}{Q_{\mathcal{A}}(N,V,T)} \int d\mathrm{x}_0\ e^{-\beta H_{\mathcal{A}}(\mathrm{x}_0)} e^{-\beta[H(\mathrm{x}_\tau(\mathrm{x}_0),\tau) - H_{\mathcal{A}}(\mathrm{x}_0)]} \tag{8.5.10}$$

$$\frac{C_N}{Q_{\mathcal{A}}(N,V,T)} \int d\mathrm{x}_0\ e^{-\beta H(\mathrm{x}_\tau(\mathrm{x}_0),\tau)} \tag{8.5.11}$$

The numerator in this expression becomes much more interesting if we perform a change of variables from x_0 to x_τ. Since the solution of Hamilton's equations for the time-dependent Hamiltonian uniquely map the initial condition x_0 onto x_t, when $t = \tau$, we have a new set of phase-space variables, and by Liouville's theorem, the phase-space volume element is preserved

$$d\mathrm{x}_\tau = d\mathrm{x}_0 \tag{8.5.12}$$

When the Hamiltonian is transformed, we find $H(\mathrm{x}_\tau, \tau) = H_{\mathcal{B}}(\mathrm{x}_\tau)$. Consequently,

$$\langle e^{-\beta \mathcal{W}_{\mathcal{AB}}} \rangle_{\mathcal{A}} = \frac{C_N}{Q(N,V,T)} \int d\mathrm{x}_\tau\ e^{-\beta H_{\mathcal{B}}(\mathrm{x}_\tau)} \tag{8.5.13}$$

$$= \frac{Q_{\mathcal{B}}(N,V,T)}{Q_{\mathcal{A}}(N,V,T)} \tag{8.5.14}$$

$$= e^{-\beta A_{\mathcal{AB}}} \tag{8.5.15}$$

thus proving the equality. The implication of the Jarzynski equality is that the work can be carried out along a reversible or irreversible path, and the correct free energy will still be obtained.

Note that due to Jensen's inequality:

$$\langle e^{-\beta \mathcal{W}_{\mathcal{AB}}} \rangle_{\mathcal{A}} \geq e^{-\beta \langle \mathcal{W}_{\mathcal{AB}} \rangle_{\mathcal{A}}} \tag{8.5.16}$$

Using Jarzynski's equality, this becomes

$$e^{-\beta A_{\mathcal{AB}}} \geq e^{-\beta \langle \mathcal{W}_{\mathcal{AB}} \rangle_{\mathcal{A}}} \tag{8.5.17}$$

which implies, as expected, that

$$A_{\mathcal{AB}} \leq \langle \mathcal{W}_{\mathcal{AB}} \rangle_{\mathcal{A}} \tag{8.5.18}$$