

# Chapter 3

## The Fluctuation Theorem

### 3.1 Background

The Second Law of Thermodynamics states that the entropy production must be greater than or equal to zero. However the Second Law was formulated for macroscopic systems observed over relatively long periods of time (over several milliseconds, seconds or more). Despite this strict limitation, the Second Law's application is often treated as being universal because the size of most systems can be considered large when compared to atomic length and time scales. However several systems of current scientific interest, such as nanomachines and protein motors, operate at length and time scales where the system cannot be considered large. At the nano- and micro-scales the thermal energy available per degree of freedom can be comparable to the work performed by the system, enabling the system to operate in reverse for short periods of time. Clearly classical thermodynamics does not apply to these small systems.

The puzzle of how the time-reversible microscopic equations of classical mechanics lead to the time-irreversible macroscopic equations of thermodynamics has been a paradox since the days of Boltzmann. Boltzmann simply side-stepped this by stating “as soon as one looks at bodies of such small dimension that they contain only very few molecules, the validity of this theorem [the Second Law of Thermodynamics and its description of irreversibility] must cease.” [1, 2] Einstein also attempted to solve this paradox and published two proofs [3, 4, 5]. However these proofs both made an incorrect assumption: “We will have to assume that more probable distributions will always follow less probable ones, that is,  $W$  always increases until the distribution

becomes constant and  $W$  has reached a maximum.” [3] No proof or justification for this assumption was provided. In 1993 the first quantitative description of entropy production in finite systems was given by the Fluctuation Theorem (FT) of Evans *et al* [6]. The FT describes the transition from time-reversible, microscopic equations of motion to irreversible macroscopic behaviour. In its most general form, the theorem provides an analytic expression for the probability that a dissipative flux flows in the direction opposite to that required by the Second Law of Thermodynamics.

The original 1993 FT paper by Evans *et al* provided a heuristic proof of the Steady State FT. In the following year, 1994, Evans and Searles [7] provided a formal proof for the FT by perturbing an equilibrium state. This form of the FT became known as the Transient FT, although its application extends beyond the transient response of the system. In 1995 Gallavotti and Cohen derived another equation, which they called the Gallavotti-Cohen Fluctuation Theorem (GCFT) [8, 9]. In 2004 another form of the GCFT, known as the extended heat theorem, was produced by von Zon and Cohen [10]. The GCFT and the extended heat theorem are different from the FT described throughout this thesis. Throughout this thesis the term FT shall apply solely to the FT derived by Evans *et al* [6, 7].

The FT has been verified in a number of Molecular Dynamics simulations [6, 7] and it is possible to demonstrate the predictions of the theorem in an experiment involving a colloidal particle and an optical trap. Wang *et al* [11] recorded the trajectories of a colloidal particle localised in a translating optical trap and evaluated the entropy production for a large number of trajectories. Wang *et al* were unable to demonstrate the FT directly, but showed that the frequency of trajectories with “entropy production” as opposed to “entropy consumption” agreed quantitatively with the integrated form of the FT. The terms “entropy production” and “entropy consumption” were used loosely: entropy is a macroscopic thermodynamic quantity and not a quantity that can be used to describe energy or work performed along a single path. Wang *et al* used these terms to refer to positive and negative values of the dissipation function, which shall be discussed later in this chapter. Further experimental evidence of the FT is presented in Chapters 4 and 5.

## 3.2 Derivation of the Fluctuation Theorem

Below are two derivations of the FT. The first is a deterministic derivation of the FT following the principles outlined by Evans and Searles in their review article [12]. The second derivation uses the stochastic approach presented in Reid *et al* [13] and in Reid [14]. Furthermore, the FT can be derived using other methods, such as using Lyapunov exponents as shown in [12].

### 3.2.1 Derivation of the FT using Deterministic Dynamics

Consider a system containing  $N$  atomistic or molecular particles, with each particle having 3 momentum coordinates  $\mathbf{p}_i \equiv (p_{xi}, p_{yi}, p_{zi})$  and 3 position coordinates  $\mathbf{q}_i \equiv (q_{xi}, q_{yi}, q_{zi})$ , where  $i$  is the particle number. This corresponds to  $6N$  variables within the system. If each variable is used to represent a spacial dimension then the system can be fully described in  $6N$ -dimensional space, or “phase space”. If the distribution of all  $\mathbf{p}_i$ ’s and  $\mathbf{q}_i$ ’s at a particular time  $s$  in phase space is known then the phase space vector at time  $s$ ,  $\mathbf{\Gamma}(s)$ , can be found by solving the equations of motion, Eqn 3.1. Additionally, one can introduce an external field  $\mathbf{F}_e$  to the system and solve the equations of motion to determine how the system responds to the external field. The equations of motion are

$$\begin{aligned}\dot{\mathbf{q}}_i &= \frac{\mathbf{P}_i}{m_i} + \mathbf{C}_i(\mathbf{\Gamma}) \cdot \mathbf{F}_e, \\ \dot{\mathbf{p}}_i &= \mathbf{F}_i(\mathbf{q}) + \mathbf{D}_i(\mathbf{\Gamma}) \cdot \mathbf{F}_e - S_i\alpha(\mathbf{\Gamma})\mathbf{p}_i,\end{aligned}\tag{3.1}$$

where  $\mathbf{C}_i(\mathbf{\Gamma})$  and  $\mathbf{D}_i(\mathbf{\Gamma})$  are second-rank tensors that describe how  $\mathbf{F}_e$  interacts with each particle,  $\mathbf{F}_i$  is the interatomic force on particle  $i$ ,  $\alpha$  is a thermostat multiplier and  $S_i$  is a switch that determines whether the particle interacts with the thermostat. The thermostat is introduced so that the system remains at a constant temperature despite the effect of external fields. Without the presence of the thermostat simulated particles in an external field keep increasing their temperature. The thermostat introduced in Eqn 3.1 is a synthetic device that does not occur in nature, but is still a useful model. The thermostating particles can be placed a large distance away from the region of interest, ensuring that the system is not “aware” of the thermostat. Furthermore, Williams *et al* [15] showed that the specific details of the thermostat had no effect on the derivation of the FT. Williams *et al* also found

that “when the number of degrees of freedom in the thermostat is large compared to the number of degrees of freedom in the nonequilibrium system, the fluctuation theorem is insensitive to the details of the thermostating mechanism”. In addition Morriss & Evans have shown that thermostats do not affect the linear response of the system [16].

In many systems of interest an initial equilibrium probability density  $f(\mathbf{\Gamma}, 0)$  can be constructed. This probability density propagates according to Liouville’s continuity equation

$$\frac{\partial f(\mathbf{\Gamma}, t)}{\partial t} = -\frac{\partial}{\partial \mathbf{\Gamma}} \cdot [\dot{\mathbf{\Gamma}} f(\mathbf{\Gamma}, t)] \quad (3.2)$$

or in Lagrangian form

$$\begin{aligned} \frac{df(\mathbf{\Gamma}, t)}{dt} &= -f(\mathbf{\Gamma}, t) \frac{d}{d\mathbf{\Gamma}} \cdot \dot{\mathbf{\Gamma}} \\ &\equiv -\Lambda(\mathbf{\Gamma}) f(\mathbf{\Gamma}, t), \end{aligned} \quad (3.3)$$

where  $\Lambda(\mathbf{\Gamma})$  is the phase space compression. The Liouville equation is used to determine how the probability density changes as a function of time.

The probability that a phase space coordinate,  $\mathbf{\Gamma}(s)$ , will be observed within an infinitesimal phase space volume of size

$$\delta V_{\Gamma} = \lim_{\delta \mathbf{q}, \delta \mathbf{p} \rightarrow 0} \delta q_{x1} \delta q_{y1} \delta q_{z1} \delta q_{x2} \dots \delta q_{zN} \delta p_{x1} \dots \delta p_{zN}$$

about the phase point  $\mathbf{\Gamma}(t)$  at time  $t$  is given by

$$P(\delta V_{\Gamma}(\mathbf{\Gamma}(t), t)) = f(\mathbf{\Gamma}(t), t) \delta V_{\Gamma}(\mathbf{\Gamma}(t)), \quad (3.4)$$

where  $f(\mathbf{\Gamma}(s), t)$  is the probability density of observing the phase space configuration  $\mathbf{\Gamma}(s)$  at time  $t$ . By integrating the Liouville equation, Eqn 3.3, the distribution at any time  $t$  can be found if the initial configuration  $f(\mathbf{\Gamma}(0), 0)$  is known,

$$f(\mathbf{\Gamma}(t), t) = \exp \left[ - \int_0^t \Lambda(\mathbf{\Gamma}(s)) ds \right] f(\mathbf{\Gamma}(0), 0). \quad (3.5)$$

By definition the number of trajectories contained within the volume  $V_{\Gamma}(\mathbf{\Gamma}(t))$  re-

mains constant as time evolves, so it is required that

$$V_{\Gamma}(\mathbf{\Gamma}(t)) = \exp \left[ \int_0^t \Lambda(\mathbf{\Gamma}(s)) ds \right] V_{\Gamma}(\mathbf{\Gamma}(0)). \quad (3.6)$$

A system's trajectory,  $\mathbf{\Gamma}(t)$ , from an initial state  $(\mathbf{q}_0, \mathbf{p}_0)$  at time 0 to a state  $(\mathbf{q}_t, \mathbf{p}_t)$  some time  $t$  later is a solution of Newton's equations of motion. As Loschmidt [17] pointed out in 1876, since these equations are time-reversible, for every trajectory  $\mathbf{\Gamma}(t)$  that satisfies the equations of motion there is a time-reversed trajectory or anti-trajectory which is also a solution to the equations. This anti-trajectory, represented by  $\mathbf{\Gamma}^*(t)$ , evolves from an initial state  $(\mathbf{q}_t, -\mathbf{p}_t)$  to a final state  $(\mathbf{q}_0, -\mathbf{p}_0)$ . Thus,  $\mathbf{\Gamma}(t)$  and  $\mathbf{\Gamma}^*(t)$  represent a pair of conjugate trajectories that are related by time-reversal symmetry according to  $\mathbf{\Gamma}(\tau) = \mathbf{\Gamma}^*(t - \tau)$ .

Consider a set of trajectories that start at time  $s = 0$  inside an infinitesimal volume  $\delta V_{\Gamma}(\mathbf{\Gamma}(0)) \equiv \delta \mathbf{q}_0 \delta \mathbf{p}_0$  about an initial state  $(\mathbf{q}_0, \mathbf{p}_0)$ , and the set of conjugate trajectories that start inside the volume element  $\delta V_{\Gamma}(\mathbf{\Gamma}^*(0)) \equiv \delta \mathbf{q}_t \delta \mathbf{p}_t$  about  $(\mathbf{q}_t, -\mathbf{p}_t)$ . Figure 3.1 illustrates these sets of trajectories in the  $(\mathbf{q}, \mathbf{p})$ -space. Every trajectory initiated within  $\delta V_{\Gamma}(\mathbf{\Gamma}(0))$  has a conjugate trajectory initiated in  $\delta V_{\Gamma}(\mathbf{\Gamma}^*(0))$ . The dissipation function is defined as the probability of observing trajectories initiated within  $\delta V_{\Gamma}(\mathbf{\Gamma}(0))$ ,  $P(\delta V_{\Gamma}(\mathbf{\Gamma}(0)), 0)$ , divided by the probability of observing trajectories initiated within  $\delta V_{\Gamma}(\mathbf{\Gamma}^*(0))$ ,

$$\frac{P(\delta V_{\Gamma}(\mathbf{\Gamma}(0)), 0)}{P(\delta V_{\Gamma}(\mathbf{\Gamma}^*(0)), 0)} \equiv \exp(\Omega_t). \quad (3.7)$$

It has also been assumed that the initial distribution of velocities is symmetric under time reversal; that is, the anti-trajectory's momentum distribution at time  $t = 0$  is equal to the forward-trajectory's distribution at time  $t$ , or  $f(\mathbf{\Gamma}^*(0), 0) = f(\mathbf{\Gamma}(t), 0)$ . The forward trajectory is related to the conjugate trajectory through a time reversal mapping. Knowing this the phase-space volume of the forward-trajectory bundle at time  $t$  must be equal to the phase-space volume of the anti-trajectory bundle at time  $t = 0$ ,  $\delta V_{\Gamma}(\mathbf{\Gamma}(t)) = \delta V_{\Gamma}(\mathbf{\Gamma}^*(0))$ . By equating the probabilities of observing the

volume elements of the trajectory and anti-trajectory bundles at time zero one gets

$$\begin{aligned} \frac{P(\delta V_{\Gamma}(\mathbf{\Gamma}(0)), 0)}{P(\delta V_{\Gamma}(\mathbf{\Gamma}^*(0)), 0)} &= \frac{f(\mathbf{\Gamma}(0), 0)\delta V_{\Gamma}(\mathbf{\Gamma}(0))}{f(\mathbf{\Gamma}^*(0), 0)\delta V_{\Gamma}(\mathbf{\Gamma}^*(0))} \\ &= \frac{f(\mathbf{\Gamma}(0), 0)\delta V_{\Gamma}(\mathbf{\Gamma}(0))}{f(\mathbf{\Gamma}(t), 0)\delta V_{\Gamma}(\mathbf{\Gamma}(t))} \\ \exp(\Omega_t) &= \frac{f(\mathbf{\Gamma}(0), 0)}{f(\mathbf{\Gamma}(t), 0)} \exp\left[-\int_0^t \Lambda(\mathbf{\Gamma}(s))ds\right], \end{aligned} \quad (3.8)$$

after substituting in Eqn 3.6.

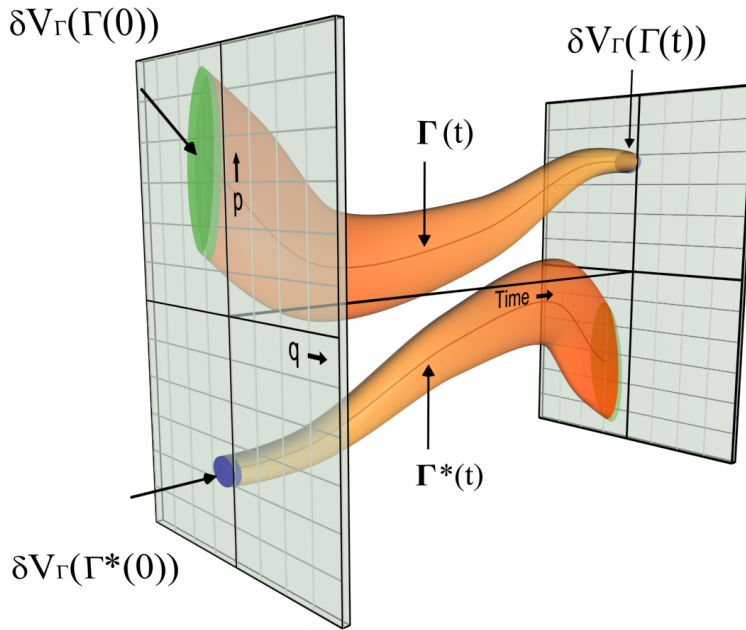


Figure 3.1: An illustration of a set of neighboring trajectories initiated in a volume element  $\delta V_{\Gamma}(\mathbf{\Gamma}(0))$  (top tube) and the corresponding set of anti-trajectories initiated in  $\delta V_{\Gamma}(\mathbf{\Gamma}^*(0))$  (lower tube) in coordinate-momentum  $(\mathbf{q}, \mathbf{p})$  and time,  $t$ , space. The ratio of the probability of observing trajectories initiated within  $\delta V_{\Gamma}(\mathbf{\Gamma}(0))$  to those in  $\delta V_{\Gamma}(\mathbf{\Gamma}^*(0))$  is a measure of the system's irreversibility. Every trajectory in the  $(\mathbf{q}, \mathbf{p})$  space corresponds to a solution of Newton's time-reversible equations of motion. Thus, for every trajectory  $\mathbf{\Gamma}(t)$  that starts at  $(\mathbf{q}_0, \mathbf{p}_0)$  and ends at  $(\mathbf{q}_t, \mathbf{p}_t)$  some time  $t$  later, there exists its time-reversed or anti-trajectory  $\mathbf{\Gamma}^*(t)$ . This anti-trajectory starts at  $(\mathbf{q}_t, -\mathbf{p}_t)$  and ends at  $(\mathbf{q}_0, -\mathbf{p}_0)$  at time  $t$ . For thermostatted Newtonian systems the probability of observing trajectories initiated in the volume element  $\delta V_{\Gamma}(\mathbf{\Gamma}(0))$  is proportional to the size of the volume element and the probability density of trajectories at  $(\mathbf{q}_0, \mathbf{p}_0)$ .

The Fluctuation Theorem is defined according to the probability of observing the dissipation function equal to a particular value  $\Omega_t = A$  as opposed to the probability

of observing the reverse process with  $\Omega_t = -A$

$$\frac{P(\Omega_t = A)}{P(\Omega_t = -A)} = \frac{\sum_{i|\Omega_{t,i}=A} P(\delta V_{\Gamma}(\mathbf{\Gamma}_i(0)), 0)}{\sum_{i|\Omega_{t,i}=-A} P(\delta V_{\Gamma}(\mathbf{\Gamma}_i(0)), 0)}. \quad (3.9)$$

The sum  $\sum_{i|\Omega_{t,i}=A}$  ensures that all trajectories with  $\Omega_t = A$  are used to compute the probability of observing  $\Omega_t = A$  and the sum  $\sum_{i|\Omega_{t,i}=-A}$  ensures that all trajectories with  $\Omega_t = -A$  are used to compute  $P(\Omega_t = -A)$ . Due to the time-reversal mapping each trajectory with  $\Omega_t = -A$  is simply an anti-trajectory when  $\Omega_t = A$ . Therefore it is possible to change the sum to include the anti-trajectories in the following manner

$$\frac{P(\Omega_t = A)}{P(\Omega_t = -A)} = \frac{\sum_{i|\Omega_{t,i}=A} P(\delta V_{\Gamma}(\mathbf{\Gamma}_i(0)), 0)}{\sum_{i|\Omega_{t,i}=A} P(\delta V_{\Gamma}(\mathbf{\Gamma}_i^*(0)), 0)}. \quad (3.10)$$

By re-arranging Eqn 3.7 and substituting into the above equation the following is obtained

$$\frac{P(\Omega_t = A)}{P(\Omega_t = -A)} = \frac{\sum_{i|\Omega_{t,i}=A} P(\delta V_{\Gamma}(\mathbf{\Gamma}_i(0)), 0)}{\sum_{i|\Omega_{t,i}=A} \exp(-\Omega_t = -A) P(\delta V_{\Gamma}(\mathbf{\Gamma}_i(0)), 0)}. \quad (3.11)$$

As  $\exp(-\Omega_t = -A)$  is a constant in the summation, it can be moved in front of the sum. After cancelling the following equation remains

$$\boxed{\frac{P(\Omega_t = A)}{P(\Omega_t = -A)} = \exp(A)}, \quad (3.12)$$

which is the definition of the Fluctuation Theorem found in the literature [11, 12, 18].

### 3.2.2 Derivation of the FT using Stochastic Dynamics

A stochastic approach is often used to characterise the trajectories of a colloidal particle. When the force of inertia is negligible the inertia-less Langevin equation [19] can be used,

$$\xi \frac{d\mathbf{r}(t)}{dt} = \mathbf{f}_{ext}(\mathbf{r}(t)) + \mathbf{g}(t). \quad (3.13)$$

The state of the colloidal system is no longer described by the set of coordinates and momenta of all constituent molecules, but is reduced to only the coordinates of the colloidal particle,  $\mathbf{r}(t) = \mathbf{r}_t$ . The external force exerted on the particle is given

by  $\mathbf{f}_{ext}(\mathbf{r}_t)$ . The many degrees of freedom associated with the solvent molecules are recast into the macroscopic material property of viscosity,  $\eta$ , and the buffeting action of the solvent molecules on the particle is represented by a fluctuating random force,  $\mathbf{g}(t)$ . The random force is assumed to be uncorrelated Gaussian noise with zero mean,  $\langle \mathbf{g}(t)\mathbf{g}(t') \rangle = 2\xi k_B T \delta(t - t')$ , and the drag coefficient for a particle of radius  $a$  is  $\xi = 6\pi\eta a$ . Unlike Newtonian dynamics, this stochastic equation cannot be used to construct conjugate pairs of trajectories through time-reversal as the random force is Markovian. As the particle's position is not unique to any given trajectory, there exists an infinite number of trajectories that originate at  $\mathbf{r}_0$  and a small subset of these arrive at a given destination  $\mathbf{r}_t$  at time  $t$ . Let  $\{\mathbf{r}_0, \mathbf{r}_t\}$  represent those stochastic trajectories that evolve from  $\mathbf{r}_0$  to  $\mathbf{r}_t$ , and let  $\{\mathbf{r}_t, \mathbf{r}_0\}$  represent a conjugate set of trajectories evolving from  $\mathbf{r}_t$  to  $\mathbf{r}_0$ . This is illustrated in Figure 3.2. Let  $\delta V_{r,r}(\{\mathbf{r}_0, \mathbf{r}_t\})$  represent an infinitesimal volume of sets of trajectories that initiate at  $\mathbf{r}_0$  and terminate at  $\mathbf{r}_t$ , then

$$\delta V_{r,r}(\{\mathbf{r}_0, \mathbf{r}_t\}) \equiv \lim_{\delta \mathbf{r}_0, \delta \mathbf{r}_t \rightarrow 0} \delta \mathbf{r}_0 \delta \mathbf{r}_t. \quad (3.14)$$

The probability of observing a stochastic trajectory with a colloidal particle located at an initial position between  $\mathbf{r}_0$  and  $\mathbf{r}_0 + \delta \mathbf{r}_0$  and a final position between  $\mathbf{r}_t$  and  $\mathbf{r}_t + \delta \mathbf{r}_t$  is then

$$P(\delta V_{r,r}(\{\mathbf{r}_0, \mathbf{r}_t\})) = P(\mathbf{r}_0, \mathbf{r}_t) \delta \mathbf{r}_0 \delta \mathbf{r}_t, \quad (3.15)$$

where the normalised distribution  $P(\mathbf{r}_0, \mathbf{r}_t)$  can be expressed using well-known probability distributions and where the volume elements are the same size,  $|\delta \mathbf{r}_0| = |\delta \mathbf{r}_t|$ .

The reversibility of a system subjected to a change at  $t = 0$  does not depend on whether Newtonian mechanics or Langevin stochastics is used to model it. Therefore the dissipation function for stochastic dynamics can be expressed, by analogy to Eqn 3.7, as

$$\Omega_t(\mathbf{r}_0, \mathbf{r}_t) = \ln \left[ \frac{P(\delta V_{r,r}(\{\mathbf{r}_0, \mathbf{r}_t\}))}{P(\delta V_{r,r}(\{\mathbf{r}_t, \mathbf{r}_0\}))} \right]. \quad (3.16)$$

Substituting in Eqn 3.15 this simplifies to

$$\Omega_t(\mathbf{r}_0, \mathbf{r}_t) = \ln \left[ \frac{P(\mathbf{r}_0, \mathbf{r}_t)}{P(\mathbf{r}_t, \mathbf{r}_0)} \right]. \quad (3.17)$$

Eqn 3.17 applies not only to single, colloidal particles, but also to a many-particle

colloidal system. In that case  $\mathbf{r}$  is a  $ND$  vector where  $N$  is the number of colloidal particles embedded in  $D$ -dimensional space. However the probability distributions of Eqn 3.17 cannot be described analytically for many-particle systems.

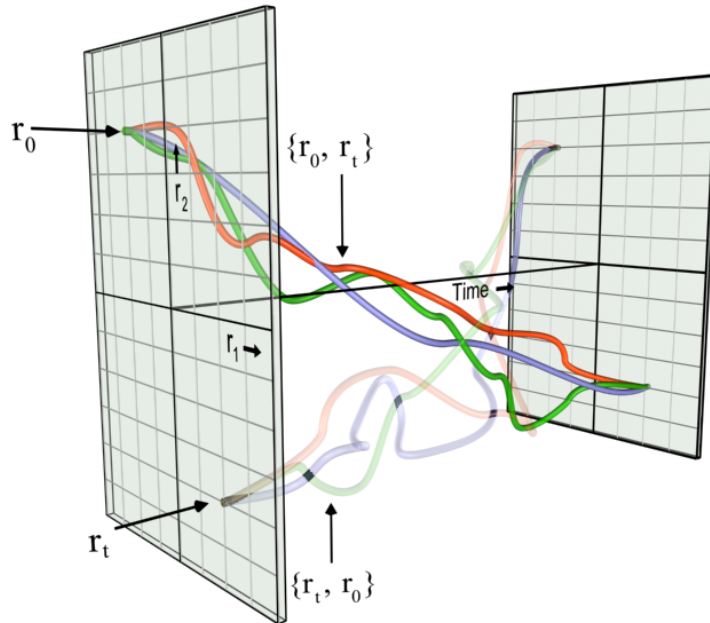


Figure 3.2: An illustration of a subset of the stochastic trajectories that initiate at  $\mathbf{r}_0$  and terminate at  $\mathbf{r}_t$ , denoted by  $\{\mathbf{r}_0, \mathbf{r}_t\}$ , and a corresponding subset of conjugate trajectories,  $\{\mathbf{r}_t, \mathbf{r}_0\}$ . Such stochastic trajectories are represented by the position of the Brownian particle(s) at time  $s$ ,  $\mathbf{r}_s$ , from  $0 < s < t$ . The degrees of freedom of the solvent molecules are reduced to viscosity and a random fluctuating force. The stochastic Langevin equation is not time-reversible and consequently, it is not possible to construct an anti-trajectory that is conjugate to any particular trajectory as is possible using Newtonian dynamics. Also, as position is not unique to any given stochastic trajectory, a bundle of trajectories is determined by both initial and terminating positions,  $\mathbf{r}_0$  and  $\mathbf{r}_t$ . This is different to deterministic trajectories where each trajectory is fully-defined by only one point,  $\Gamma(s)$ , at any time  $s$  along its trajectory.

The probability density associated with observing a set of trajectories of duration  $t$  for which  $\Omega_t = A$  is

$$P(\Omega_t = A) = \int d\mathbf{r}_0 d\mathbf{r}_t \delta(\Omega_t(\mathbf{r}_0, \mathbf{r}_t) - A) P(\mathbf{r}_0, \mathbf{r}_t). \quad (3.18)$$

Likewise, the density associated with a trajectory with  $\Omega_t = -A$  is

$$P(\Omega_t = -A) = \int d\mathbf{r}_0 d\mathbf{r}_t \delta(\Omega_t(\mathbf{r}_0, \mathbf{r}_t) + A) P(\mathbf{r}_0, \mathbf{r}_t) \quad (3.19)$$

$$P(\Omega_t = -A) = \int d\mathbf{r}_0 d\mathbf{r}_t \delta(\Omega_t(\mathbf{r}_t, \mathbf{r}_0) - A) P(\mathbf{r}_0, \mathbf{r}_t) \quad (3.20)$$

where  $\Omega_t(\mathbf{r}_0, \mathbf{r}_t) = -\Omega_t(\mathbf{r}_t, \mathbf{r}_0)$  has been used. Using the definition of  $\Omega_t$  from Eqn 3.17,  $P(\Omega_t = -A)$  can be rewritten as

$$P(\Omega_t = -A) = \int d\mathbf{r}_0 d\mathbf{r}_t \delta(\Omega_t(\mathbf{r}_t, \mathbf{r}_0) - A) \exp(-\Omega_t(\mathbf{r}_t, \mathbf{r}_0)) P(\mathbf{r}_t, \mathbf{r}_0) \quad (3.21)$$

Thus

$$\frac{P(\Omega_t = A)}{P(\Omega_t = -A)} = \frac{\int d\mathbf{r}_0 d\mathbf{r}_t \delta(\Omega_t(\mathbf{r}_0, \mathbf{r}_t) - A) P(\mathbf{r}_0, \mathbf{r}_t)}{\int d\mathbf{r}_0 d\mathbf{r}_t \delta(\Omega_t(\mathbf{r}_t, \mathbf{r}_0) - A) \exp(-\Omega_t(\mathbf{r}_t, \mathbf{r}_0)) P(\mathbf{r}_t, \mathbf{r}_0)}. \quad (3.22)$$

Recognising that the integration variables,  $\mathbf{r}_0$  and  $\mathbf{r}_t$ , can be written arbitrarily, the FT is obtained.

$$\boxed{\frac{P(\Omega_t = A)}{P(\Omega_t = -A)} = \exp(A)}. \quad (3.23)$$

### 3.3 The Second-Law Inequality

The FT is often described as a Second Law-like theorem. In some experimental systems  $\Omega_t$  can be related to the accumulated entropy production,  $\Sigma_t$ , from time  $s = 0$  to  $s = t$  in the system. In these particular cases the FT reduces exactly to the Second Law for large systems observed over long times. This can be shown directly from the FT, which has been re-written as

$$\frac{P(\Omega_t = -A)}{P(\Omega_t = A)} = \exp(-\Omega_t). \quad (3.24)$$

The FT tells us that the probability of observing a negative dissipation function decreases exponentially when compared to the probability of observing a positive dissipation function. As the values of  $\Omega_t$  get larger the probability of observing positive dissipation becomes overwhelmingly likely. In the limiting case of  $\Omega_t \rightarrow \infty$  it is easy to show from Eqn 3.24 that  $P(\Omega_t = -A) = 0$ . This simply states that the dissipation function will not have negative values in large systems, thus reducing to Second Law-like behaviour.

It is also possible to prove that the ensemble average of the dissipation function is always positive using the FT [18]. Starting with the definition

$$\langle \Omega_t \rangle = \int_{-\infty}^{\infty} A P(\Omega_t = A) dA \quad (3.25)$$

and splitting the integral to obtain

$$\begin{aligned}
 \langle \Omega_t \rangle &= \int_0^{\infty} A P(\Omega_t = A) dA + \int_{-\infty}^0 A P(\Omega_t = A) dA \\
 &= \int_0^{\infty} A P(\Omega_t = A) dA + \int_0^{\infty} -A P(\Omega_t = -A) dA \\
 &= \int_0^{\infty} A (P(\Omega_t = A) - P(\Omega_t = -A)) dA
 \end{aligned}$$

By substituting in the FT for  $P(\Omega_t = -A)$  one gets

$$\langle \Omega_t \rangle = \int_0^{\infty} A (P(\Omega_t = A) - \exp(-A) P(\Omega_t = A)) dA$$

which simplifies to

$$\langle \Omega_t \rangle = \int_0^{\infty} A P(\Omega_t = A) (1 - \exp(-A)) dA$$

and then to

$$\langle \Omega_t \rangle = \langle \Omega_t (1 - \exp(-\Omega_t)) \rangle_{\Omega_t \geq 0} \geq 0 \quad \forall t. \quad (3.26)$$

This equation shows that the *average* dissipation function is always positive. The exponential term inside the brackets,  $\exp(-\Omega_t)$ , can only take on values between 0 and 1 as  $\Omega_t \geq 0$ . As such the term  $(1 - \exp(-\Omega_t))$  can only be  $\geq 0$ . This ensures that the value of the entire function is  $\geq 0$ .

By knowing that the ensemble average of the dissipation function equates to entropy in some experimental systems,  $\langle \Omega_t \rangle = \Sigma_t$ , it is clear that the FT predicts a positive entropy, and hence Second Law behaviour. In other systems where  $\langle \Omega_t \rangle \neq \Sigma_t$  there is still a positive quantity that mimics entropy. It is for these reasons that the FT is referred to as a Second Law-like theorem.

### 3.4 The Integrated Fluctuation Theorem

An integrated form of the Fluctuation Theorem (IFT) can easily be derived and is experimentally very useful. It considers the probability ratios of the trajectories

with positive and negative  $\Omega_t$  values rather than ratios of individual values. This theorem is particularly useful when the data is limited and statistical sampling is insufficient to show the FT directly. The probability of observing  $\Omega_t < 0$  as opposed to  $\Omega_t > 0$  is given by

$$\frac{P(\Omega_t < 0)}{P(\Omega_t > 0)} = \frac{\int_0^{\infty} dA P(\Omega_t = -A)}{\int_0^{\infty} dA P(\Omega_t = A)}. \quad (3.27)$$

Substituting in the FT for  $P(\Omega_t = -A)$  yields

$$\frac{P(\Omega_t < 0)}{P(\Omega_t > 0)} = \frac{\int_0^{\infty} dA \exp(-A) P(\Omega_t = A)}{\int_0^{\infty} dA P(\Omega_t = A)}, \quad (3.28)$$

which is the definition of an ensemble average taken over all trajectories where  $\Omega_t > 0$ . Therefore the IFT can be written as

$$\boxed{\frac{P(\Omega_t < 0)}{P(\Omega_t > 0)} = \langle \exp(-\Omega_t) \rangle_{\Omega_t > 0}}. \quad (3.29)$$

Alternately a second form of the IFT can be derived. Starting from Eqn 3.27, and realising that you can re-write it is as

$$\frac{P(\Omega_t < 0)}{P(\Omega_t > 0)} = \frac{\int_{-\infty}^0 dA P(\Omega_t = A)}{\int_{-\infty}^0 dA P(\Omega_t = -A)}. \quad (3.30)$$

Then substitute the FT for  $P(\Omega_t = -A)$  to obtain

$$\frac{P(\Omega_t < 0)}{P(\Omega_t > 0)} = \frac{\int_{-\infty}^0 dA P(\Omega_t = A)}{\int_{-\infty}^0 dA \exp(-A) P(\Omega_t = A)}. \quad (3.31)$$

This equation simplifies to

$$\frac{P(\Omega_t < 0)}{P(\Omega_t > 0)} = \langle \exp(-\Omega_t) \rangle_{\Omega_t < 0}^{-1}. \quad (3.32)$$

This form of the equation is not as valuable as Eqn 3.29 due to the lower frequency of observing anti-trajectories. Consequently many more trajectories (experimental or simulated) need to be collected to ensure that this method works with acceptable error margins.

### 3.5 The Dissipation Function

As has been shown in the previous section there are two alternate definitions for the dissipation function depending on whether one uses stochastic or deterministic derivations. For the remainder of this thesis  $\Omega_t(\Gamma)$  will be used to represent the dissipation function derived deterministically,  $\Omega_t(\mathbf{r}_0, \mathbf{r}_t)$  has been shortened to  $\Omega_t(\mathbf{r})$  and will be used to represent the dissipation function derived stochastically, and  $\Omega_t$  will be used when the description applies to all derivations of the dissipation function.

The dissipation function,  $\Omega_t$ , is a measure of the thermodynamic irreversibility of a system. Specifically it quantifies the amount of energy dissipated to the surroundings. It is an extensive parameter that increases with system size and generally scales with observation time. Furthermore its ensemble average  $\langle \Omega_t \rangle$  is always positive. When applied to large systems observed over long periods of time the probability of observing trajectories with  $\Omega_t < 0$  decreases exponentially and the irreversible behaviour of the Second Law is recovered.

If the probability of observing a trajectory bundle is equal to the probability of observing the anti-trajectory bundle, that is  $P(\delta V_\Gamma(\Gamma(0)), 0) = P(\delta V_\Gamma^*(\Gamma(0)), 0)$  or  $P(\delta V_{\mathbf{r},\mathbf{r}}(\{\mathbf{r}_0, \mathbf{r}_t\})) = P(\delta V_{\mathbf{r},\mathbf{r}}(\{\mathbf{r}_t, \mathbf{r}_0\}))$ , then the system is thermodynamically reversible and  $\Omega_t = 0$ . This measure of a system's irreversibility is consistent with our qualitative observations on how events proceed in time. To illustrate this point, consider a colloidal particle residing in an optical trap whose strength increases instantaneously from  $k_0$  to  $k_1$  at  $t = 0$ . The properties of the particle are such that it undergoes Brownian motion, but is confined within the optical trap. Furthermore, consider a set of trajectories characterised by the particle spiraling into the harmonic trap to become more tightly confined to its center. A corresponding set of anti-trajectories is characterised by the particle spiraling out and escaping away from the trap center, despite the increase in trapping strength. These anti-trajectories are less likely to be observed than their time-forward counterparts, and this probability imbalance is a measure of the system's irreversibility. If these same trajectories are

explored in the limit of  $t \rightarrow 0$ , there is little difference in the frequency of observation of these two sets of trajectories and it is possible to say that such short trajectories are nearly reversible. However, as the time of the particle's trajectory increases, the probabilities of observing forward-spiralling trajectories and their conjugate anti-trajectories become more unbalanced. On average, the anti-trajectories become increasingly infrequent with trajectory duration, and  $\langle \Omega_t \rangle$  becomes larger, reflecting the growing irreversibility of the system.

### 3.6 Steady State FT

In the literature there are two different labels for the FT, depending upon how the theorem is applied. The Transient Fluctuation Theorem or TFT is simply Eqn 3.12 applied to transient systems that evolve from a known initial equilibrium state towards a final (equilibrium or non-equilibrium) steady-state. The Steady-State Fluctuation Theorem or SSFT refers to the steady-state application of the theorem, where the dissipation function is evaluated over trajectory segments of duration  $t$ , sampled wholly under nonequilibrium steady-state conditions. When  $\Omega_t$  is evaluated for steady-state trajectories, the theorem is said to hold only in the long time limit,

$$\lim_{t \rightarrow \infty} \frac{P(\Omega_t^{ss} = -A)}{P(\Omega_t^{ss} = A)} = \exp(-A), \quad (3.33)$$

where the term  $\Omega_t^{ss}$  is the steady state approximation of  $\Omega_t$  (as shown below). The asymptotic limit in the SSFT is a result of *approximations* made in the dissipation function,  $\Omega_t$ . When it is possible to express  $\Omega_t$  *exactly*, the asymptotic limit is no longer needed and the operative theorem under steady-state conditions is the FT. This is explored more fully in Chapter 5, page 75. Thus, while the literature and its nomenclature might indicate that there are two different theorems, the FT is general and applicable to both transient and steady-state conditions. The labels SSFT is used in this thesis to indicate the application of the asymptotic limit in Eqn 3.33.

As  $\Omega_t(\mathbf{\Gamma})$  requires that the relative probabilities of trajectories be made under initial, equilibrium conditions it is not possible to construct exact expressions for  $\Omega_t(\mathbf{\Gamma})$  for trajectory segments of duration  $t$  that are wholly at a nonequilibrium steady-state. However, when the dissipation function is extensive with time, an approximate steady-state dissipation function can be constructed in the following

way.  $\Omega_t(\mathbf{\Gamma})$  can be cast in terms of its instantaneous rate of change,  $\Omega(s)$  at time  $s$ , accumulated from an initial equilibrium state at time  $s = 0$  to some arbitrary time  $t$

$$\Omega_t(\mathbf{\Gamma}) = \int_0^\tau ds \Omega(s) + \int_\tau^t ds \Omega(s). \quad (3.34)$$

Here  $\tau$  has been introduced as an arbitrary ‘‘cut-off’’ time that is sufficiently large so that the system can be regarded as being in steady-state for  $s > \tau$ . This means that  $\Omega_t$  is cast as a sum of transient and steady-state contributions. The steady-state contribution is identified with the steady-state dissipation function,  $\Omega_t^{ss}$ , which can be used to approximate  $\Omega_t$  with an error of order  $\tau$ .

$$\Omega_t(\mathbf{\Gamma}) = \int_0^\tau ds \Omega(s) + \Omega_t^{ss}. \quad (3.35)$$

$$\Omega_t(\mathbf{\Gamma}) \approx \mathcal{O}(\tau) + \Omega_t^{ss} \quad (3.36)$$

It is instructive to express these dissipation functions as time-averages,  $\bar{\Omega}_t = \Omega_t/t$ , such that

$$\bar{\Omega}_t \approx \bar{\Omega}_t^{ss} + \mathcal{O}\left(\frac{\tau}{t}\right). \quad (3.37)$$

This shows that the error invoked by approximating  $\bar{\Omega}_t$  with  $\bar{\Omega}_t^{ss}$  vanishes in the long time limit as  $\tau/t$ . However, the fluctuations in  $\bar{\Omega}_t^{ss}$  also vanish in the long time limit and, in order that the SSFT be of any utility, it is necessary that these fluctuations vanish more slowly than  $\mathcal{O}(\tau/t)$ . The measure  $\bar{\Omega}_t^{ss}$  along the steady-state portion of a trajectory is

$$\bar{\Omega}_t^{ss} \equiv \frac{1}{t} \int_0^t ds \Omega(s), \quad (3.38)$$

which can be re-expressed as a sum of measures taken along contiguous trajectory segments of duration  $\Delta t$

$$\bar{\Omega}_t^{ss} \equiv \frac{1}{t} \sum_i^{t/\Delta t} \int_{(i-1)\Delta t}^{i\Delta t} ds \Omega(s) \quad (3.39)$$

$$\equiv \frac{1}{t} \sum_i^{t/\Delta t} \Omega_{i\Delta t}. \quad (3.40)$$

If  $\Delta t$  is larger than the longest correlation time in the system then the sum  $\sum \Omega_{i\Delta t}$  contains independent measures and, consequently, the variance in the sum is proportional to the number of measures or  $t/\Delta t$ . The factor  $1/t$  in front of the sum

decreases the variance of the sum by a factor  $t^2$ . Thus the standard deviation of the measure  $\bar{\Omega}_t^{ss}$  along a steady-state portion of a trajectory diminishes as  $1/\sqrt{t}$ , slower than  $\mathcal{O}(\frac{\tau}{t})$ . Therefore  $\Omega_t$  can be approximated by the steady-state dissipation function  $\Omega_t^{ss}$  in the FT (see Eqn 3.12), leading to the SSFT

$$\lim_{t \rightarrow \infty} \frac{P(\Omega_t^{ss} = A)}{P(\Omega_t^{ss} = -A)} = \exp(A). \quad (3.41)$$

In this way the SSFT is an approximation to the FT that is accurate in the long time limit. That is, the transient contribution to the dissipation function becomes negligible before the fluctuations in  $\Omega_t^{ss}$  vanish. For some systems described using stochastic dynamics it is possible to construct distributions of trajectories that are wholly in a non-equilibrium steady-state. In such cases it is not necessary to apply the SSFT as the FT can be used directly. However if approximations are made to  $\Omega_t(\mathbf{r})$  then the SSFT will be required.

An integrated form of the SSFT also exists and is derived in the same manner as the IFT:

$$\lim_{t \rightarrow \infty} \frac{P(\Omega_t^{ss} < 0)}{P(\Omega_t^{ss} > 0)} = \langle \exp(-\Omega_t^{ss}) \rangle_{\Omega_t^{ss} > 0}. \quad (3.42)$$

Again the average on the RHS is an ensemble average taken over all trajectories where  $\Omega_t^{ss} > 0$ .

### 3.7 References

- [1] L. Boltzman, Rejoinder to the heat theoretical considerations of Mr. E. Zermelo (1896).
- [2] E. Broda, *Ludwig Boltzmann, Man-Physicist-Philosopher* (Ox Bow Press, 1983).
- [3] A. Einstein, *Annalen der Physik (Leipzig)* **11**, 170 (1903).
- [4] A. Einstein, *Annalen der Physik (Leipzig)* **14**, 354 (1904).
- [5] A. Pais, *'Subtle is the Lord'... the Science and the Life of Albert Einstein* (Oxford University Press, 1983).
- [6] D. Evans, E. Cohen, G. Morriss, *Phys. Rev. Lett.* **71**, 2401 (1993).

- 
- [7] D. Evans, D. Searles, *Phys. Rev. E* **50**, 1645 (1994).
- [8] G. Gallavotti, E. G. D. Cohen, *Phys. Rev. Lett.* **74**, 2694 (1995).
- [9] G. Gallavotti, E. G. D. Cohen, *J. Stat. Phys.* **80**, 931 (1995).
- [10] R. van Zon, E. G. D. Cohen, *Phys. Rev. E* **69**, 56121 (2004).
- [11] G. Wang, E. Sevick, E. Mittag, D. Searles, D. Evans, *Phys. Rev. Lett.* **89**, 050601 (2002).
- [12] D. J. Evans, D. J. Searles, *Adv. Phys.* **51**, 1529 (2002).
- [13] J. C. Reid, D. M. Carberry, G. M. Wang, E. M. Sevick, D. J. Searles, D. J. Evans, *Phys. Rev. E* **70**, 16111 (2004).
- [14] J. C. Reid, Thesis dissertation (in production).
- [15] S. R. Williams, D. J. Searles, D. J. Evans, *Phys. Rev. E* **70**, 066113 (2004).
- [16] D. Evans, G. Morriss, *Statistical Mechanics of Nonequilibrium Liquids* (Academic Press, 1990).
- [17] J. Loschmidt, *J. Sitzungsber. der kais. Akad. d. W. Math. Naturw. II* **73**, 128 (1876).
- [18] D. J. Searles, D. J. Evans, *Aust. J. Chem.* **57**, 1119 (2004).
- [19] M. Doi, S. Edwards, *the Theory of Polymer Dynamics* (Clarendon Press, 1988).

