Nonequilibrium Work Relations and Equilibrium Thermodynamics

"When one gets to the molecules the distinction between heat and work disappears because both are [ultimately molecular] energy." J.C. Maxwell, Nature, **17**, 257(1878)

In previous chapters devoted to the Fluctuation, Dissipation and Relaxation Theorems once we set the dynamics running at the initial time, or perhaps at time 0^+ , at no stage did we change the state of the underlying equilibrium thermodynamic state. If at any time we ceased doing work on a driven system, the system would relax back to the thermodynamic equilibrium state we started from at t = 0. From $t = 0^+$ onwards the equilibrium state for the system with the dissipative field set to zero, did not change with time. In equilibrium state specified by the zero field dynamics is unchanging and for T-mixing systems, is unique. In driven systems the system starts in equilibrium, but is driven out of equilibrium by the dissipative field. If the dissipative field is subsequently set to zero, the system will return to the *initial* equilibrium state.

The clearest indication that the dissipation is purely dissipative is evidenced in the Nonequilibrium Partition Identity, (§3.3). If the integrated dissipation contained a reversible component, ΔW_{rev} then the left hand side of the NPI, (3.3.1), would have to equal exp $[-\Delta W_{rev}]$ rather than unity.

Another feature of the dynamics studied previously was that the equations of motion, at least after $t = 0^+$, were autonomous. In the present chapter we will discuss nonautonomous systems in which either the Hamiltonian or some thermodynamic state variables change non-autonomously during the dynamics.

1

In classical thermodynamics, free energy differences between two equilibrium states are determined using the work carried out along a quasistatic (*i.e.* equilibrium, reversible) pathway connecting the two equilibrium states. Of course in classical thermodynamics one only performs these measurements in the so-called thermodynamic limit where all intensive thermodynamic quantities are independent of the system size.

In 1997 Jarzynski discovered the first of a set of new fluctuation relations that used nonequilibrium path integrals, measured along an ensemble of *nonequilibrium* pathways to provide *equilibrium* thermodynamic information about *small* systems. In order for these approaches to be useful the system size *must* be small because the methods rely on fluctuations and the observation of phase space trajectories that are the time reversed conjugate trajectories to the most probable trajectories (*i.e.* they are the *most* improbable trajectories possible, within the specified phase space domain). Although this requirement for small system size may be seen as something of a disadvantage, it turns out to be an *essential* advantage for studying the thermodynamics of small nano-systems, something the classical thermodynamics could not do.

The fundamental reason why the Jarzynski Equality works is because if we write the nonequilibrium work as a sum of the purely reversible thermodynamic work and the purely irreversible work, the irreversible work satisfies a Nonequilibrium Partition Identity leading directly to the Jarzynski Equality. We show this in detail in this chapter – see (8.6.9).

The Crooks Fluctuation Theorem (CFT) and the Jarzynski Equality (JE) were originally developed for determining the difference in free energy of canonical equilibrium states at the same temperature, however we present a very general formalism for deriving nonequilibrium free energy relations that can be applied in a very wide variety of circumstances.

Later in the chapter we show how to derive Clausius' Theorem directly from mechanical considerations. In 1854 Clausius proved his Theorem by *assuming* the Second "Law" of Thermodynamics. In fact Clausius' statement of the Second "Law" is perhaps the most commonly used form of the Second "Law". The fact that we can now prove this from the laws of mechanics, completely changes the logical structure of statistical thermodynamics. The "Laws" of thermodynamics in fact cease being

2

laws and instead become theorems provable from the laws of mechanics. No longer is thermodynamics in apparent contradiction to time reversible microscopic dynamics.

8.1 GENERALIZED CROOKS FLUCTUATION THEOREM (GCFT)

We consider two closed *N*-particle systems: 1, 2. These systems may have the same or different Hamiltonians, temperatures or volumes; it does not matter. Nor does the ensemble matter: microcanonical, canonical or isothermal isobaric. A protocol, and the corresponding time-dependent dynamics, is then defined that will eventually transform equilibrium system 1 into equilibrium system 2. The systems are distinguished by introducing a parameter, $\lambda(t)$, which takes on a value λ_1 in system 1 and λ_2 in system 2, and the transformation is also parameterised through $\lambda(t)$ with $\lambda(0) = \lambda_1$ and $\lambda(\tau) = \lambda_2$. The equations of motion are therefore non-autonomous (*i.e.* they depend *explicitly* on absolute time).

Definition

We define a *generalised dimensionless "work"*, $\Delta X_{\tau}(\Gamma)$, for a trajectory of duration τ , originating from the phase point Γ as,

$$\exp[\Delta X_{\tau}(\Gamma)] = \lim_{d\Gamma \to 0} \frac{p_{eq,1}(d\Gamma)Z(\lambda_1)}{p_{eq,2}(dS^{\tau}\Gamma))Z(\lambda_2)},$$

$$= \frac{f_{eq,1}(\Gamma)d\Gamma Z(\lambda_1)}{f_{eq,2}(S^{\tau}\Gamma)dS^{\tau}\Gamma Z(\lambda_2)}, \forall \Gamma \in D_1$$
(8.1.1)

where $Z(\lambda_i)$ is the partition function for the system *i*=1,2 and D_1 is the accessible phase space domain for system 1, (*e.g.* coordinates in a fixed special range (-L,+L), and momenta range from $(-\infty,+\infty)$). In equation (8.1.1), $d\Gamma$ is an infinitesmal phase space volume centred on Γ and $dS^{\tau}\Gamma$ an infinitesimal phase volume centred on $S^{\tau}\Gamma$. Without loss of generality we assume both equilibrium distribution functions are even functions of the momentum. This implies that we are not moving relative to both systems. If the system is not canonical the *partition function* Z_i is just the normalization factor for the equilibrium distribution function $f_{eq}(\Gamma)$ and $f_{eq}(\Gamma) = \exp[F(\Gamma)]/Z$ where $F(\Gamma)$ is some real single valued phase function.

Although the physical significance of the generalized work, *X*, might seem obscure at this point, we will show that for particular choices of dynamics and ensemble, it is related to important thermodynamic properties and when it is evaluated along quasi-static paths it is in fact a path independent state function.

Before proceeding further with the analysis, it is useful to consider precisely what the generalized work is dependent upon. Firstly it is a function of the *equilibrium* states 1 and 2. This occurs via the equilibrium distributions appearing in (8.1.1) and also the partition functions for those states – see (8.1.1). Secondly it is a function of the endpoints of the possibly *nonequilibrium* phase space trajectory that takes phase Γ to $S^{\tau}\Gamma$. It is also a function of how much heat is gained or lost from the system over the duration of that trajectory. This heat loss determines the ratio of phase space volumes $d\Gamma/dS^{\tau}\Gamma$. Lastly it is a function of the duration of the trajectories τ .

The probability of observing ensemble members within the infinitesimal phase volume $d\Gamma$, centred on the phase vector Γ , in the initial equilibrium distribution function, $f_{eq,1}(\Gamma)$ is $p_{eq,1}(d\Gamma, 0) = f_{eq,1}(\Gamma)d\Gamma$.

It is very important to note that the time τ is the time at which the parametric change in λ is complete. This means that at time τ this system is *not* necessarily in equilibrium: $f(\Gamma; 0) = f_{eq,1}(\Gamma)$ but $f(\Gamma; \tau) \neq f_{eq,2}(\Gamma)$ in general. The generalized work is defined with respect to two different *equilibrium* distributions and the end points of *finite time* phase space trajectories: $S^s \Gamma : 0 \le s \le \tau$. The Equilibrium Relaxtion Theorem (§5) says that if the system is T-mixing and if the initial equilibrium distribution is an even function of the momenta then, for the purposes of computing averages of physical observeables, $\lim_{t \to \infty} f(\Gamma; t) = f_{eq,2}(\Gamma)$.

5

Definition

In order for $\Delta X_{\tau}(\Gamma)$ to be well defined, $\forall \Gamma \in D_1$, then $S^{\tau}\Gamma \in D_2$ and both $f_{eq,1}(\Gamma) \neq 0$ and $f_{eq,2}(S^{\tau}\Gamma) \neq 0$. This is known as the *ergodic consistency for the generalised work*.

We identify $||\partial S^{\tau} \Gamma / \partial \Gamma||$ as the Jacobian determinant and note that

$$\left\|\frac{\partial S^{\tau} \Gamma}{\partial \Gamma}\right\| = \frac{dS^{\tau} \Gamma}{d\Gamma}.$$
(8.1.2)

The GCFT considers the probability, $p_{eq,f}(\Delta X_t = B \pm dB)$, of observing values of $\Delta X_t = B \pm dB$ for forward trajectories starting from the initial equilibrium distribution 1, $f_1(\Gamma; 0) = f_{eq,1}(\Gamma)$, and the probability, $p_{eq,r}(\Delta X_t = -B \mp dB)$, of observing $\Delta X_t = -B \pm dB$ for reverse trajectories but starting from the equilibrium distribution given by $f_{eq,2}(\Gamma)$, of system 2.

Consider two equilibrium ensembles from which initial trajectories can be selected with known equilibrium distributions: $f_{eq,1}(\Gamma)$ and $f_{eq,2}(\Gamma)$.

If initially we select phases from $f_{eq,1}(\Gamma)$, employ a particular protocol (f) and corresponding time-dependent dynamics, defined by a parameter $\lambda_f(s)$ with $\lambda_f(0) = \lambda_1$ and $\lambda_f(\tau) = \lambda_2$, then the probability that the phase variable defined in (8.1.1) takes on the value B is given by:

$$p_{eq,1}(\Delta X_{\tau,f} = B \pm dB) = \int_{\Delta X_{\tau,f} = B \pm dB} d\Gamma f_{eq,1}(\Gamma) .$$
(8.1.3)

If initially we select phases from $f_{eq,2}(\Gamma)$ with a particular protocol (*r*) which is the time-reverse of (*f*), $\lambda_r(s) = \lambda_f(\tau - s)$ and corresponding time-dependent dynamics, so $\lambda_r(0) = \lambda_2$ and $\lambda_r(\tau) = \lambda_1$ then the probability that the phase variable defined in (8.1.1) takes on the value -B is given

by:
$$p_{eq,2}(\Delta X_{\tau,f} = -B \mp dB) = \int_{\Delta X_{\tau,f} = -B \mp dB} d\Gamma f_{eq,2}(\Gamma).$$

We note that a trajectory starting at point Γ , and evolved forward in time with the forward protocol to the point $S^{\tau}\Gamma$, will be related by a time reversal mapping to a trajectory starting at $M^{T}S^{\tau}\Gamma$ and evolving with the time-reverse protocol. If $S_{f/r}^{\tau}$ is the time evolution operator with forward/reverse protocol:

$$M^{T}S_{r}^{\tau}M^{T}S_{f}^{\tau}\Gamma=\Gamma.$$
(8.1.4)

Now we look at the ratio of these two probabilities:

$$\frac{p_{eq,1}(\Delta X_{\tau,f} = B \pm dB)}{p_{eq,2}(\Delta X_{\tau,r} = -B \mp dB)} = \frac{\int_{\Delta X_{\tau,f}(\Gamma) = B \pm dB} d\Gamma f_{eq,1}(\Gamma)}{\int_{\Delta X_{\tau,r}(\Gamma) = -B \mp dB} d\Gamma f_{eq,2}(\Gamma)}.$$
(8.1.5)

If $\Delta X_{\tau,f}(\mathbf{\Gamma}) = B$ then from (8.1.1) we know that $\Delta X_{\tau,r}(S^{\tau}\mathbf{\Gamma}) = -B$. We can therefore write (8.1.5) as:

$$\frac{p_{eq,1}(\Delta X_{\tau,f} = B \pm dB)}{p_{eq,2}(\Delta X_{\tau,r} = -B \mp dB)} = \frac{\int_{\Delta X_{\tau,f}(\Gamma) = B \pm dB} d\Gamma f_{eq,1}(\Gamma)}{\int_{\Delta X_{\tau,r}(\Gamma) = -B \mp dB} d\Gamma f_{eq,2}(\Gamma)}$$

$$= \frac{\int_{\Delta X_{\tau,f}(\Gamma) = B \pm dB} d\Gamma f_{eq,1}(\Gamma)}{\int_{\Delta X_{\tau,r}(M^{T}S^{T}\Gamma) = -B \mp dB} dM^{T}S^{T}\Gamma f_{eq,2}(M^{T}S^{T}\Gamma)}$$

$$= \frac{\int_{\Delta X_{\tau,f}(\Gamma) = B \pm dB} d\Gamma f_{eq,1}(\Gamma)}{\int_{\Delta X_{\tau,f}(\Gamma) = B \pm dB} d\Gamma exp[-\Delta X_{\tau,f}(\Gamma)]f_{eq,1}(\Gamma)Z(\lambda_{1})/Z(\lambda_{2})}$$

$$= \exp[B]\frac{Z(\lambda_{2})}{Z(\lambda_{1})}$$
(8.1.6)

Definition

Equation (8.1.6) is the *Generalized Crooks Fluctuation Relation* (GCFR) and its derivation is the Generalized Crooks Fluctuation Theorem (GCFT).

We note that at time τ the system that is evolving from $f_{eq,1}(\Gamma)$ will not have a relaxed *to* the equilibrium distribution $f_{eq,2}(\Gamma)$ (or vice versa). In fact complete relaxation never takes place in finite time – see §5.3,4. We can compute the change in the generalized work going from a time τ to time $\tau + s$. During this interval there is no parametric change and the system simply relaxes towards the equilibrium state, 2. From equation (8.1.1) we see that,

$$\begin{aligned} \exp[\Delta X_{\tau+s}(\Gamma) - \Delta X_{\tau}(\Gamma)] &= \frac{f_{eq,1}(\Gamma)d\Gamma Z(\lambda_1)}{f_{eq,2}(S^{\tau+s}\Gamma)dS^{\tau+s}\Gamma Z(\lambda_2)} \frac{f_{eq,2}(S^{\tau}\Gamma)dS^{\tau}\Gamma Z(\lambda_2)}{f_{eq,1}(\Gamma)d\Gamma Z(\lambda_1)} \\ &= \frac{f_{eq,2}(S^{\tau}\Gamma)dS^{\tau}\Gamma}{f_{eq,2}(S^{\tau+s}\Gamma)dS^{\tau+s}\Gamma}, \quad \forall s > 0 \end{aligned}$$

$$(8.1.7)$$

If we look at the second line of (8.1.7), we recognize that it is simply the integrated dissipation function (3.1.2), $\Omega_{eq,2,s}(S^{\tau}\Gamma)$ defined in the Evans-Searles Fluctuation Theorem, for *equilibrium* system 2, evaluated at a phase $S^{\tau}\Gamma$, and integrated for a time *s*. It is important to note that both the numerator and the denominator of (8.1.7) involve *forward* time integrations from system 2 equilibrium (*i.e.* there's no forward and reverse as in (8.1.1)). Therefore,

$$[\Delta X_{\tau+s}(\Gamma) - \Delta X_{\tau}(\Gamma)] \equiv \Delta X_s(S^{\tau}\Gamma)$$

= $\ln \frac{f_{eq,2}(S^{\tau}\Gamma)}{f_{eq,2}(S^{\tau+s}\Gamma)} - \int_0^{\tau} ds \Lambda(S^{\tau+s}\Gamma)$ (8.1.8)
= $\Omega_{eq,2,s}(S^{\tau}\Gamma) = 0, \forall \Gamma \in D_2, \forall s > 0$

The last line is identically zero because the dissipation function $\Omega_{eq}(\Gamma)$ for all equilibrium systems is identically zero – see §3.7 - and we know from the

Equilibrium Relaxation Theorem that the system does eventually relax to the unique, ergodic, dissipationless equilibrium state of system 2!

8.2 GENERALIZED JARZYNSKI EQUALITY

The Generalized Jarzynski Equality (GJE), can be thought of as the analogue of the NonEquilibrium Partition Identity (NPI) evaluated for the generalized work. We say "analogue" because the introduction of forward and reverse paths in the definition of the generalized work is quite different from the use of forward only paths, for the dissipation function.

The derivation of the GJE, from the GCFT is trivial. The simplest approach is to obtain GJE by integration of the GCFT, eq. (8.1.6):

$$\left\langle \exp[-\Delta X_{\tau}(\mathbf{\Gamma})] \right\rangle_{eq,1} = \int_{-\infty}^{+\infty} dB \ p_f(\Delta X_{\tau} = B) \exp(-B)$$

$$= \int_{-\infty}^{+\infty} dB \ p_r(\Delta X_{\tau} = -B) \frac{Z(\lambda_2)}{Z(\lambda_1)}$$

$$= \frac{Z(\lambda_2)}{Z(\lambda_1)}$$

$$(8.2.1)$$

If the two states had the same free energies (8.2.1) would, *superficially*, look almost identical to the NPI. The proof is line by line analogous to that given in §3.3 for the NPI. We also note that with the change of variables the domain of integration may change.

From the first line of eq. (8.2.1), it is clear that trajectories for which the value of ΔX_{τ} is negative have a contribution to the ensemble average that is exponentially enhanced. Therefore in order to obtain numerical convergence of the ensemble average, it is important that these trajectories are sufficiently well sampled. Many recent studies have addressed this issue and have developed algorithms to improve convergence [53-65]. If the averaging process is not sufficiently exhaustive for these possibly extremely rare events to be observed, (8.1.6) & (8.2.1) will give incorrect results. This observation has an immediate impact on the calculation of free energy differences in the thermodynamic limit. This difference *must* be calculated in finite systems for a series of system sizes and then extrapolated to obtain the value in the thermodynamic limit. If you apply GCFT or GJE to extremely large systems, one will *never* observe the required fluctuations and incorrect estimates will be inferred.

As is the case for NPI the GJE can be derived straightforwardly from GCFR (8.1.6). Here we compute the relevant average directly:

$$\begin{split} \left\langle \exp[-\Delta X_{\tau}(\mathbf{\Gamma})] \right\rangle_{eq,1} &= \int_{D_{1}} d\mathbf{\Gamma} f_{eq,1}(\mathbf{\Gamma}) \frac{f_{eq,2}(S^{\tau}\mathbf{\Gamma})Z(\lambda_{2})dS^{\tau}\mathbf{\Gamma}}{f_{eq,1}(\mathbf{\Gamma})Z(\lambda_{1})d\mathbf{\Gamma}} \\ &= \int_{D_{2}} dS^{\tau}\mathbf{\Gamma} f_{eq,2}(S^{\tau}\mathbf{\Gamma}) \frac{Z(\lambda_{2})}{Z(\lambda_{1})} \quad , \tag{8.2.2} \end{split}$$
$$= \frac{Z(\lambda_{2})}{Z(\lambda_{1})}$$

where the brackets $\langle ... \rangle_{eq,1}$ denote an equilibrium ensemble average over the initial equilibrium distribution. We also note that with the change of variables the domain of integration may change.

The GJE, is very widely applicable. It relates the ensemble average of the exponential of *nonequilibrium* path integrals to *equilibrium* thermodynamic free energy differences.

The validity of (8.2.2) requires that $\forall \Gamma \in D_1, f_{eq,1}(\Gamma) \neq 0$, implies $\forall S^{\tau}\Gamma \in D_2, f_{eq,2}(S^{\tau}\Gamma) \neq 0$. This is the *ergodic consistency condition for the generalized work*.

Equations (8.2.1,2) are very general, and they even apply to stochastic dynamics (see reference 3). Obviously the paths do not need to be quasi-static paths as in traditional thermodynamics. These equations are independent of the particular protocol, provided ergodic consistency holds. In fact it is possible to average over the initial ensemble *and* a set of protocols, since the final answer is protocol or path independent.

As is the case for the NonEquilibrium Partition Identity (NPI) the GJE can be, as we have seen, proved from the GCFR. However, the reverse is not true because the Fluctuation Relations contain more information than either the NPI or GJE.

As we mentioned earlier, like the GCFT, the application of the GJE relies on the observation of improbable fluctuations. In order to yield reliable estimates of free energy differences one must sample the trajectories that are the conjugate antitrajectories of the most probable trajectories. This means that these formulae are of limited use for computing free energy differences in the thermodynamic, or large system, limit.

8.3 MINIMUM AVERAGE GENERALIZED WORK

We now derive a further simple corollary of the GJE. From equation (8.2.2), we see that,

$$\frac{Z(\lambda_{2})}{Z(\lambda_{1})} = \left\langle \exp[-\Delta X_{\tau}] \right\rangle_{1}$$

$$= \exp[-\left\langle \Delta X_{\tau} \right\rangle_{1}] \left\langle \exp[-\Delta X_{\tau} + \left\langle \Delta X_{\tau} \right\rangle_{1}] \right\rangle$$

$$\geq \exp[-\left\langle \Delta X_{\tau} \right\rangle_{1}] \left\langle 1 - \Delta X_{\tau} + \left\langle \Delta X_{\tau} \right\rangle_{1} \right\rangle$$

$$= \exp[-\left\langle \Delta X_{\tau} \right\rangle_{1}]$$
(8.3.1)

In deriving this relation we have used the fact that⁶ $e^x \ge 1 + x$, $\forall x \in \mathbb{R}$. Taking the logarithms of both sides and then multiplying both sides by -1:

$$\left\langle \Delta X_{\tau} \right\rangle_{1} \ge \ln \left[\frac{Z(\lambda_{1})}{Z(\lambda_{2})} \right].$$
 (8.3.2)

This is clearly the analogue of the Second Law Inequality⁵ for systems of changing free energy. Some authors refer to work inequalities like (8.3.2) as the Clausius Inequality⁷, however we reserve that term for *cyclic* inequalities of the *heat*, since as Planck remarked⁸, "this is the form of the Second Law first enunciated by Clausius".

In actual systems the right hand side will turn out to be a dimensionless free energy difference. For example if systems 1, 2 are canonical and at the same temperature and have the same number of particles and volume, as we will see later: $\ln[Z_1/Z_2] = \beta \Delta A_{21} = \beta(A_2 - A_1)$ and $\Delta X = \beta \int_0^{\tau} ds W(s)$ where *W* denotes the work (*i.e.* the change internal energy minus that change caused by the heat). A_i is the Helmholtz free energy of system *i*. The minimum average work inequality implies in this case $\Delta W_{21} \ge \Delta A_{21}$. The minimum work is expended if the path is reversible or quasi-static, in which case that work is in fact the difference in the Helmholtz free energies divided by k_BT .

If the parametric protocol takes us around a closed cycle that is defined in terms of the parameter $\lambda(t)$ we see that since by definition $Z_1 / Z_2 = 1$:

$$\oint ds \langle \dot{X}(s) \rangle = \oint \langle dX \rangle \ge 0 . \tag{8.3.3}$$

The ensemble average of the cyclic integral of the generalized work is non-negative.

Definition

We call equation (8.3.3) the *nonequilibrium cyclic inequality for the generalized work*.

Equation (8.3.3) for the generalised work is very different from the corresponding cyclic integral of the heat. For the work, we simply execute the protocol cycle. For the heat, as we will soon see, we have to complete the cycle many times and wait until the system settles into a periodic response – not all systems do! – before we can apply the cyclic integral of the heat.

The reason for the difference is if we execute the cycle once along a nonquasi-static path, when the parameter reaches the value λ_2 , even though the system is not yet in equilibrium there is by (8.1.8), no further change in the work during the long relaxation process. The heat exchanged does change during this relaxation however!

Definitions

Because the dynamics is microscopically reversible the cyclic integral can only be zero if the cycle is *thermodynamically* reversible – what we term *quasi-static*. A pathway is traversed quasi-statically if the average work for a forward path is equal and opposite the average work for the reversed path.

The cyclic integral of the generalized work for a quasi-static cycle is zero. The proof is obvious. The fact that the cyclic integral of the generalized work is zero $\oint dX = 0$, also implies that:

$$\int_{qs}^{f} dX = independent \ of \ path$$
(8.3.4)

where the subscript "qs" denotes the fact that the integral is for a quasi-static or thermodynamically reversible pathway. The proof of (8.3.4) is obvious. Construct a reversible cycle $i \rightarrow f, f \rightarrow i$. The cyclic integral must be zero, so if we vary the pathway for the return leg $f \rightarrow i$ we must always get the same value for the integrated reversible work, independently of the precise path.

Definition

Finally we can see that if the integral of the generalized work for paths is independent of the pathway that integral must be a *state function* (*i.e.* a function only of the initial and final states of the system). In fact this is why the seemingly abstract generalized work defined in equation (8.1.1) is so important. The generalized work for a thermodynamically reversible pathway is always a history and path independent function of the thermodynamic states of the system at the end points of the path.

Example

What we have proved above is true if quasi-static paths exist but we have not shown they do. Consider the case where the parameter $\lambda(t)$ is equal to the strain and suppose we wish to strain the crystal of volume *V*, though an angle $\delta\gamma$. For simplicity suppose that our protocol is to strain the crystal at a constant rate $\dot{\gamma}$. The time taken to increase the strain from zero to $\delta\gamma$ is $\delta\gamma/\dot{\gamma}$. For small strains, we expect that,

$$\lim_{\dot{\gamma},\gamma\to 0} \left\langle P_{xy}(t) \right\rangle = -G_0 \gamma(t) - \eta_{0^+} \dot{\gamma}, \quad 0 < t < \frac{\delta \gamma}{\dot{\gamma}}$$
(8.3.5)

where G_0 is the zero frequency elastic shear modulus which will in the end give the quasi-static or macroscopically reversible work and η_{0+} is the limiting zero frequency shear viscosity of the crystal. [Note: You cannot speak of *the* zero frequency shear viscosity of a solid!] The generalized work turns out to be

$$\begin{split} \lim_{\dot{\gamma}\to 0} \langle \Delta X \rangle &= -\lim_{\dot{\gamma}\to 0} \beta V \int_{0}^{\delta \gamma/\dot{\gamma}} ds \left\{ P_{xy}(s) \right\} \dot{\gamma} \\ &= \lim_{\dot{\gamma}\to 0} \beta V \int_{0}^{\delta \gamma/\dot{\gamma}} ds \left[G_{0} \dot{\gamma}^{2} s + \eta_{0^{+}} \dot{\gamma}^{2} \right] \\ &= \lim_{\dot{\gamma}\to 0} \beta V \left[\frac{G_{0} \dot{\gamma}^{2} \delta \gamma^{2}}{2 \dot{\gamma}^{2}} + \frac{\eta_{0^{+}} \dot{\gamma}^{2} \delta \gamma}{\dot{\gamma}} \right] \\ &= \lim_{\dot{\gamma}\to 0} \beta V \left[\frac{G_{0} \delta \gamma^{2}}{2} + \eta_{0^{+}} \dot{\gamma} \delta \gamma \right] \\ &= \frac{\beta V G_{0} \delta \gamma^{2}}{2} \end{split}$$

$$(8.3.6)$$

The second viscous term is always positive independent of the sign of $\delta\gamma$, $\dot{\gamma}$ whereas the first term is reversible. It can be positive or negative. Obviously if we strain the crystal though a cycle the first term will then also vanish while for finite strain rates the second term will satisfy the inequality (8.3.3).

8.4 NONEQUILIBRIUM WORK RELATIONS FOR THERMAL PROCESSES

We wish to consider a realistic model of a system that is driven away from equilibrium by a reservoir whose temperature is changing. For this case the simple parametric change in the Hamiltonian or external field usually employed in the derivation of the GJE or the GCFT, is not applicable and care is needed in developing the physical mechanisms.

Here we could address this issue by considering a system of interest, containing some very slowly relaxing constituents such as soft matter or pitch, in contact with a rapidly relaxing reservoir. The reservoir may be formed from a copper block or another highly thermally conductive material. Changing the temperature of the reservoir (say with a thermostatically controlled heat exchanger) then drives the system of interest out of equilibrium. The change in temperature is slow enough that the reservoir may be treated to high accuracy, as undergoing a quasi-static temperature change. The slowly relaxing system of interest is far from equilibrium. We employ the GCFT and the GJE to describe this system. Importantly the quantities that appear in the theory are physically measurable variables.

Another mechanism for achieving the required result would be (following Planck⁸) to have a set of large equilibrium thermostats that can be thermally coupled to the system of interest in a protocol sequence. If these thermostats are large they can be regarded as being in thermal equilibrium. If they are sufficiently remote from the system of interest there is no way the system of interest can "know" the precise mathematical details of how heat is ultimately taken from or added to the system of interest.

For convenience from a theoretical perspective, we choose the Nosé-Hoover thermostatting mechanism⁹ and the equations of motion, including the thermostat multiplier are then:

17

$$\dot{\mathbf{q}}_{i} = \frac{\mathbf{p}_{i}}{m}$$

$$\dot{\mathbf{p}}_{i} = \mathbf{F}_{i}(\mathbf{q}) - S_{i}(\alpha(\mathbf{\Gamma})\mathbf{p}_{i} + \gamma_{th})$$

$$\dot{\mathbf{x}} = \left(\frac{\sum_{i=1}^{N} S_{i}\mathbf{p}_{i} \cdot \mathbf{p}_{i} / m}{3(N_{th} - 1)k_{B}T(t)} - 1\right) \frac{1}{\tau_{\alpha}^{2}}$$
(8.4.1)

where τ_{α} is an arbitrary Nosé-Hoover time constant. The value of T(t) is the target temperature of the thermostat and $S_i = 0,1$ is a switch that controls which particles are coupled to the Nosé-Hoover thermostat - $\sum_{i=1}^{N} S_i = N_{th}$. In our model the particles that are coupled to the thermostat can be taken to be remote from the system of interest. This ensures that the particles in the system of interest are ignorant of the precise details of this unphysical thermostat. These thermostatted particles are also subject to a fluctuating force γ_{th} that is chosen to ensure that the total momentum of the

thermostatted particles
$$\sum_{i=1}^{N} S_i \mathbf{p}_i \equiv \mathbf{p}_{th} = \mathbf{0}$$
 is identically zero.

The extended, time-dependent internal energy is,

$$H_E(\Gamma, \alpha, t) = H_0(\Gamma) + \frac{3(N_{th} - 1)}{2} k_B T(t) \alpha^2 \tau_{\alpha}^2 \text{ and the extended phase space of the}$$

system is $\Gamma' = (\Gamma, \alpha)$. The phase continuity equation (2.4.9), states: $df / dt = -\Lambda f$, and using (8.4.1) it is easy to show that,

$$k_B T \Lambda = k_B T \left(\frac{\partial}{\partial \Gamma} \cdot \dot{\Gamma} + \frac{\partial}{\partial \alpha} \dot{\alpha} \right) = -3(N_{th} - 1)k_B T \alpha = -\dot{Q}^{therm}$$
, where \dot{Q}^{therm} is the rate of

decrease in H_E due to the thermostat or equivalently the rate of increase of energy by the external thermostat. From the relaxation theorem (§5), the unique equilibrium distribution function for this system at a fixed temperature *T* is then

$$f_{eq}(\mathbf{\Gamma};T,\alpha) = \frac{\tau_{\alpha}\sqrt{3(N_{th}-1)/(2\pi)}}{Z_{c}(T)} \exp(-\beta H_{E}(\mathbf{\Gamma},T,\alpha))\delta(\mathbf{p}_{th}) \qquad (8.4.2)$$

where $Z_c(T)$ is the canonical partition function and, $\lambda(t) \equiv T(t)$.

We now consider applying the GCFR, eq. (8.1.6), when a thermal rather than a mechanical process occurs. Consider a thermostatted system of *N* particles whose target kinetic temperature is changed from T_1 to T_2 over a period $0 < t < \tau$. We do not change the Hamiltonian during this process. For simplicity we consider a canonical ensemble for the two equilibrium states (8.4.2), and use the equations of motion (8.4.1). The temperature dependence of the reservoir is achieved by making the Nosé-Hoover target temperature T(t) in (8.4.1) a time dependent parameter.

From (8.1.1,8.4.1) we see that the generalized dimensionless work is

$$\Delta X_{\tau}(\Gamma';0,\tau) = \beta_2 H_E(S^{\tau}\Gamma') - \beta_1 H_E(\Gamma') + \int_0^{\tau} dt \ \beta(t) \dot{Q}_{therm}(S^{t}\Gamma')$$
(8.4.3)

where $\beta(t) = 1/(k_B T(t))$ is the inverse, time-dependent target temperature. (Note: as noted above $dE = dW - dQ_{therm} = dW + dQ_{soi}$ because, following Planck, our change in the heat refers to the thermal reservoir (therm) rather than to the system of interest.) Now if we take the derivative of the extended Hamiltonian while the temperature is changing, but with no other external agent acting on the system, we obtain using (8.4.1),

$$\frac{d}{dt}H_E(S'\Gamma') = -\dot{Q}_{therm}(S'\Gamma') + \frac{3}{2}(N_{therm} - 1)k_B\dot{T}(t)\alpha^2(t)\tau_\alpha^2.$$
(8.4.4)

We then obtain

$$\frac{d}{dt}[\beta(t)H_E(S'\Gamma')] = -\beta(t) \left[H_0(S'\Gamma')\frac{\dot{T}(t)}{T(t)} + \dot{Q}(S'\Gamma') \right], \qquad (8.4.5)$$

and combining (8.4.3) & (8.4.5), the generalised "power" for a change in the target temperature with time is,

$$\dot{X}(S'\Gamma) = \dot{\beta}(t)H_0(S'\Gamma).$$
(8.4.6)

Note that the right hand side of (8.4.6) only depends upon physical variables and not the unphysical thermostat multiplier α or the extended Hamiltonian. Eq. (8.2.1) then becomes

$$\left\langle \exp\left(-\int_{0}^{\tau} dt \ \dot{\beta}(t) H_{0}(S'\Gamma)\right) \right\rangle_{1} = \frac{Z_{c,2}}{Z_{c,1}} = \exp\left[-\beta_{2}A_{2} + \beta_{1}A_{1}\right].$$
 (8.4.7)

Definition

In (8.4.7) we have defined the *Helmholtz Free energy* in terms of the logarithm of the canonical partition function:

$$A_i \equiv -k_B T \ln\left(\int d\Gamma \exp(-\beta H_i(\Gamma))\right), \quad i = 1,2$$
(8.4.8)

For quasi-static processes the exponent of the left hand side of (8.4.7) has no fluctuations, one can use (8.4.8) to show that,

$${}_{qs} \int_{0}^{\tau} dt \, \dot{\beta}(t) U(t) = {}_{qs} \int_{0}^{\tau} dt \, \dot{T}(t) \frac{d}{dT} [\beta(t) A(t)]$$

$$= {}_{qs} \int_{0}^{\tau} dt \, \frac{d}{dt} [\beta(t) A(t)] \qquad .. \qquad (8.4.9)$$

$$= {}_{\beta_2} A_2 - {}_{\beta_1} A_1$$

where the subscript "qs" denotes that the integrals are for quasistatic processes only. In deriving (8.4.9) we have used the fact that the internal energy is just the canonical average of the Hamiltonian. For this case we see that the dimensionless "power" is the rate of change of dimensionless Helmholtz free energy.

For temperature changes at finite rates, the thermodynamic temperature of the system of interest cannot be defined and the kinetic temperature of the system of interest may not be equal to the temperature of the thermal reservoir. Nonetheless (8.4.7) can still be used to compute changes in the free energy of the system of interest as specified by (8.4.1).

If one constructs an algorithm (8.4.1) to accomplish some thermal transformation $(N_1, V_1, T_1) \rightarrow (N_1, V_1, T_2)$ then (8.4.6) gives a precise microscopic form for the generalised "work" appearing in the classical thermodynamic path integral for the free energy change. Although the quasistatic path integral expression is unique, the nonequilibrium expression is certainly not. This is because there are infinitely many protocols that accomplish the required change. Nonetheless each of these expressions gives identical values for the free energy difference.

8.5 CLAUSIUS' INEQUALITY AND THERMODYNAMIC TEMPERATURE

We now turn our attention away from work, to heat. As before we consider a periodic protocol. However for the heat (and unlike work), we can only deduce useful results if the system responds periodically to the cyclic protocol. We note that if we periodically cycle a given protocol, not all systems will respond periodically. The necessary and sufficient conditions for the system to respond periodically are not known. Clausius' Inequality only applies if, in the long time limit $(t \rightarrow \infty)$, the average system response is periodic.

Consider a system with the equations of motion given by (8.4.1). If we now substitute equation (8.4.3) into (8.3.3) and apply it to a *periodic cycle* after any cyclic transients have decayed, we can deduce that,

$$\lim_{t \to \infty} \oint_P ds \left\langle \dot{X}(t+s) \right\rangle = \lim_{t \to \infty} \oint_P ds \left\langle \beta(t+s) \dot{Q}_{therm}(t+s) \right\rangle = \lim_{t \to \infty} \oint_P \left\langle \frac{dQ_{therm}}{k_B T} \right\rangle \ge 0, \quad (8.5.1)$$

where we use the notation: $\oint_P ds$ to denote the cyclic integral of a periodic function. Because the cycle is periodic, in (8.4.3) the change in $\langle \beta H_E \rangle$ around the cycle is identically zero. In this equation *t* is the time that you start the cyclic integral. Thus for periodic cycles (8.3.3) reduces to a cyclic integral of the heat divided by the target temperature.

In more usual notation equation (8.5.1) implies that, in the large system limit, $N \rightarrow \infty$, where fluctuations are negligible we obtain the very well-known Clausius Inequality⁸:

$$\lim_{N \to \infty} \lim_{t \to \infty} \oint_{P} \frac{dQ_{therm}}{T} \ge 0.$$
(8.5.2)

In these equations (8.5.1,2), regardless of the Nosé-Hoover time constant, the time dependent temperature is the instantaneous value of the *target temperature* of the Nosé-Hoover thermostat. At any instant the numerical value of the target temperature is in fact the equilibrium thermodynamic temperature that the entire system would relax to, if at that same moment, this target temperature was fixed at its current value and the entire system is allowed to relax to thermodynamic equilibrium. We know that this is so from the Relaxation Theorem for T-mixing systems (§5). We will often use the description that, the temperature appearing in (8.5.1,2) is at any instant of time, the equilibrium thermodynamic temperature of the *underlying* equilibrium state.

If the cycle is traversed rapidly so that the system of interest is not in thermodynamic equilibrium, the actual thermodynamic temperature of the system of interest is of course, not defined.

If the thermostat is composed of a large Hamiltonian region coupled to the system of interest and a remote Nosé-Hoover thermostatted region, we can argue that the precise details of the thermostat cannot possibly be "known" to the system of interest and are therefore unimportant.

If the thermostat is comparable in size to the system of interest and if the cycle is traversed quickly, both the system of interest and the thermostat will be away from equilibrium. At any point in the cycle there is a profound difference between the nonequilibrium state generated by Gaussian isokinetic and Nosé-Hoover thermostats. However, for both types of thermostat (8.5.1,2) take the same form. For Gaussian thermostats the change in the kinetic temperature of the thermostat is instantaneous whereas for Nosé-Hoover thermostats there is a variable phase lag $\sim \tau_{\alpha}$ in equation (8.4.1). (The value of this feedback time constant is completely arbitrary.) The *only* "temperature" any of these systems have in common is the equilibrium thermodynamic temperature of the *underlying* equilibrium state. At any point in the cycle the precise nature of the nonequilibrium state (*e.g.* the instantaneous average pressure or energy) is highly dependent on the phase lag τ_{α} , or whether the thermostat is Gaussian or Nosé-Hoover like.

In Planck's discussion of Clausius' Inequality⁹, at any instant in the cycle, T is the equilibrium thermodynamic temperature of the particular large equilibrium reservoir with which the system of interest is currently in contact with.

23

Clausius' Thermodynamic Inequality (8.5.2), is of course only exact in the thermodynamic limit and in small systems it can occasionally be violated as in (8.5.1). The probability ratio that for a finite system the work integral takes on a value A compared to -A can be computed from a time dependent version of the fluctuation theorem.

Equations (8.5.1,2) show that on average we cannot construct a perpetual motion machine of the second kind. A perpetual motion machine of the second kind would require that $\oint_P \langle dQ^{therm} / T \rangle < 0$ so that abient heat from the reservoir is converted into useful work in the system of interest. Thus the proof of (8.5.1,2) constitutes a direct mechanical proof of Clausius' statement of the Second "Law" of thermodynamics.

If the cycle is reversible we can apply equation (8.5.2) to the forward cycle and to the reversed cycle that must have the same value for magnitude of the integral but opposite sign. The only possible value for both integrals for reversible cycles, is therefore zero:

$$\lim_{N \to \infty} {}_{qs} \oint \frac{dQ_{therm}}{T} = 0.$$
(8.5.3)

The subscript *qs* denotes a quasi-static cycle. We note that a quasi-static cycle cannot have any transients and is always periodic.

Applying the same arguments as we did for the quasi-static cyclic integral of the generalized work shows that the quasi-static integral from state 1 to state 2:

$$\lim_{N \to \infty} {}_{qs} \int_{1}^{2} \frac{dQ_{\text{therm}}}{T} \equiv S_{\text{therm,2}} - S_{\text{therm,1}} = -S_{\text{soi,2}} + S_{\text{soi,1}}$$
(8.5.4)

is independent of the path from state 1 to 2. **Definition**

In equation (8.5.4) S_{soi} denotes the equilibrium *entropy* of the system of interest. The state function S_{therm} , defined in equation (8.5.4) is known as the *equilibrium entropy* of the thermostatting reservoir.

If we combine equations (8.4.3,8) for quasi static processes in the thermodynamic, we see that in the thermodynamic limit where fluctuations are negligible

$$\lim_{N \to \infty} {}_{qs} \int_{i}^{f} \frac{\left\langle dQ^{therm} \right\rangle}{T} = -\left(\frac{A_2 - \left\langle H_{E,2} \right\rangle}{T_2}\right) + \left(\frac{A_1 - \left\langle H_{E,1} \right\rangle}{T_1}\right) = -S_{soi,2} + S_{soi,1}$$
(8.5.5)

(Note: We could remove the ensemble averages because in the thermodynamic limit fluctuations vanish.) Now if we compare (8.5.4) and (8.5.5) we see that in the thermodynamic limit, the equilibrium entropy of the system of interest *S*, the Helmholtz free energy *A* and the energy $\langle H_E \rangle$ must be related by the equation:

$$A = \langle H_E \rangle - TS = U - TS , \qquad (8.5.6)$$

where U is the internal energy of the system of interest. we have used the fact that the heat gained by the thermostat is equal and opposite to the heat gained by the system of interest.

We can now also reinterpret (8.4.3,6,8) as

$$dU = dQ + dW = TdS + dW \tag{8.5.7}$$

where all quantities (the change in the heat dQ which is not a state function, the change in the work dW, and the change in the equilibrium entropy dS) refer to the system of interest. Using (8.5.6) we find that the change in the Helmholtz free energy of the system of interest is given by:

$$dA = -SdT + dW \tag{8.5.8}$$

Equations (8.5.6,7,8) are immediately recognised as the conventional equations of classical thermodynamics for quasistatic processes in the thermodynamic limit. However unlike classical thermodynamics our equations came directly from the laws of mechanics and the axiom of causality. We did not need to *assume* Clausius' Inequality; we proved it. Indeed the Helmholtz free energy was defined using the logarithm of the partition function – a statistical mechanical expression rather than thermodynamic expression. Equation (8.5.6) is usually taken as the definition of the Helmholtz free energy. In our exposition (8.4.8) is the definition of the Helmholtz free energy and (8.5.6) a derived property.

Equation (8.5.4) tells us another very important piece of information. The integration factor for the heat in quasi-static processes is the time dependent thermodynamic temperature. The fact that the equilibrium temperature is the integrating factor for the heat ultimately comes from the form of the canonical equilibrium distribution function. The Equilibrium Relaxation Theorem (§5) says that this distribution is *unique* for T-mixing systems. The consequence of this is that the integrating factor for the heat is also *unique*.

We do not give the proof here but had we used isokinetic dynamics rather than Nosé-Hoover dynamics, the integration factor would have turned out to be the instantaneous kinetic temperature of the thermal reservoir.

In our proof of the Equilibrium Relaxation Theorem ($\S5$) for systems in contact with a heat bath⁵, we proved (\$5.6) that the Helmholtz free energy and the partition function are related by (8.4.8). We also proved for equilibrium systems, the equivalence of the target kinetic temperature employed in a Nosé-Hoover thermostat and the thermodynamic temperature for canonical systems. These proofs took the classical thermodynamic relations between the Helmholtz free energy, the internal energy, the entropy, volume and temperature as *given*. Now we have arrived at a completely new logical position. We have *proved* the Zeroth (\$5.10), and Second "Laws" of Thermodynamics (\$8.5); the latter in the form of Clausius' Equality (8.5.1,2). We take the First Law of thermodynamics as also given by the laws of mechanics. This means that logically we have now constructed thermodynamics (\$8.5,7,8) without any assumptions except the laws of mechanics, the assumption of T-mixing, ergodic consistency and the axiom of causality. Clausius of course proved his Theorem *assuming* the Second "Law" of Thermodynamics. He *assumed* that the

26

construction of a perpetual motion machine of the second kind, is impossible. Our proof requires no such assumption. This is quite a different epistemological point of view from that used by Clausius in 1854.

8.6 PURELY DISSIPATIVE GENERALIZED WORK

Definition

We define a generalised dimensionless, purely dissipative "work", $\Delta Y_{\tau}(\Gamma)$, for a trajectory of duration τ , originating from the phase point Γ , under this dynamics as in [43],

$$\exp[\Delta Y_{\tau}(\Gamma)] = \lim_{d\Gamma \to 0} \frac{p_{eq,1}(d\Gamma)}{p_{eq,2}(dS^{\tau}\Gamma)}$$

$$\equiv \frac{f_{eq,1}(\Gamma)d\Gamma}{f_{eq,2}(S^{\tau}\Gamma)dS^{\tau}\Gamma}, \forall \Gamma \in D$$
(8.6.1)

The derivation of results for this quantity is very similar to that for the generalized dimensionless work (§8.1,2,3) so we will quickly give a summary of the main results without rehearsing the proofs.

As before the time τ , is the time at which the parametric change in λ is complete. This means that at time τ this system is *not* necessarily at equilibrium.

Following the same procedure that led to (8.2.2) we see that:

$$\left\langle \exp[-\Delta Y_{\tau}(\Gamma)] \right\rangle_{1} = \int d\Gamma \ f_{1}(\Gamma;0) \frac{f_{2}(S^{\tau}\Gamma;0)dS^{\tau}\Gamma}{f_{1}(\Gamma;0)d\Gamma}$$

$$= 1$$
(8.6.2)

where the brackets $\langle ... \rangle_1$ denote an equilibrium ensemble average over the initial distribution. This equation is the analogue of the NonEquilibrium Partition Identity. The validity of (8.6.2) requires that there is an integrable region in the phase space of the final equilibrium distribution for which $f_2(S^{\tau}\Gamma; 0) \neq 0$, that is

 $\int dS^{\tau} \mathbf{\Gamma} f_2(S^{\tau} \mathbf{\Gamma}; 0) \neq 0$. This is the ergodic consistency condition for the generalized work.

The FR for the purely dissipative work considers the probability, $p_{eq,l}(\Delta Y_{\tau} = B)$, of observing values of $\Delta Y_{\tau} = B \pm dB$ for forward trajectories starting from the initial equilibrium distribution $f_{eq,l}(\Gamma,0)$, and the probability $p_{eq,2}(\Delta Y_{\tau} = -B)$ of observing $\Delta Y_{\tau} = -B \pm dB$ for reverse trajectories but starting from the equilibrium given by $f_{eq,2}(\Gamma,0)$. Proof of this GCFR closely resembles the proof of the ESFT and the GCFT (8.1.6):

$$\frac{p_{eq,1}(\Delta Y_{\tau,f} = B \pm dB)}{p_{eq,2}(\Delta Y_{\tau,r} = -B \mp dB)} = \frac{\int_{\Delta Y_{\tau,f}(\Gamma) = B \pm dB} d\Gamma f_{eq,1}(\Gamma)}{\int_{\Delta Y_{\tau,r}(\Gamma) = -B \mp dB} d\Gamma f_{eq,2}(\Gamma)}$$
$$= \frac{\int_{\Delta Y_{\tau,f}(\Gamma) = B \pm dB} d\Gamma f_{eq,1}(\Gamma)}{\int_{\Delta Y_{\tau,r}(M^{T}S^{T}\Gamma) = -B \mp dB} dM^{T}S^{T}\Gamma f_{eq,2}(M^{T}S^{T}\Gamma)}$$
$$= \frac{\int_{\Delta X_{\tau,f}(\Gamma) = B \pm dB} d\Gamma f_{eq,1}(\Gamma)}{\int_{\Delta Y_{\tau,f}(\Gamma) = B \pm dB} d\Gamma \exp[-\Delta X_{\tau,f}(\Gamma)]f_{eq,1}(\Gamma)}$$
$$= \exp[B]$$
(8.6.3)

Again we can see the change in the purely dissipative work *after* the parametric changes are complete:

$$\exp[\Delta Y_{\tau+s}(\Gamma) - \Delta Y_{\tau}(\Gamma)] = \frac{f_{eq,2}(S^{\tau}\Gamma)dS^{\tau}\Gamma}{f_{eq,2}(S^{\tau+s}\Gamma)dS^{\tau+s}\Gamma(\tau+s)}$$
(8.6.4)

which is again recognized as being the dissipation function for the second equilibrium. Therefore we have,

$$[\Delta Y_{\tau+s}(\Gamma) - \Delta Y_{\tau}(\Gamma)] \equiv \Delta Y_s(S^{\tau}\Gamma)$$

$$= \Omega_{eq,2,s}(S^{\tau}\Gamma) = 0, \forall \Gamma \in D, \forall s > 0$$
(8.6.5)

This leads to the shortened form for the FR for the dimensionless purely dissipative work:

$$\frac{p_{eq,1}(\Delta Y_{\tau} = B)}{p_{eq,2}(\Delta Y_{\tau} = -B)} = \exp[B].$$
(8.6.6)

The derivation of the JE for the irreversible work is trivial:

$$\left\langle \exp[-\Delta Y_{\tau}(\Gamma)] \right\rangle_{eq,1} = \int_{D_{1}} d\Gamma f_{eq,1}(\Gamma) \frac{f_{eq,2}(S^{\tau}\Gamma) dS^{\tau}\Gamma}{f_{eq,1}(\Gamma) d\Gamma}$$
$$= \int_{D_{2}} dS^{\tau}\Gamma f_{eq,2}(S^{\tau}\Gamma)$$
(8.6.7)
$$= 1.$$

Now we are in a position to give a simple but informative derivation of the generalized Jarzynski equality. Comparing (8.1.1) with (8.6.1) we see that the generalized work is related to the purely dissipative work by the relation

$$\exp[\Delta X_{\tau}(\Gamma)] = \exp[\Delta Y_{\tau}(\Gamma)] \frac{Z_{1}}{Z_{2}}$$
(8.6.8)

This means that,

$$\langle \exp[-\Delta X_{\tau}] \rangle_{1} = \langle \exp[-\Delta Y_{\tau}] \rangle_{1} \frac{Z_{2}}{Z_{1}}$$

$$= \frac{Z_{2}}{Z_{1}}$$
(8.6.9)

which is identical to the GJE (8.2.1). For isothermal processes this becomes rather simple: $\Delta X_{\tau} \rightarrow \beta \Delta W_{\tau}$, $\Delta Y_{\tau} \rightarrow \beta \Delta W_{irr,\tau}$ and $\beta \Delta W_{\tau} = \beta \Delta W_{irr,\tau} + \beta \Delta A$.

The generalized purely irreversible work $\Delta Y_{irr,\tau}$, has very similar properties to the dissipation function. However there are important differences. States 1,2 must be in equilibrium and equation (8.6.1) refers to forward and reverse trajectories whereas the dissipation function (as in (8.6.5)) only refers to forward processes.

Nevertheless the dissipative work shares many analogous properties. From the Jarzynski Equality it is easy to compute a bound on the work for a thermodynamic process [39]:

$$1 = \langle \exp[-\Delta Y_{\tau}] \rangle_{1}$$

$$= \exp[-\langle \Delta Y_{\tau} \rangle_{1}] \langle \exp[-\Delta Y_{\tau} + \langle \Delta Y_{\tau} \rangle_{1}] \rangle \qquad (8.6.10)$$

$$\geq \exp[-\langle \Delta Y_{\tau} \rangle_{1}] \langle 1 - \Delta Y_{\tau} + \langle \Delta Y_{\tau} \rangle_{1} \rangle = \exp[-\langle \Delta Y_{\tau} \rangle_{1}]$$

In deriving this result we have used the fact that, $e^x \ge 1 + x$, $\forall x$. The above equation implies that the ensemble average of the purely irreversible dimensionless work is positive except for quasistatic processes:

$$\left\langle \Delta Y_{\tau} \right\rangle_{1} \ge 0 \,. \tag{8.6.11}$$

This is formally analogous to the Second Law Inequality. If a process is reversible the change in the work for the forward path must be equal and opposite to that for the reverse path and the only way this can occur is if the change in the work is zero. This shows that the work is *purely* dissipative as claimed when it was defined. For quasi-static averages

$$\left\langle \Delta Y \right\rangle_{qs} = 0 \tag{8.6.12}$$

In spite of these similarities between the purely irreversible work and the dissipation function, their respective definitions show that they are quite different. This is also evidenced by the fact that the proof of the Second Law Inequality for the purely irreversible work has to be different from that for the dissipation function.

8.7 CROOKS FLUCTUATION THEOREM, AND THE JARZYNSKI EQUALITY

We now give an example of how to apply the GCFT, the GJE and the GCI to an actual statistical mechanical ensemble and system of dynamics. We show that these very general results lead to the canonical forms of the Crooks Fluctuation Theorem (CFT) and the Jarzynski Equality (JE) for the transformation between initial and final equilibrium states with the same values for temperature, volume and number of particles, T,V,N. We assume all systems are T-mixing over the respective phase space domains. The relevant equilibrium distribution function is the canonical distribution function – see §5.3,

$$f(\Gamma;0) = \frac{\exp\left[-\beta H_0(\Gamma)\right]}{Z_c}, \forall \Gamma \in D.$$
(8.7.1)

Definition

We *define* a quantity that we call the *Helmholtz free energy*, A, which is related to logarithm of the canonical partition function, Z_c – see §5.3,4,6,

$$A(\lambda) \equiv -k_{B}T \ln Z_{c}(\lambda)$$

$$= -k_{B}T \ln(\int d\Gamma \exp(-\beta H_{0}(\Gamma, \lambda)))$$
(8.7.2)

In order to transform from the initial equilibrium state, with $\lambda = \lambda_1 = \lambda(0)$ to the final equilibrium state with $\lambda = \lambda_2 = \lambda(\tau)$, the functional form of the system's Hamiltonian may vary parametrically, over the period $0 < t < \tau$, for example

$$H_0(\mathbf{\Gamma}, \lambda(t)) = \sum_{i=1}^N p_i^2 / (2m) + \Phi(\mathbf{q}, \lambda(t)) \text{ where } \Phi(\mathbf{q}, \lambda(t)) \text{ is the interparticle potential.}$$

For $t > \tau$, the Hamiltonian's parametric dependence is fixed at $H_0(\Gamma, \lambda(\tau))$. Over the times $0 < t < \tau$, the ensemble is driven arbitrarily far away from equilibrium, and if the transformation is halted at $t = \tau$, because the system is T-mixing, the system will eventually relax to the unique, new, equilibrium state.

Using eq. (8.1.1), the generalised "work" becomes:

$$\Delta X_{\tau} = \beta [H_0(S^{\tau} \Gamma, \lambda(\tau)) - H_0(\Gamma, \lambda(0))] + \ln \left[\frac{\delta \Gamma}{\delta S^{\tau} \Gamma(\tau)}\right])$$

$$= \beta [H_0(S^{\tau} \Gamma, \lambda(\tau)) - H_0(\Gamma, \lambda(0))] - \int_0^{\tau} ds \Lambda(S^{s} \Gamma))$$

$$= \beta [H_0(S^{\tau} \Gamma, \lambda(\tau)) - H_0(\Gamma, \lambda(0)) + \Delta Q_{\tau}]$$

$$= \beta \Delta W_{\tau}.$$

(8.7.3)

The final equality is obtained from the First "Law" of Thermodynamics, and the equations of motion must satisfy $AI\Gamma$.

Definition

So the generalized dimensionless "work" is in the canonical case, identifiable as β times the *work* performed over a period of time τ . The latter is the change in energy *minus* the change in energy due solely to the exchange of heat:

 $\Delta W = \Delta U + \Delta Q_{therm} = \Delta U_{soi} - \Delta Q_{soi}$. [Remember \dot{Q}_{therm} is defined Eq. (8.4.3), as the heat increase of the thermostat which is the negative of the heat increase of the system.]

We can write this as a standard thermodynamic expression for the so-called First "Law" of thermodynamics

$$dU_{soi} = dQ_{soi} + dW_{soi} \tag{8.7.4}$$

where the subscript "sys" denotes: internal energy of the system, absorbed by the system or work done on the system respectively. Thus the First "Law" of thermodynamics simply states that energy that in this case is the sum of work and heat, is conserved.

If we combine (8.7.4) with (8.5.4) we obtain for quasi static processes,

$$dU_{soi} = TdS_{soi} + pdV_{soi}$$

where we have assumed the work is performed by changing the system volume dV_{soi} , against a non-zero pressure, p.

We note that if at the end of the protocol, $t = \tau$, the system is not in equilibrium it does not matter. Any subsequent relaxation processes will have *no* effect on ΔW . The change in the energy is *exactly* compensated for by the change in the heat, leaving the work unchanged; exactly as proved for the generalized work in (8.1.6).

For a system where the phase space is extended due to the introduction of additional dynamical variables such as the volume or those associated with the thermostat (such as in the case of Nosé-Hoover dynamics [[67]], as detailed below), the work becomes $\Delta W_{\tau} = H_E(S^{\tau}\Gamma', \lambda(\tau)) - H_E(\Gamma', \lambda(0)) - \Delta Q_{\tau}$ where H_E is the Hamiltonian of the extended system [49].

Using (8.1.6) and (8.7.3), the CFT is given as

$$\frac{p_1(\Delta W_{\tau} = B)}{p_2(\Delta W_{\tau} = -B)} = \exp[\beta B] \frac{Z_{c,2}}{Z_{c,1}} = \exp[-\beta(\Delta A - B)]$$
(8.7.5)

where $\Delta A = A_2 - A_1 = A(\lambda(\tau)) - A(\lambda(0))$, and using eq. (8.2.2), the JE is:

$$\left\langle \exp(-\beta \Delta W_{\tau}) \right\rangle = \frac{Z_{c,2}}{Z_{c,1}} = \exp(-\beta \Delta A).$$
 (8.7.6)

The same results are obtained for the canonical distribution when the dynamics are thermostatted by a Gaussian thermostat [68], a Nosé-Hoover thermostat [49] or the dynamics are adiabatic (*i.e.* unthermostatted). For other ensembles and transformations (8.1.1) does not necessarily refer to a work (e.g. see [42,43,46]).

If we now evaluate the cyclic work integral (8.3.3) to the case of constant temperature systems we have

$$\oint \langle dW \rangle = \oint \langle dU_{sys} \rangle - \oint \langle dQ_{sys} \rangle$$

$$= \oint \langle dU_{sys} \rangle + \oint \langle dQ_{therm} \rangle \ge 0$$
(8.7.7)

where as usual the cycle is defined as a cycle in the parameter $\lambda(t)$. The validity of (8.7.7) independent of whether the cycle is periodic. It only requires that the parameter $\lambda(s)$ returns to its initial value. The system at the end of this cycle may have an internal energy that on average is different from its average initial value. This is because if the cycle begins from an equilibrium state but is traversed quickly then at

the end of the cycle the system may be in a nonequilibrium state with an internal energy that is different from the initial equilibrium value.

When the parameter completes a cycle $\lambda(\tau) = \lambda(0)$, the subsequent change in the internal energy dU_{sys} is identical to the heat absorbed by the system dQ_{sys} from the thermostat and there is no further change in the work. So even though, at the end of a parametric cycle, the system is not yet in equilibrium and subsequent thermal relaxation will still to take place changing U_{sys} , the CI (8.4.7) is still valid.

If we run the cycle in reverse (*i.e.* we reverse the direction of the protocol) if the process is reversible then the cyclic integral will take on the opposite sign (by definition) but still obey the inequality (8.4.7). The only way this can be true is if for reversible systems the cyclic integral of the work is zero.

Definition

We call equation (8.7.7) the nonequilibrium cyclic work inequality.

If and only if, the cycle is *periodic* (*i.e.* the system has periodically cycled many times through a periodic protocol and *if* the initial transients have decayed to zero) then $\oint_{P} \langle dU_{sys} \rangle = 0$ and we see that

$$\oint_{P} \langle dQ \rangle \ge 0 \tag{8.7.8}$$

Further, if we now take the thermodynamic limit we see that for large systems the heat absorbed by a large heat reservoir over one periodic cycle cannot be negative. If the process is conducted reversibly then both the forward periodic cycle and the reverse periodic cycle must both be non-negative and therefore both must integrate to zero. In summary: if we surround our system of interest with a *large* equilibrium thermal reservoir at temperature T, and if we do a cycle of work defined by some protocol $\lambda(t)$. We see that over the cycle on average the thermostat absorbs heat rather than looses it.

In small systems where the cycles are of limited duration there *will* be instances when $\oint dW < 0$! Our result says that *on average* CI holds.

8.8 ENTROPY REVISITED

We now look again at equation (22). If our system is subject to a periodic thermal protocol and if the system settles into a periodic cycle, the ensemble averaged heat absorbed by the thermostat (dQ_{therm}) is non-negative $\lim_{t\to\infty} \oint_P \langle dQ_{therm}/T \rangle \ge 0$. If the sign was reversed we would have been able to construct a perpetual motion machine of the second kind. So we have given a proof of the Second "Law" of thermodynamics since Clausius' statement of that "Law" refers to the impossibility of constructing such a machine.

There is a complementary inequality for the system of interest (soi) namely

$$\lim_{t \to \infty} \oint_{P} \left\langle dQ_{soi}/T \right\rangle \le 0.$$
(8.8.1)

If we combine the system of interest and the thermostat we see that for the combined system, the Universe, we find that for all periodic cycles, nonequilibrium and quasistatic

$$\lim_{t \to \infty} [\oint_{P} \langle dQ_{soi}/T \rangle + \lim_{t \to \infty} \oint_{P} \langle dQ_{therm}/T \rangle]$$

=
$$\lim_{t \to \infty} \oint_{P} \langle (dQ_{soi} - dQ_{soi})/T \rangle \qquad (8.8.2)$$

= 0

The second line follows because firstly $dQ_{soi} = -dQ_{therm}$ and secondly, at any point in the cycle the temperature of the underlying equilibrium states for both the thermostat and the system of interest are by the Zeroth Law, equal because both systems, by definition, are in thermal contact. You cannot exchange heat between the two systems without thermal contact!

This appears to be completely consistent with the constancy of the Gibbs entropy for an isolated Hamiltonian system. This also appears to be the final resolution of Orr and Buckingham's objection to Clausius' Inequality when it is a strict inequality. Our derivation gives meaning to the temperature when the system is not in true thermodynamic equilibrium. The temperature is just the equilibrium thermodynamic temperature on the underlying equilibrium state. However because this is identical for the system of interest and the reservoir equation (8.8.2) is the inescapable result.

In spite of (8.8.2) we still have a full set of inequalities for dissipation, for the work and the Clausius Inequality for the reservoir, that each meet of expectations from what is called the Second "Law" of thermodynamics. The difference is that our inequalities are mathematically well defined not just for quasistatic processes but also for strictly nonequilibrium processes too.

8.9 FOR THERMOSTATTED SYSTEMS THE HELMHOLTZ FREE ENERGY IS A CONSTANT OF THE MOTION

Consider a thermostatted system in contact with a isokinetic heat bath¹³. The bath could be much larger than the system of interest, in which case the heat bath could be approximated as being in thermodynamic equilibrium while the system of interest which is in thermal contact with the bath, relaxes towards equilibrium. The heat bath could also be of similar size to the system of interest and therefore it may also be out of equilibrium.

There are no external dissipative fields applied to the system but the initial distribution for the system of interest is not the equilibrium distribution. The nonequilibrium system of interest is relaxing towards equilibrium. From equation (2) since the phase space expansion term is non-zero, the change in the Gibbs entropy for the combined system of interest and reservoir is

$$\dot{S}_{G}(t) = k_{B} \int d\mathbf{\Gamma} f(\mathbf{\Gamma}; t)) \frac{\partial}{\partial \mathbf{\Gamma}} \cdot \dot{\mathbf{\Gamma}} = -k_{B} \left\langle 3N_{th} \alpha(t) \right\rangle, \qquad (8.9.1)$$

where N_{th} is the number of thermostatted particles and α is the usual isokinetic thermostat multiplier⁵. We also know that the rate of change of the total energy of the system of interest and heat bath is

$$\left\langle \dot{H}_{0}(t) \right\rangle = -2K_{th} \left\langle \alpha(t) \right\rangle$$

$$\equiv -3N_{th}k_{B}T_{th} \left\langle \alpha(t) \right\rangle$$
(8.9.2)

where T_{th} is the kinetic temperature of the equilibrium reservoir. However since the system and the reservoir are in thermal contact, T_{th} is of course also equal to the underlying equilibrium thermodynamic temperature of the system of interest and the reservoir.

If we generalize the Helmholtz free energy so that it takes the form for the composite system

$$A \equiv U - T_{th}S \tag{8.9.3}$$

From equations (8.9.1,2,3) we deduce that the Helmholtz free energy for the composite system is a constant of the motion:

$$\dot{A} = \left\langle \dot{H}_0 \right\rangle - T_{th} \dot{S} = 0 \tag{8.9.4}$$

This is precisely the type of system treated by the Relaxation Theorem. In that case however the time integral of the ensemble average of the dissipation is positive for all times and the instantaneous dissipation eventually decays towards zero everywhere in phase space (*i.e.* In the infinite time limit, the system must have relaxed to equilibrium.).

8.10 EXPERIMENTAL DEMONSTRATIONS OF CROOKS AND JARZYNSKI RELATIONS

Deb to update?

The first tests of the JE and CFR were by Liphardt *et al.* [21] who used optical tweezers to extend a DNA-RNA hybrid chain, measuring the work required as the extension proceeded. As well as demonstrating the ability of observing fluctuations that would allow the JE and CFR to be applied, it led to the use of the JE as an experimental tool for studying protein folding and for generating free energy landscapes.

More recently, Collin *et al.* [20] carried out an experiment using the CFR to determine the difference in free energies of an RNA molecule and a mutant that differs by one base pair. The CFR was shown to be useful far from equilibrium where insufficient sampling hampers convergence of the JE.

Hummer and Szabo [104] demonstrated that in single molecule stretching experiments, the JE provides an expression for the work at different times, whereas from an experimental point of view it is of more interest to know the free energy difference between states at different extensions of the molecule. They show how this can be obtained and apply it in experiments.

Douarche *et al.* [105] have verified the CFR and JE for fluctuations in the work of a mechanical oscillator that is in contact with a reservoir and driven by a large external field.

In the future it will be interesting to see how the relationships can be used to advantage in experimental studies or interpretation of experimental results. In this vein, Noy has used [106] the JE to benefit in interpretation of experimental results of chemical force microscopy where the probes of atomic force microscopy are functionalised.