Advances in Physics, 2002, Vol. 51, No. 7, 1529–1585

The Fluctuation Theorem

Denis J. Evans*
Research School of Chemistry, Australian National University, Canberra, ACT 0200 Australia

and Debra J. Searles
School of Science, Griffith University, Brisbane, Qld 4111 Australia

[Received 1 February 2002; revised 8 April 2002; accepted 9 May 2002]

Abstract

The question of how reversible microscopic equations of motion can lead to irreversible macroscopic behaviour has been one of the central issues in statistical mechanics for more than a century. The basic issues were known to Gibbs. Boltzmann conducted a very public debate with Loschmidt and others without a satisfactory resolution. In recent decades there has been no real change in the situation. In 1993 we discovered a relation, subsequently known as the Fluctuation Theorem (FT), which gives an analytical expression for the probability of observing Second Law violating dynamical fluctuations in thermostatted dissipative non-equilibrium systems. The relation was derived heuristically and applied to the special case of dissipative non-equilibrium systems subject to constant energy 'thermostatting'. These restrictions meant that the full importance of the Theorem was not immediately apparent. Within a few years, derivations of the Theorem were improved but it has only been in the last few of years that the generality of the Theorem has been appreciated. We now know that the Second Law of Thermodynamics can be derived assuming ergodicity at equilibrium, and causality. We take the assumption of causality to be axiomatic. It is causality which ultimately is responsible for breaking time reversal symmetry and which leads to the possibility of irreversible macroscopic behaviour.

The Fluctuation Theorem does much more than merely prove that in large systems observed for long periods of time, the Second Law is overwhelmingly likely to be valid. The Fluctuation Theorem *quantifies* the probability of observing Second Law violations in small systems observed for a short time. Unlike the Boltzmann equation, the FT is completely consistent with Loschmidt's observation that for time reversible dynamics, every dynamical phase space trajectory and its conjugate time reversed 'anti-trajectory', are both solutions of the underlying equations of motion. Indeed the standard proofs of the FT explicitly consider conjugate pairs of phase space trajectories. Quantitative predictions made by the Fluctuation Theorem regarding the probability of Second Law violations have been confirmed experimentally, both using molecular dynamics computer simulation and very recently in laboratory experiments.

Contents PAGE

* To whom correspondence should be addressed. e-mail: evans@rsc.anu.edu.au

Advances in Physics ISSN 0001–8732 print/ISSN 1460–6976 online
http://www.tandf.co.uk/journals
DOI: 10.1080/0001873021015513 3

1. Introduction

1.1. *Overview*

Linear irreversible thermodynamics is a macroscopic theory that combines Navier–Stokes hydrodynamics, equilibrium thermodynamics and Maxwell's postulate of local thermodynamic equilibrium. The resulting theory predicts in the near equilibrium regime, where local thermodynamic equilibrium is expected to be valid, that there will be a 'spontaneous production of entropy' in non-equilibrium systems. This spontaneous production of entropy is characterized by the entropy source strength, σ, which gives the rate of spontaneous production of entropy per unit volume. Using these assumptions it is straightforward to show [1] that

$$\int d\mathbf{r}\, \sigma(\mathbf{r}, t) = \int d\mathbf{r} \left(\sum J_i(\mathbf{r}, t) X_i(\mathbf{r}, t) \right) > 0, \tag{1.1}$$

where $J_i(\mathbf{r}, t)$ is one of the Navier–Stokes hydrodynamic fluxes (e.g. the stress tensor, heat flux vector, ...) at position $\mathbf{r}$ and time t and X_i is the thermodynamic force which is conjugate to $J_i(\mathbf{r}, t)$ (e.g. strain rate tensor divided by the absolute temperature or the gradient of the reciprocal of the absolute temperature, ... respectively). As discussed in reference [1], equation (1.1) is a consequence of exact conservation laws, the Second Law of Thermodynamics and the postulate of local thermodynamic equilibrium.

The conservation laws (of energy, mass and momentum) can be taken as given. The postulate of local thermodynamic equilibrium can be justified by assuming

analyticity of thermodynamic state functions arbitrarily close to equilibrium.† Assuming analyticity, then local thermodynamic equilibrium is obtained from a first order expansion of thermodynamic properties in the irreversible fluxes $\{X_i\}$. We take this 'postulate' as highly plausible—especially on physical grounds.

However, the rationalization of the Second Law of Thermodynamics is a different issue. The question of how irreversible macroscopic behaviour, as summarized by the Second Law of Thermodynamics, can be derived from reversible microscopic equations of motion has remained unresolved ever since the foundation of thermodynamics. In their 1912 Encyclopaedia article [3] the Ehrenfests made the comment: *Boltzmann did not fully succeed in proving the tendency of the world to go to a final equilibrium state ... The very important irreversibility of all observable processes can be fitted into the picture: The period of time in which we live happens to be a period in which the H-function of the part of the world accessible to observation decreases. This coincidence is not really an accident, it is a precondition for the existence of life.* The view that irreversibility is a result of our special place in space–time is still widely held [4]. In the present Review we will argue for an alternative, less anthropomorphic, point of view.

In this Review we shall discuss a theorem that has come to be known as the Fluctuation Theorem (FT). This 'Theorem' is in fact a group of closely related Fluctuation Theorems. One of these theorems states that in a time reversible, thermostatted, ergodic dynamical system, if $\Sigma(t) = -\beta J(t) F_e V = \int_V \mathrm{d}V\, \sigma(\mathbf{r}, t)/k_B$ is the total (extensive) irreversible entropy production rate, where V is the system volume, F_e an external dissipative field, J is the dissipative flux, and $\beta = 1/k_B T$ where T is the absolute temperature of the thermal reservoir coupled to the system and k_B is Boltzmann's constant, then in a non-equilibrium steady state the fluctuations in the time averaged irreversible entropy production $\bar{\Sigma}_t \equiv (1/t) \int_0^t \mathrm{d}s\, \Sigma(s)$, satisfy the relation:

$$\lim_{t\to\infty} \frac{1}{t} \ln \frac{p(\bar{\Sigma}_t = A)}{p(\bar{\Sigma}_t = -A)} = A. \tag{1.2}$$

The notation $p(\bar{\Sigma}_t = A)$ denotes the probability that the value of $\bar{\Sigma}_t$ lies in the range A to $A + \mathrm{d}A$ and $p(\bar{\Sigma}_t = -A)$ denotes the corresponding probability $\bar{\Sigma}_t$ lies in the range $-A$ to $-A - \mathrm{d}A$. The equation is valid for external fields, F_e, of *arbitrary* magnitude. When the dissipative field is weak, the derivation of (1.2) constitutes a proof of the fundamental equation of linear irreversible thermodynamics, namely equation (1.1).

Loschmidt objected to Boltzmann's 'proof' of the Second Law, on the grounds that because dynamics is time reversible, for every phase space trajectory there exists a conjugate time reversed antitrajectory [5] which is also a solution of the equations of motion.‡ If the initial phase space distribution is symmetric under time reversal symmetry (which is the case for all the usual statistical mechanical ensembles) then it was then argued that the Boltzmann H-function (essentially the negative of the

† See: *Comments on the Entropy of Nonequilibrium Steady States* by D. J. Evans and L. Rondoni, Festschrift for J. R. Dorfman [2].

‡ *Apparently, if the instantaneous velocities of all of the elements of any given system are reversed, the total course of the incidents must generally be reversed for every given system.* Loschmidt, reference [5], page 139.

dilute gas entropy), could not decrease monotonically as predicted by the Boltzmann H-theorem.

However, Loschmidt's observation does not deny the possibility of deriving the Second Law. One of the proofs of the Fluctuation Theorem given here, explicitly considers bundles of conjugate trajectory and antitrajectory pairs. Indeed the existence of conjugate bundles of trajectory and antitrajectory segments is central to the proof. By considering the *measure* of the initial phases from which these conjugate bundles originate, we derive a Fluctuation Theorem which confirms that for large systems, or for systems observed for long times, the Second Law of Thermodynamics is likely to be satisfied with overwhelming (exponential) likelihood.

The Fluctuation Theorem is really best regarded as a set of closely related theorems. One reason for this is that the theorem deals with *fluctuations*, and since one expects the statistics of fluctuations to be different in different statistical mechanical ensembles, there is a need for a set of different, but related theorems. A second reason for the diversity of this set of theorems is that some theorems refer to non-equilibrium steady state fluctuations, e.g. (1.2), while others refer to transient fluctuations. If transient fluctuations are considered, the time averages are computed for a finite time from a zero time where the initial distribution function is assumed to be known: for example it could be one of the equilibrium distribution functions of statistical mechanics.

Even when the time averages are computed in the steady state, they could be computed for an ensemble of experiments that started from a known, ergodically consistent, distribution in the (long distant) past or, if the system is ergodic, time averages could be computed at different times during the course of a *single* very long phase space trajectory†. As we shall see, the Steady State Fluctuation Theorems (SSFT) are asymptotic, being valid in the limit of long averaging times, while the corresponding Transient Fluctuation Theorems (TFT) are *exact* for arbitrary averaging times. The TFT can therefore be written, $[p(\bar{\Sigma}_t = A)]/[p(\bar{\Sigma}_t = -A)] = \exp[At]$, $\forall\, t > 0$.

We can illustrate the SSFT expressed in equation (1.2) very simply. Suppose we consider a shearing system with a constant positive strain rate, $\gamma \equiv \partial u_x/\partial y$, where u_x is the streaming velocity in the x-direction. Suppose further that the system is of fixed volume and is in contact with a heat bath at a fixed temperature T. Time averages of the xy-element of the pressure tensor, $\bar{P}_{xy,t}$, are proportional to the negative of the time-averaged entropy production. A histogram of the fluctuations in the time-averaged pressure tensor element could be expected as shown in figure 1.1. In accord with the Second Law, the mean value for $\bar{P}_{xy,t}$ is negative. The distribution is approximately Gaussian. As the number of particles increases or as the averaging time increases we expect that the variance of the histogram would decrease.

For the parameters studied in this example, the wings of the distribution ensure that there is a significant probability of finding data for which the time averaged entropy production is negative. The SSFT gives a mathematical relationship for the ratio of peak heights of pairs of data points which are symmetrically distributed about zero on the x-axis, as shown in figure 1.1. The SSFT says that it becomes exponentially likely that the value of the time-averaged entropy production will be positive rather than negative. Further, the argument of this exponential grows

† The equivalence of these two averages is the definition of an ergodic system.

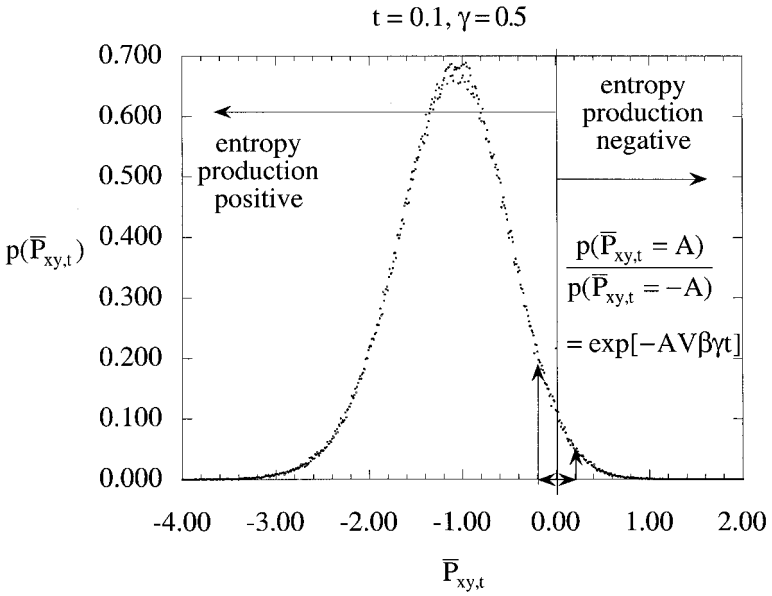


Figure 1.1. A histogram showing fluctuations in the time-averaged shear stress for a system undergoing Couette flow.

linearly with system size and with the duration of the averaging time. In either the large system or long time limit the SSFT predicts that the Second Law will hold absolutely and that the probability of Second Law violations will be zero.

If $\langle \ldots \rangle_{\bar{\Sigma}_t > 0}$ denotes an average over all fluctuations in which the time-integrated entropy production is positive, then one can show that from the transient form of equation (1.2), that

$$\left[\frac{p(\bar{\Sigma}_t > 0)}{p(\bar{\Sigma}_t < 0)}\right] = \langle \exp(-\bar{\Sigma}_t t) \rangle_{\bar{\Sigma}_t < 0} = \langle \exp(-\bar{\Sigma}_t t) \rangle^{-1}_{\bar{\Sigma}_t > 0} > 1 \tag{1.3}$$

gives the ratio of probabilities that for a finite system observed for a finite time, the Second Law will be satisfied rather than violated (see section 4.5). The ratio increases approximately exponentially with increased time of observation, t, *or* with system size (since Σ is extensive). [There is a corresponding steady state form of (1.3) which is valid asymptotically, in the limit of long averaging times.] We will refer to the various transient or steady state forms of (1.3) as transient or steady state, Integrated Fluctuation Theorems (IFTs).

The Fluctuation Theorems are important for a number of reasons:

(1) they quantify probabilities of violating the Second Law of Thermodynamics;
(2) they are verifiable in a laboratory;
(3) the SSFT can be used to derive the Green–Kubo and Einstein relations for linear transport coefficients;
(4) they are valid in the nonlinear regime, far from equilibrium, where Green–Kubo relations fail;
(5) local versions of the theorems are valid;

(6) stochastic versions of the theorems have been derived [6–11];
(7) TFT and SSFT can be derived using the traditional methods of non-equilibrium statistical mechanics and applied to ensembles of transient or steady state trajectories;
(8) the Sinai–Ruelle–Bowen (SRB) measure from the modern theory of dynamical systems can be used to derive an SSFT for a single very long dynamical trajectory characteristic of an isochoric, constant energy steady state;
(9) FTs can be derived which apply *exactly* to transient trajectory segments while SSFTs can be derived which apply asymptotically ($t \to \infty$) to non-equilibrium steady states;
(10) FTs can be derived for dissipative systems under a variety of thermodynamic constraints (e.g. thermostatted, ergostatted or unthermostatted, constant volume or constant pressure), and
(11) a TFT can be derived which proves that an ensemble of non-dissipative purely Hamiltonian systems will with overwhelming likelihood, *relax* from any arbitrary initial (non-equilibrium) distribution towards the appropriate equilibrium distribution.

Point (11). is the analogue of Boltzmann's H-theorem and can be thought of as a proof of Le Chatelier's Principle [12, 13].

In this Review we will concentrate on the ensemble versions of the TFT and SSFT. A detailed account of the application of the SRB measure to the statistics of a single dynamical trajectory has been given elsewhere by Gallavotti and Cohen (GC) [14, 15]. However, it is true to say that for this more strictly *dynamical* derivation of the SSFTs there are many unanswered questions. For example, essentially nothing is known of the application of the SRB measure and GC methods to dynamical trajectories which are characteristic of systems under various macroscopic thermodynamic constraints (e.g. constant temperature or pressure). All the known results seem to be applicable only to isochoric, constant energy systems. Also an hypothesis which is essential to the GC proof of the SSFT, the so-called *chaotic hypothesis*, is little understood in terms of how it applies to dynamical systems that occur in nature. FT have also been developed for general Markov processes by Lebowitz and Spohn [7] and a derivation of FT using the Gibbs formalism has been considered in detail by Maes and co-workers [8–10].

1.2. *Reversible dynamical systems*

A typical experiment of interest is conveniently summarized by the following example. Consider an electrical conductor (a molten salt for example) subject at say $t = 0$, to an applied electric field, $\mathbf{E}$. We wish to understand the behaviour of this system from an atomic or molecular point of view. We assume that classical mechanics gives an adequate description of the dynamics. Experimentally we can only control a small number of variables which specify the initial state of the system. We might only be able to control the initial temperature $T(0)$, the initial volume $V(0)$ and the number of atoms in the system, N, which we assume to be constant. The microscopic state of the system is represented by a phase space vector of the coordinates and momenta of all the particles, in an exceedingly high dimensional space—phase space—$\{\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_N, \mathbf{p}_1, \ldots, \mathbf{p}_N\} \equiv (\mathbf{q}, \mathbf{p}) \equiv \Gamma$ where $\mathbf{q}_i, \mathbf{p}_i$ are the position and conjugate momentum of particle i. There are a huge number of initial

microstates $\Gamma(0)$, that are consistent with the initial macroscopic specification of the system $(T(0), V(0), N)$.

We could study the macroscopic behaviour of the macroscopic system by taking just one of the huge number of microstates that satisfy the macroscopic conditions, and then solving the equations of motion for this single microscopic trajectory. However, we would have to take care that our microscopic trajectory $\Gamma(t)$, was a *typical* trajectory and that it did not behave in an exceptional way. The best way of understanding the macroscopic system would be to select a set of N_Γ initial phases $\{\Gamma_j(0), j = 1, \ldots, N_\Gamma\}$ and compute the time dependent properties of the macroscopic system by taking a time dependent average $\langle A(t)\rangle$ of a phase function $A(\Gamma)$ over the *ensemble* of time evolved phases

$$\langle A(t)\rangle = \sum_{j=1}^{N_\Gamma} A(\Gamma_j(t))/N_\Gamma.$$

Indeed, repeating the experiment with initial states that are consistent with the specified initial conditions is often what an experimentalist attempts to do in the laboratory. Although the concept of ensemble averaging seems natural and intuitive to experimental scientists, the use of ensembles has caused some problems and misunderstandings from a more purely mathematical viewpoint.

Ensembles are well known to equilibrium statistical mechanics, the concept being first introduced by Maxwell. The use of ensembles in non-equilibrium statistical mechanics is less widely known and understood.† For our experiment it will often be convenient to choose the initial ensemble which is represented by the set of phases $\{\Gamma_j(0), j = 1, \ldots, N_\Gamma\}$, to be one of the standard ensembles of equilibrium statistical mechanics. However, sometimes we may wish to vary this somewhat. In any case, in all the examples we will consider, the initial ensemble of phase vectors will be characterized by a *known* initial N-particle distribution function, $f(\Gamma, t)$, which gives the probability, $f(\Gamma, t)\,\mathrm{d}\Gamma$, that a member of the ensemble is within some small neighbourhood $\mathrm{d}\Gamma$ of a phase Γ at time t, after the experiment began.

The electric field does work on the system causing an electric current, $\mathbf{I}$, to flow. We expect that at an arbitrary time t after the field has been applied, the ensemble averaged current $\langle \mathbf{I}(t)\rangle$ will be in the direction of the field; that the work performed on the system by the field will generate heat—Ohmic heating, $\langle \mathbf{I}(t)\rangle \cdot \mathbf{E}$; and that there will be a 'spontaneous production of entropy' $\langle \Sigma(t)\rangle = \langle \mathbf{I}(t) \cdot \mathbf{E}/T(t)\rangle$. It will frequently be the case that the electrical conductor will be in contact with a heat reservoir which fixes the temperature of the system so that $T(t) = T(0) = T$, $\forall t$. The particles in this system constitute a typical time reversible dynamical system.

We are interested in an number of problems suggested by this experiment:

(1) How do we reconcile the 'spontaneous production of entropy', with the time reversibility of the microscopic equations of motion?
(2) For a given initial phase $\Gamma_j(0)$ which generates some time dependent current $\mathbf{I}_j(t)$, can we generate Loschmidt's conjugate *antitrajectory* which has a time-reversed electric current?
(3) Is there anything we can say about the deviations of the behaviour of individual ensemble members, from the average behaviour?

† For further background information on non-equilibrium statistical mechanics see reference [16].

In general, it is convenient to consider equations of motion for an N-particle system, of the form,

$$\left.\begin{aligned}\dot{\mathbf{q}}_i &= \frac{\mathbf{p}_i}{m} + \mathbf{C}_i(\Gamma)\cdot\ \mathbf{F}_e \\ \dot{\mathbf{p}}_i &= \mathbf{F}_i(\mathbf{q}) + \mathbf{D}_i(\Gamma)\cdot\ \mathbf{F}_e - S_i\alpha(\Gamma)\mathbf{p}_i,\end{aligned}\right\} \tag{1.4}$$

where $\mathbf{F}_e$ is the dissipative external field that couples to the system via the phase functions $\mathbf{C}(\Gamma)$ and $\mathbf{D}(\Gamma)$, $\mathbf{F}_i(\mathbf{q}) = -\partial\Phi(\mathbf{q})/\partial\mathbf{q}_i$ is the interatomic force on particle i (and $\Phi(\mathbf{q})$ is the interparticle potential energy), and the last term $-S_i\alpha(\Gamma)\mathbf{p}_i$ is a deterministic time reversible thermostat used to add or remove heat from the system [16]. The thermostat multiplier is chosen using Gauss's Principle of Least Constraint [16], to fix some thermodynamic constraint (e.g. temperature or energy). The thermostat employs a switch, S_i, which controls how many and which particles are thermostatted.

The model system could be quite realistic with only some particles subject to the external field. For example, some fluid particles might be charged in an electrical conduction experiment, while other particles may be chemically distinct, being solid at the temperatures and densities under consideration. Furthermore these particles may form the thermal boundaries or walls which thermostat and 'contain' the electrically charged particles fluid particles inside a conduction cell. In this case $S_i = 1$ only for wall particles and $S_i = 0$ for all the fluid particles. This would provide a realistic model of electrical conduction.

In other cases we might consider a homogeneous thermostat where $S_i = 1, \forall i$. It is worth pointing out that as described, equations (1.4) are time reversible and heat can be both absorbed and given out by the thermostat. However, in accord with the Second Law of Thermodynamics, in dissipative dynamics the *ensemble averaged* value of the thermostat multiplier is positive at all times, no matter how short, $\langle\alpha(t)\rangle > 0, \forall t > 0$.

One should not confuse a real thermostat composed of a very large (in principle, infinite) number of particles with the purely mathematical—albeit convenient—term α. In writing equation (1.4) it is assumed that the momenta $\mathbf{p}_i$ are peculiar (i.e. measured relative to the local streaming velocity of the fluid or wall). The thermostat multiplier may be chosen, for instance, to fix the internal energy of the system

$$H_0 \equiv \sum_{i:S_i=0}\left[p_i^2/2m + 1/2\sum_j \Phi(\mathbf{q})\right],$$

in which case we speak of ergostatted dynamics, or we can constrain the peculiar kinetic energy of the wall particles

$$K_W \equiv \sum_{S_i=1} p_i^2/2m = d_C N_W k_B T_w/2, \tag{1.5}$$

with $N_W = \sum S_i$, in which case we speak of isothermal dynamics. The quantity T_W defined by this relation is called the kinetic temperature of the wall, and d_C is the Cartesian dimension of the system. For homogeneously thermostatted systems, T_W becomes the kinetic temperature of the whole system and N_W becomes just the number of particles N, in the whole system.

For ergostatted dynamics, the thermostat multiplier, α, is chosen as the instantaneous solution to the equation,

$$\dot{H}_0(\Gamma) \equiv -\mathbf{J}(\Gamma)V \cdot \mathbf{F}_e - 2K_W(\Gamma)\alpha(\Gamma) = 0, \tag{1.6}$$

where $\mathbf{J}$ is the dissipative flux due to $\mathbf{F}_e$ defined as

$$\dot{H}_0^{ad} \equiv -\mathbf{J}V \cdot \mathbf{F}_e \equiv -\sum \left[\frac{\mathbf{p}_i}{m} \cdot \mathbf{D}_i - \mathbf{F}_i \cdot \mathbf{C}_i\right] \cdot \mathbf{F}_e, \tag{1.7}$$

$\dot{H}_0^{ad}$ is the adiabatic time derivative of the internal energy and V is the volume of the system. Equation (1.6) is a statement of the First Law of Thermodynamics for an ergostatted non-equilibrium system. The energy removed from (or added to) the system by the ergostat must be balanced instantaneously by the work done on (or removed from) the system by the external dissipative field, $\mathbf{F}_e$. For ergostatted dynamics we solve (1.6) for the ergostat multiplier and substitute this phase function into the equations of motion. For thermostatted dynamics we solve an equation which is analogous to (1.6) but which ensures that the kinetic temperature of the walls or system, is fixed [16]. The equations of motion (1.4) are reversible where the thermostat multiplier is defined in this way.

One might object that our analysis is compromised by our use of these artificial (time reversible) thermostats. However, the thermostat can be made arbitrarily remote from the system of physical interest [17]. If this is the case, the system cannot 'know' the precise details of how entropy was removed at such a remote distance. This means that the results obtained for the system using our simple mathematical thermostat must be the same as those we would infer for the same system surrounded (at a distance) by a real physical thermostat (say with a huge heat capacity). These mathematical thermostats may be unrealistic, however in the final analysis they are very convenient but ultimately irrelevant devices.

Using conventional thermodynamics, the total rate of entropy absorbed (or released!) by the ergostat is the energy absorbed by the ergostat divided by its absolute temperature,

$$\Sigma(t) = 2K_W(\Gamma)\alpha(\Gamma)/T_W(t) = d_C N_W k_B \alpha(t) = -\mathbf{J}(t)V \cdot \mathbf{F}_e/T_W(t). \tag{1.8}$$

The entropy flowing into the ergostat results from a continuous generation of entropy in the dissipative system.

The exact equation of motion for the N-particle distribution function is the time reversible Liouville equation

$$\frac{\partial f(\Gamma, t)}{\partial t} = -\frac{\partial}{\partial \Gamma} \cdot [\dot{\Gamma} f(\Gamma, t)], \tag{1.9}$$

which can be written in Lagrangian form,

$$\frac{\mathrm{d}f(\Gamma, t)}{\mathrm{d}t} = -f(\Gamma, t)\frac{\mathrm{d}}{\mathrm{d}\Gamma} \cdot \dot{\Gamma} \equiv -\Lambda(\Gamma) f(\Gamma, t). \tag{1.10}$$

This equation simply states that the time reversible equations of motion conserve the number of ensemble members, N_Γ. The presence of the thermostat is reflected in the phase space compression factor, $\Lambda(\Gamma) \equiv \partial\dot{\Gamma} \cdot /\partial\Gamma$, which is to first order in N, $\Lambda = -d_C N_W \alpha$. Again one might wonder about the distinction between Hamiltonian dynamics of realistic systems, where the phase space compression factor is identically zero and artificial ergostatted dynamics where it is non-zero. However, as Tolman pointed out [18], in a purely Hamiltonian system, the neglect of 'irrelevant' degrees of freedom (as in thermostats or for example by neglecting solvent degrees of freedom in a colloidal or Brownian system) inevitably results in a non-zero phase

space compression factor for the remaining 'relevant' degrees of freedom. Equation (1.8) shows that there is an exact relationship between the entropy absorbed by an ergostat and the phase space compression in the (relevant) system.

1.3. *Example: SLLOD equations for planar Couette flow*

A very important dynamical system is the standard model for planar Couette flow—the so-called SLLOD equations for shear flow. Consider N particles under shear. In this system the external field is the shear rate, $\partial u_x/\partial y = \gamma(t)$ (the y-gradient of the x-streaming velocity), and the xy-element of the pressure tensor, P_{xy}, is the dissipative flux, J [16]. The equations of motion for the particles are given by the the so-called thermostatted SLLOD equations,

$$\dot{\mathbf{q}}_i = \mathbf{p}_i/m + \mathbf{i}\gamma y_i, \quad \dot{\mathbf{p}}_i = \mathbf{F}_i - \mathbf{i}\gamma p_{yi} - \alpha \mathbf{p}_i. \tag{1.11}$$

Here, $\mathbf{i}$ is a unit vector in the positive x-direction. At arbitrary strain rates these equations give an exact description of adiabatic (i.e. unthermostatted) Couette flow. This is because the adiabatic SLLOD equations for a step function strain rate $\partial u_x(t)/\partial y = \gamma(t) = \gamma\Theta(t)$, are equivalent to Newton's equations after the impulsive imposition of a linear velocity gradient at $t = 0$ (i.e. $\mathrm{d}\mathbf{q}_i(0^+)/\mathrm{d}t = \mathrm{d}\mathbf{q}_i(0^-)/\mathrm{d}t + \mathbf{i}\gamma y_i$) [16]. There is thus a remarkable subtlety in the SLLOD equations of motion. If one starts at $t = 0^-$, with a canonical ensemble of systems then at $t = 0^+$, the SLLOD equations of motion transform this initial ensemble into the local equilibrium ensemble for planar Couette flow at a shear rate γ. The adiabatic SLLOD equations therefore give an *exact* description of a boundary driven thermal transport process, although the shear rate appears in the equations of motion as a fictitious (i.e. unnatural) external field. This was first pointed out by Evans and Morriss in 1984 [19].

At low Reynolds number, the SLLOD momenta, $\mathbf{p}_i$, are peculiar momenta and α is determined using Gauss's Principle of Least Constraint to keep the internal energy, $H_0 = \Sigma p_i^2/2m + \Phi(\mathbf{q})$, fixed [16]. Thus, for a system subject to pair interactions†

$$\Phi(\mathbf{q}) = \sum_{i=1}^{N-1} \sum_{j>i}^{N} \phi(q_{ij}),$$

$$\alpha = -\gamma \left[\sum_{i=1}^{N} p_{xi} p_{yi}/m - 1/2 \sum_{i,j}^{N} x_{ij} F_{yij} \right] \Big/ \sum_{i=1}^{N} \mathbf{p}_i^2/m$$

$$\equiv -P_{xy}\gamma V \Big/ \sum_{i=1}^{N} \mathbf{p}_i^2/m = -P_{xy}\gamma V/2K(\mathbf{p}), \tag{1.12}$$

where F_{yij} is the y-component of the intermolecular force exerted on particle i by j and $x_{ij} \equiv x_j - x_i$. The corresponding isokinetic form for the thermostat multiplier is,

$$\alpha = \frac{\sum_{i}^{N} \mathbf{F}_i \cdot \mathbf{p}_i - \gamma \left[\sum_{i=1}^{N} p_{xi} p_{yi}/m \right]}{\sum_{i=1}^{N} \mathbf{p}_i^2/m}. \tag{1.13}$$

† We limit ourselves to pair interactions only for reasons of simplicity.

The ergostatted and thermostatted SLLOD equations of motion, (1.11), (1.12), (1.13), are time reversible [16]. In the weak flow limit these equations yield the correct Green–Kubo relation for the linear shear viscosity of a fluid [16]. We have also proved that in this limit, the linear response obtained from the equations of motion, or equivalently from the Green–Kubo relation are identical to leading order in N the number of particles. In the far-from-equilibrium regime, Brown and Clarke [20] have shown that the results for homogeneously thermostatted SLLOD dynamics are indistinguishable from those for boundary thermostatted shear flow, up to the limiting shear rate above which a steady state for boundary thermostatted systems is not stable.†

1.4. *Lyapunov instability*

The Lyapunov exponents are used in dynamical systems theory to characterize the stability of phase space trajectories. If one imagines two systems that evolve in time from phase vectors $\Gamma_1(0), \Gamma_2(0)$ which initially are very close together $|\Gamma_1(0) - \Gamma_2(0)| \equiv \delta\Gamma(0) \to 0$, then one can ask how the separation between these two systems evolves in time. Oseledec's Theorem says for non-integrable systems under very general conditions, that the separation vector asymptotically grows or shrinks *exponentially* in time. Of course this does not happen for integrable systems, but then most real systems are not integrable. A system is said to be chaotic if the separation vector asymptotically *grows* exponentially with time. Most systems in Nature are chaotic: the world weather and high Reynolds Number flows are chaotic. In fact all systems that obey thermodynamics are chaotic. In 1990 the first of a remarkable set of relationships between phase space stability measures (i.e. Lyapunov exponents) and thermophysical properties were discovered by Evans *et al.* [21] and Gaspard and Nicolis [22]. More recently Lyapunov exponents have been used to assign dynamical probabilities to the observation of phase space trajectory segments [14, 15, 23]. This is something quite new to statistical mechanics where hitherto probabilities had been given (only for equilibrium systems!) on the basis of the value of the Hamiltonian (i.e. the weights are static).

Suppose the equations of motion (1.4), are written

$$\dot{\mathbf{\Gamma}} = \mathbf{G}(\mathbf{\Gamma}, t). \tag{1.14}$$

It is trivial to see that the equation of motion for an infinitesimal phase space separation vector, $\mathrm{d}\mathbf{\Gamma}$, can be written as:

$$\mathrm{d}\dot{\mathbf{\Gamma}} = \mathbf{T}(\mathbf{\Gamma}, t) \cdot \mathrm{d}\mathbf{\Gamma}, \tag{1.15}$$

where $\mathbf{T} \equiv \partial\mathbf{G}(\mathbf{\Gamma}, t)/\partial\mathbf{\Gamma}$ is the stability matrix for the flow. The propagation of the tangent vectors is therefore given by,

$$\mathrm{d}\mathbf{\Gamma}(t) = \mathbf{L}(t) \cdot \mathrm{d}\mathbf{\Gamma}(0), \tag{1.16}$$

where the propagator is:

$$\mathbf{L}(t) = \exp_{\mathrm{L}}\left(\int_0^t \mathrm{d}s\, \mathbf{T}(s)\right) \tag{1.17}$$

† Entropy production is extensive $= O(N)$ while entropy absorption by the thermostat $= O(N^{2/3})$. So for any given system there is a limiting shear rate beyond which boundary thermostatting is not possible.

and $\exp_L$ is a left time-ordered exponential. The time evolution of these tangent vectors is used to determine the Lyapunov spectrum for the system. The Lyapunov exponents thus represent the rates of divergence of nearby points in phase space.

If $d\Gamma_i(0)$ is an eigenvector of $\mathbf{L}(t)^T \cdot \mathbf{L}(t)$ and if the Lyapunov exponents are defined as [24]:

$$\{\lambda_i; i = 1, \ldots, 2dN\} = \lim_{t\to\infty} \frac{1}{2t} \ln [\text{eigenvalues } (\mathbf{L}(t)^T \cdot \mathbf{L}(t))], \tag{1.18}$$

then the Lyapunov exponents describe the growth rates of the set of orthogonal tangent vectors (eigenvectors of $(\mathbf{L}(t)^T \cdot \mathbf{L}(t))$), $\{d\Gamma_i(t); i = 1, 2d_C N\}$,

$$\begin{aligned}\lim_{t\to\infty} \frac{1}{2t} \ln \frac{|d\Gamma_i(t) \cdot d\Gamma_i(t)|}{|d\Gamma_i(0) \cdot d\Gamma_i(0)|} &= \lim_{t\to\infty} \frac{1}{2t} \ln \frac{|d\Gamma_i(0)^T \cdot \mathbf{L}(t)^T \cdot \mathbf{L}(t) \cdot d\Gamma_i(0)|}{|d\Gamma_i(0) \cdot d\Gamma_i(0)|} \\ &= \frac{1}{2t} \ln \frac{|d\Gamma_i(0)^T \cdot \exp[2\lambda_i t]\mathbf{1} \cdot d\Gamma_i(0)|}{|d\Gamma_i(0) \cdot d\Gamma_i(0)|} \\ &= \lambda_i, \quad i = 1, \ldots, 2d_C N. \end{aligned} \tag{1.19}$$

By convention the exponents are ordered such that $\lambda_1 > \lambda_2 > \cdots > \lambda_{2d_C N}$. It can be shown that the Lyapunov exponents are independent of the metric used to measure phase space lengths.

In order to calculate the Lyapunov spectrum, one does not normally use equation (1.18). Benettin *et al*. developed a technique whereby the finite but small displacement vectors are periodically rescaled and orthogonalized during the course of a solution of the equations of motion [25, 26]. Hoover and Posch [27] pointed out that this rescaling and orthogonalization can be carried out continuously by introducing constraints to the equations of motion of the tangent vectors [28]. With this modification, orthogonality and tangent vector length are maintained at all times during the calculation.

In theory, the $2dN$ eigenvalues of the real symmetric matrix $\mathbf{L}(t)^T \cdot \mathbf{L}(t)$ can also be used to calculate the Lyapunov spectrum in the limit $t \to \infty$. Since $\mathbf{L}$ is dependent only on the mother trajectory, calculation of the Lyapunov exponents from the eigenvalues of $\mathbf{L}(t)^T \cdot \mathbf{L}(t)$ does not require the solution of $2dN$ tangent trajectories as in the methods mentioned in the previous paragraph. However, after a short time, numerical difficulties are encountered using this method due to the enormous difference in the magnitude of the eigenvalues of the $\mathbf{L}(t)^T \cdot \mathbf{L}(t)$ matrix.† The use of QR decompositions (where where $\mathbf{L} = \mathbf{Q} \cdot \mathbf{R}$ and $\mathbf{R}$ is a real upper triangular matrix with positive diagonal elements and $\mathbf{Q}$ is a real orthogonal matrix) reduces this problem [24, 29]. Use of the QR decomposition is equivalent to the reorthogonalization/rescaling of the displacement vectors in the scheme discussed above [30].

We note that the Lyapunov exponents are only defined in the long time limit and if the simulated *non-equilibrium* fluid does not reach a steady state, the exponents will not converge to constant values. It is useful for the purposes of this work to define time-dependent exponents as:

$$\{\lambda_i(t; \Gamma(0)); i = 1, \ldots, 2dN\} = \frac{1}{2t} \ln \{\text{eigenvalues } [\mathbf{L}(t; \Gamma(0))^T \cdot \mathbf{L}(t; \Gamma(0))]\}. \tag{1.20}$$

† It rapidly becomes an illconditioned matrix.

Unlike the Lyapunov exponents, these finite time exponents will depend on the initial phase space vector, $\Gamma(0)$ and the length of time over which the tangent vectors are integrated, and we therefore will refer to them as finite-time, local Lyapunov exponents.

The systems considered here are chaotic: they have at least one positive Lyapunov exponent. This means that (except for a set of zero measure) points that are initially close will diverge after some time, and therefore information on the initial phase space position of the trajectory will be lost. Points that are initially close will eventually span the accessible phase space of the system. The Lyapunov exponents of an equilibrium (Hamiltonian) system sum to zero, reflecting the phase space conservation of these system, whereas for systems in thermostatted steady states, the sum is negative. This indicates that the phase space collapses onto a lower dimensional attractor in the original phase. The set of Lyapunov exponents, can be used to calculate the dimension of phase space accessible to a non-equilibrium steady state. The Kaplan–Yorke dimension of the accessible phase space is defined as

$$D^{\mathrm{KY}} = n^{\mathrm{KY}} + \sum_{i=1}^{n^{\mathrm{KY}}} \lambda_i / |\lambda_{n^{\mathrm{KY}}+1}|,$$

where n^{KY} is the largest integer or which

$$\sum_{i=1}^{n^{\mathrm{KY}}} \lambda_i > 0.$$

As we shall see, for Second Law satisfying steady states this dimension is always less than the ostensible dimension of phase space, $d_{\mathrm{C}}N$. Furthermore, an exact relationship between this dimensional reduction and the limiting small field transport coefficient, has recently been proved [31].

2. Liouville derivation of FT

2.1. *The transient FT*

The probability $p(\delta V_\Gamma(\Gamma(t), t))$, that a phase Γ, will be observed within an infinitesimal phase space volume of size

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

about $\Gamma(t)$ at time t, is given by,

$$p(\delta V_\Gamma(\Gamma(t), t)) = f(\Gamma(t), t)\delta V_\Gamma(\Gamma(t), t), \tag{2.1}$$

where $f(\Gamma(t), t)$ is the normalized phase space distribution function at the phase $\Gamma(t)$ at time t. Since the Liouville equation (1.9), is valid for all phase points Γ, it is also valid for the phase $\Gamma(t)$ which has evolved at time t from from $\Gamma(0)$ at $t = 0$. Integrating the resultant ordinary differential equation gives the Lagrangian form (1.10) of the Kawasaki distribution function [32]:

$$f(\Gamma(t), t) = \exp\left[-\int_0^t \Lambda(\Gamma(s))\,\mathrm{d}s\right] f(\Gamma(0), 0). \tag{2.2}$$

Now consider the set of initial phases inside the volume element of size $\delta V_\Gamma(\Gamma(0),0)$ about $\Gamma(0)$. At time t, these phases will occupy a volume $\delta V_\Gamma(\Gamma(t),t)$. Since by definition, the number of ensemble members within a comoving phase volume is conserved, equation (2.2) implies,

$$\delta V_\Gamma(\Gamma(t),t) = \exp\left[\int_0^t \Lambda(\Gamma(s))\,\mathrm{d}s\right]\delta V_\Gamma(\Gamma(0),0). \tag{2.3}$$

The exponential on the right hand side of (2.3) gives the relative phase space volume contraction along the trajectory, from $\Gamma(0)$ to $\Gamma(t)$.

Our aim is to determine the ratio of probabilities of observing bundles of trajectory segments and their conjugate bundles of antisegments. For any phase space trajectory segment, an antisegment can be constructed using a time reversal mapping, $M^{\mathrm{T}}(\mathbf{q},\mathbf{p}) \equiv (\mathbf{q},-\mathbf{p})$. We will refer to the trajectory starting at $\Gamma(0)$ and ending at $\Gamma(t)$ as $\Gamma(0;t)$. If we advance time from 0 to $t/2$ using the equations of motion (such as (1.4)), we obtain $\Gamma(t/2) = \exp[\mathrm{i}L(\Gamma(0),F_\mathrm{e})t/2]\Gamma(0)$ where the phase Liouvillean, $\mathrm{i}L(\Gamma,F_\mathrm{e})$, is defined as $\mathrm{i}L(\Gamma,F_\mathrm{e})\ldots = [\dot{\mathbf{q}}(\Gamma,F_\mathrm{e})\cdot\partial/\partial\mathbf{q} + \dot{\mathbf{p}}(\Gamma,F_\mathrm{e})\cdot\partial/\partial\mathbf{p}]\ldots$. Continuing to time t gives $\Gamma(t) = \exp[\mathrm{i}L(\Gamma(t/2),F_\mathrm{e})t/2]\Gamma(t/2) = \exp[\mathrm{i}L(\Gamma(0),F_\mathrm{e})t]\Gamma(0)$.

As discussed previously [32], a time-reversed trajectory segment $\Gamma^*(0;t)$ that is initiated at time zero, and for which $\Gamma^*(0;t) = M^{\mathrm{T}}(\Gamma(0;t))$, can be constructed by applying a time-reversal mapping at the midpoint of $\Gamma(t/2)$ and propagating forward and backward in time from this point for a period of $t/2$ in each direction. At time zero, this generates $\Gamma^*(0) = \exp[-\mathrm{i}L(\Gamma^*(t/2),F_\mathrm{e})t/2]\Gamma^*(t/2) = M^{\mathrm{T}}\exp[\mathrm{i}L(\Gamma(t/2),F_\mathrm{e})t/2]\Gamma(t/2) = M^{\mathrm{T}}\Gamma(t)$. See reference [32] for further details. The point $\Gamma^*(0)$ is related to the point $\Gamma(t)$ by a time-reversal mapping. This provides us with an algorithm for finding initial phases which will subsequently generate the conjugate antisegments. Since the Jacobian of the time-reversal mapping is unity, $\delta V_\Gamma(\Gamma^*(t/2),t/2) = \delta V_\Gamma(\Gamma(t/2),t/2)$, the measure of the phase volume $\delta V_\Gamma(\Gamma(t),t)$ is equal to that of $\delta V_\Gamma(\Gamma^*(0),0)$. The ratio of the probabilities of observing the two volume elements at time zero is:

$$\begin{aligned}\frac{p(\delta V_\Gamma(\Gamma(0),0))}{p(\delta V_\Gamma(\Gamma^*(0),0))} &= \frac{f(\Gamma(0),0)\delta V_\Gamma(\Gamma(0),0)}{f(\Gamma^*(0),0)\delta V_\Gamma(\Gamma^*(0),0)} \\ &= \frac{f(\Gamma(0),0)}{f(\Gamma(t),0)}\exp\left[-\int_0^t \Lambda(\Gamma(s))\,\mathrm{d}s\right].\end{aligned} \tag{2.4}$$

It is worth listing the assumptions used in deriving equation (2.4):

(1) The initial distribution $f(\Gamma,0)$ is symmetric under the time reversal mapping $(f(\Gamma,0) = f(M^{\mathrm{T}}(\Gamma),0))$† [Note: The initial phase space distribution does *not* have to be an equilibrium distribution.];

---

† If this is not the case, a more general form of equation (2.4) and hence the FT (2.6) can still be obtained. Equation (2.4) becomes

$$\frac{P(\delta V_\Gamma(0),(0))}{p(\delta V_\Gamma{}^*(\Gamma^*(0),(0))} = \frac{f(\Gamma(0),0)}{f(M^{\mathrm{T}}(\Gamma(t)),0)}\exp\left[-\int_0^t \Lambda(\Gamma(s))\,\mathrm{d}s\right].$$

Furthermore, alternative reversal mappings to the time reversal map M^{T} (such as the Kawasaki map [16, 32]) may be necessary to generate the conjugate trajectories in some situations—see section 6.5 and reference [8].

(2) The equations of motion (1.4), must be reversible;†
(3) The initial ensemble and the subsequent dynamics are *ergodically consistent*:

$$f(M^{\mathrm{T}}[\Gamma(t)],0) \neq 0,\ \forall \Gamma(0). \tag{2.5}$$

Ergodic consistency (2.5) requires that the initial ensemble must actually *contain* time reversed phases of all possible trajectory end points. Ergodic consistency would be violated for example, if the initial ensemble was microcanonical but the subsequent dynamics was adiabatic and therefore did not preserve the energy of the system.‡

It is convenient to define a dissipation function $\Omega(\Gamma)$,

$$\begin{aligned}\int_0^t \mathrm{d}s\ \Omega(\Gamma(s)) &\equiv \ln\left[\frac{f(\Gamma(0),0)}{f(\Gamma(t),0)}\right] - \int_0^t \Lambda(\Gamma(s))\,\mathrm{d}s \\ &= \bar{\Omega}_t t.\end{aligned} \tag{2.6}$$

We can now calculate the probability ratio for observing a particular time averaged value A, of the dissipation function $\bar{\Omega}_t$ and its negative, $-A$. This is achieved by dividing the initial phase space into subregions $\{\delta V_\Gamma(\Gamma_i); i = 1,\ldots\}$ centred on an initial set of phases $\{\Gamma_i(0); i = 1,\ldots\}$. The probability ratio can be obtained by calculating the corresponding ratio of probabilities that the system is found *initially* in those subregions which *subsequently* generate bundles of trajectory segments with the requisite time average values of the dissipation function. Thus the probability of observing the complementary time average values of the dissipation function is given by the ratio of generating the *initial* phases from which the *subsequent* trajectories evolve. We now sum over all subregions for which the time-averaged dissipation function takes on the specified values,

$$\begin{aligned}\ln\frac{p(\bar{\Omega}_t = A)}{p(\bar{\Omega}_t = -A)} &= \ln\frac{\sum_{i|\bar{\Omega}_{t,i}=A} p(\delta V_\Gamma(\Gamma_i(0),0))}{\sum_{i|\bar{\Omega}_{t,i}=-A} p(\delta V_\Gamma(\Gamma_i(0),0))} \\ &= \ln\frac{\sum_{i|\bar{\Omega}_{t,i}=A} p(\delta V_\Gamma(\Gamma_i(0),0))}{\sum_{i|\bar{\Omega}_{t,i}=A} p(\delta V_\Gamma(\Gamma_i^*(0),0))} \\ &= \ln\frac{\sum_{i|\bar{\Omega}_{t,i}=A} p(\delta V_\Gamma(\Gamma_i(0),0))}{\sum_{i|\bar{\Omega}_{t,i}=A} \frac{f(\Gamma_i(t),0)}{f(\Gamma_i(0),0)} \exp\left[\int_0^t \Lambda(\Gamma_i(s))\,\mathrm{d}s\right] p(\delta V_\Gamma(\Gamma_i(0),0))}\end{aligned}$$

† Note that the looser condition, that will still lead to equations (2.4) and (2.6), is that the reverse trajectory *must exist*. This enables the proof to be extended to stochastic dynamics [6, 7].

‡ Jarzynski [33] and Crooks [34, 35] treat cases where the dynamics is not ergodically consistent and thereby obtain expressions for Helmholtz free energy differences between different systems. This work has been widely applied and and extended, see for example [36–38].

$$= \ln \frac{\sum_{i|\bar{\Omega}_{t,i}=A} p(\delta V_\Gamma(\Gamma_i(0),0))}{\sum_{i|\bar{\Omega}_{t,i}=A} \exp(-\bar{\Omega}_{t,i}t)p(\delta V_\Gamma(\Gamma_i(0),0))}$$

$$= At. \tag{2.7}$$

The notation $\sum_{i|\bar{\Omega}_{t,i}=A}$ is used to indicate that the sum is carried out on subvolumes for which $\bar{\Omega}_t = A$. In (2.7) we carry out the time-reversal mapping to obtain the second equality, then substitute equations (2.4) and (2.6). The final equality is obtained by recognizing that since the summation is only carried out over trajectory segments with particular values of $\bar{\Omega}_t$, the exponential term is common and can be removed from the summation.

We have now completed our derivation of the Transient Fluctuation Theorem (TFT):

$$\frac{p(\bar{\Omega}_t = A)}{p(\bar{\Omega}_t = -A)} = \exp[At]. \tag{2.8}$$

The *form* of the above equation applies to any valid ensemble/dynamics combination, although the precise *expression* for $\bar{\Omega}_t$ (2.6) is dependent on the ensemble and dynamics.

The original derivation of the TFT was for homogeneously ergostatted dynamics carried out over an initial ensemble that was microcanonical. In this simple case $\Omega(\Gamma) = -\Lambda(\Gamma) = d_C N\alpha$. From equation (1.8) we see that the *microcanonical* TFT can be written as

$$\frac{p(\overline{[\beta J]}_t F_e = A)}{p(\overline{[\beta J]}_t F_e = -A)} = \exp[-AVt]. \tag{2.9}$$

If the equations of motion are the homogeneously ergostatted SLLOD equations of motion for planar Couette flow, the dissipative flux J is just the xy-element of the pressure tensor, P_{xy}, and the external field is the strain rate, γ, and we have

$$\frac{p(\overline{[\beta P_{xy}]}_t \gamma = A)}{p(\overline{[\beta P_{xy}]}_t \gamma = -A)} = \exp[-AVt]. \tag{2.10}$$

We note that in this case the dissipation function Ω, is precisely the (dimensionless) thermodynamic entropy production (since it is equal to the work done on the system by the external field, $-P_{xy}(\Gamma)\gamma V$, divided by the absolute temperature, $k_B T(\Gamma)$), and also (because the system is at constant energy), is equal to the entropy absorbed from the system by the thermostat, $(\alpha(\Gamma)\sum p_i^2/mk_B T)$.

If the strain rate is positive then in accord with the Second Law of Thermodynamics P_{xy} should be negative (since the shear viscosity is positive). The TFT is consistent with this. In equation (2.10), if A is negative then the right hand side is positive and therefore the TFT predicts the negative time-averaged values of P_{xy} will be much more probable than the corresponding positive values. Further, since P_{xy} and the strain rate are intensive, for a fixed value of the strain rate it becomes exponentially more unlikely to observe positive values for P_{xy} as either the system size or the observation time is increased. In either the large time or the large system limit, the Second Law will not be violated at all.

2.2. *The steady state FT and ergodicity*

We note that in the TFT, time averages are carried out from $t = 0$, where we have an initial distribution $f(\Gamma, 0)$, to some arbitrary later time t—see equation (2.6). One can make the averaging time arbitrarily long. For sufficiently long averaging times t, we might *approximate* the time averages in (2.8) by performing the time average not from $t = 0$ but from some later time $\tau_R \ll t$,

$$\bar{\Omega}(\tau_R, t) \equiv \frac{1}{t - \tau_R} \int_{\tau_R}^{t} \mathrm{d}s\, \Omega(s). \tag{2.11}$$

Using this approximation for time averages required in (2.8) and noting that,

$$\bar{\Omega}_t \equiv \frac{1}{t} \int_0^t \mathrm{d}s\, \Omega(s) = \frac{1}{t - \tau_R} \int_{\tau_R}^{t} \mathrm{d}s\, \Omega(s) + O(\tau_R/t) \approx \bar{\Omega}(\tau_R, t), \tag{2.12}$$

we can derive an asymptotic form of the FT,

$$\lim_{t/t_R \to \infty} \frac{1}{t} \ln \frac{p(\bar{\Omega}(\tau_R, t) = A)}{p(\bar{\Omega}(\tau_R, t) = -A)} = A. \tag{2.13}$$

If the system is thermostatted in some way and if after some finite transient relaxation time τ_R , it comes to a non-equilibrium *steady state*, then (2.13) is in fact an asymptotic Steady State Fluctuation Theorem (SSFT)

$$\lim_{t/\tau_R \to \infty} \frac{1}{t} \ln \frac{p(\bar{\Omega}_{t,\mathrm{ss}} = A)}{p(\bar{\Omega}_{t,\mathrm{ss}} = -A)} = A. \tag{2.14}$$

In this equation $\bar{\Omega}_{t,\mathrm{ss}}$ denotes the fact that the time averages are only computed after the relaxation of initial transients (i.e. in a non-equilibrium steady state). It is understood that the probabilities are computed over an *ensemble* of long trajectories which initially (at some long time in the past) were characterized by the distribution $f(\Gamma, 0)$ at $t = 0$.

We often expect that the non-equilibrium steady state is unique or *ergodic*. When this is so, steady state time averages and statistics are independent of the initial starting phase at $t = 0$. Most of non-equilibrium statistical mechanics is based on the assumption that the systems being studied are ergodic. For example, the Chapman Enskog solution of the Boltzmann equation is based on the tacit assumption of ergodicity. Experimentally, one does not usually measure transport coefficients as ensemble averages: almost universally transport coefficients are measured as time averages, although experimentalists often employ repeated experiments under identical macroscopic conditions in order to determine the statistical uncertainties in their measured time averages. They would not expect that the results of their measurements would depend on the initial (un-specifiable!) microstate. Arguably, the clearest indication of the ubiquity of non-equilibrium ergodicity, is that empirical data tabulations assume that transport coefficients are single valued functions of the macrostate: (N,V,T) and possibly the strength of the dissipative field. The tacit assumption of non-equilibrium ergodicity is so widespread that it is frequently forgotten that it is in fact an *assumption*. The necessary and sufficient conditions for ergodicity are not known. However, if the initial ensemble used to obtain equation (2.14) is the equilibrium ensemble generated by the dynamics when the non-

equilibrium driving force is removed,† and the system is ergodic then the probabilities referred to in the SSFT (2.14) can be computed not only over an *ensemble* of trajectories, but also over segments along a single exceedingly long phase space trajectory.

This is the version of the Fluctuation Theorem first derived (heuristically) by Evans *et al*. in reference [23] and later more rigorously by Gallavotti and Cohen [14, 15].

3. Lyapunov derivation of FT

The original statement of the SSFT by Evans *et al*. [23] was justified using heuristic arguments for the probability of escape of trajectory segments from phase space tubes (i.e. infinitesimal, fixed radius tubes surrounding steady state phase space trajectory segments). A more rigorous derivation of the theorem, based on similar arguments but invoking the Markov partitioning of phase space and the Sinai–Ruelle–Bowen measure was given by Gallavotti and Cohen [14, 15]. However, even this derivation is not completely rigorous because they had to introduce the *Chaotic Hypothesis* in order to complete the proof [14, 15]. The Chaotic Hypothesis has not, and we believe probably cannot be, proven for realistic systems.

We now show how to derive an FT rigourously using escape rate arguments and *Lyapunov weights*. This derivation completely avoids the difficulties of the Chaotic Hypothesis. As we will see, this new derivation employs a partitioning of phase space which is analogous in many respects to the Markov Partition employed by Gallavotti and Cohen. Although our new derivation is rigorous it leads to an exact Transient Fluctuation Theorem rather than an asymptotic Steady State FT.

The probability of escape from infinitesimal phase space trajectory tubes is controlled by the sum of all the finite-time local positive Lyapunov exponents, defined in equation (1.20). Previous Lyapunov derivations of the SSFT [14, 15, 23], assumed either that the initial probability distribution was uniform (e.g. microcanonical), or if non-uniform, that variations in the initial density could be ignored at long times and the asymptotic escape rate would always be dominated by the exponential of the sum of positive Lyapunov exponents. Here we show that this is not the case and that consistent with the Liouville derivation of the SSFT (section 2.2), the steady state FT does indeed depend on the initial ensemble and the dynamics of the system. For an isoenergetic system, the results obtained are identical to those obtained previously for this system [14, 15, 23, 39, 40].

Consider an ensemble of systems which is initially characterized by a distribution $f(\Gamma, 0)$. As before (section 2.1), we assume that the initial distribution is symmetric under the time reversal mapping. Suppose that a phase space trajectory evolves from Γ_0 at $t = 0$ to $\Gamma_0(t)$ at time t. We call this trajectory the mother trajectory. We also consider the evolution of a set of neighbouring phase points, $\Gamma(0)$, that begin at time

† This requires more than ergodic consistency (see equation (2.5)) that is required to generate the TFT and the ensemble version of the SSFT. It means for example, if the steady state is isoenergetic, then the microcanonical ensemble must be used as the initial ensemble—a canonical initial distribution is ergodically consistent with isoenergetic dynamics, but would not be suitable for generation of the *dynamic* version of the SSFT, because it would generate a set of isoenergetic steady states with different energies. This condition can be expressed by stating that there is a unique steady state for the selected combination of initial ensemble and dynamics.

zero within some fixed region of size determined by $d\Gamma$: $0 < \Gamma_\alpha(0) - \Gamma_{0,\alpha}(0) = \delta\Gamma_\alpha(0) < d\Gamma$, $\forall\,\alpha = 1, \ldots, 2d_C N$ (Γ_α is the αth component of the phase space vector Γ, and $\Gamma_{0,\alpha}$ is the αth component of the vector Γ_0), and that are within the region surrounding the mother at least at time t, so $0 < \delta\Gamma_\alpha(t) < d\Gamma$, $\forall\,\alpha$. Because of Lyapunov instability, most initial points that are within this initial region will diverge from the tube at a later time (see figure 3.1). The probability $p(\Gamma_0(0, t; d\Gamma))$, that initial phases start in the mother tube and stay within that tube is given by,

$$p(\Gamma_0(0, t; d\Gamma)) \propto d\Gamma^{6N} f(\Gamma_0, 0) \exp\left[-\sum_{\lambda_i > 0} \lambda_i(t; \Gamma_0) t \right], \tag{3.1}$$

where $\lambda_i(t; \Gamma_0)$ is the finite-time, local Lyapunov exponent defined in equation (1.20).

Since the system is assumed to be time reversible, there will be a set of antitrajectories which are also solutions of the equations of motion. We use the notation $\Gamma^* = M^{\mathrm{T}}(\Gamma)$ to denote the time reversal mapping of a phase point. From figure 3.1 and the properties of the time reversal mapping we know that $M^{\mathrm{T}}\Gamma_0^*(t) \equiv \Gamma_0$ and $M^{\mathrm{T}}\Gamma_0^*(0) \equiv \Gamma_0(t)$. Further from the time reversibility of the dynamics the set of positive Lyapunov exponents for the antitrajectory $\{\lambda_i(t; \Gamma_0^*); \lambda_i > 0\}$ is identical to minus the set of negative exponents for the conjugate forward trajectory,

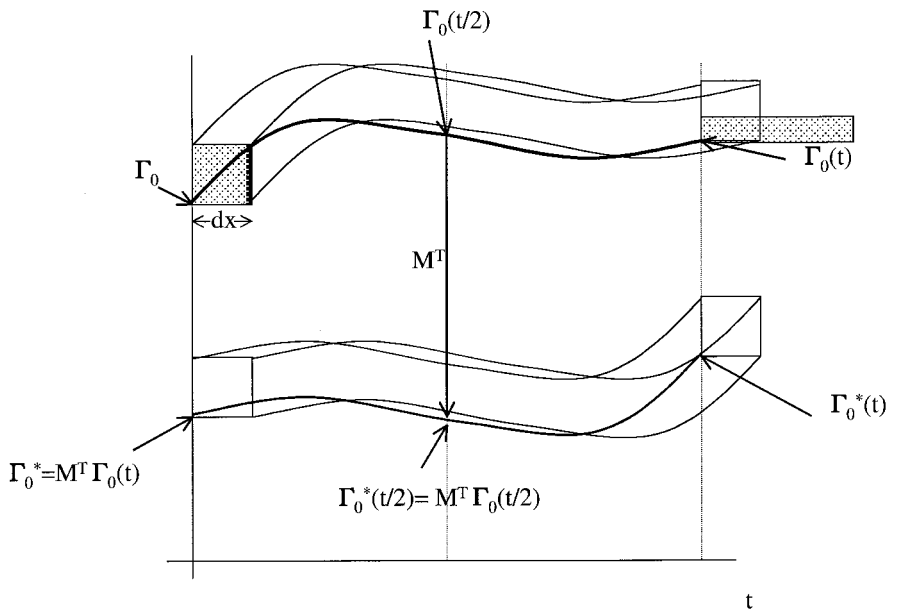


Figure 3.1. A schematic diagram showing how a trajectory and its conjugate evolve. The square region emanating from $\Gamma_0(0)$ has axes aligned with the eigenvectors of the tangent vector propagator matrix $\mathbf{L}(t)^{\mathrm{T}} \cdot \mathbf{L}(t)$. The shaded region thus shows where initial points in this region will propagate to at time t. For illustrative purposes we assume a two-dimensional ostensible phase space and that there is one positive time-dependent local Lyapunov exponent (in the x-direction) and one negative time-dependent local Lyapunov exponent (in the y-direction).

$$\{\lambda_i(t;\Gamma_0^*);\lambda_i>0\}=-\{\lambda_i(t;\Gamma_0);\lambda_i<0\}. \tag{3.2}$$

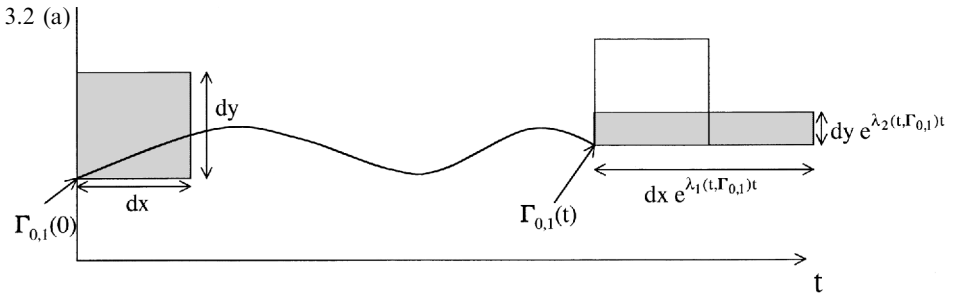
Before considering the Lyapunov derivation of the SSFT, it is useful to consider the computation of phase space averages of a variable A. The ensemble average, $\langle\bar{A}_t\rangle$, of the trajectory segment time average, $\bar{A}_t(\Gamma)$, of an arbitrary phase function $A(\Gamma)$, can be written as,

$$\langle\bar{A}_t\rangle=\int \mathrm{d}\Gamma\, f(\Gamma,0)\bar{A}_t(\Gamma). \tag{3.3}$$

We can partition the initial ostensible phase space into $2d_CN$-dimensional phase volume elements that are formed by the set of orthogonal eigenvectors of $\mathbf{L}(t;\Gamma(0))^{\mathrm{T}}\cdot\mathbf{L}(t;\Gamma(0))$ projected from the initial mother phase points $\{\Gamma_0(0)\}$. By careful construction of the partition, or mesh, we are able to ensure that each point in phase space is associated with a single mother phase—that is, it is within a region about a mother phase point $0<\delta\Gamma_\alpha(t)<\mathrm{d}\Gamma,\forall\alpha$, at least at time t [41]. It is assumed that the phase volume elements are sufficiently small that any curvature in the direction of the eigenvectors can be ignored. In practice this phase space can be constructed as shown in figure 3.2. In this figure we assume that there is *no* curvature in the direction of the eigenvectors over the region considered: this limit will be approached as $\mathrm{d}\Gamma\to 0$.

It should also be noted that although this diagram considers one expanding and one contracting eigendirection, there is no reason that an equal number of positive and negative exponents must exist for this construction to be used, and one or more Lyapunov exponents may be equal to zero. The structure of the steady state is irrelevant, so it is not necessary for the steady state to be Anosov.†

An arbitrary initial mother phase point is selected and the set of points that are within the tube defined by $0<\delta\Gamma_\alpha(t)<\mathrm{d}\Gamma,\ \forall\alpha$ are identified. These points are considered to belong to the first region in the partition. From equation (3.1) it is clear that the volume occupied by these points at $t=0$ is



† Compare this with the Chaotic Hypothesis employed by Gallavotti and Cohen [14, 15].

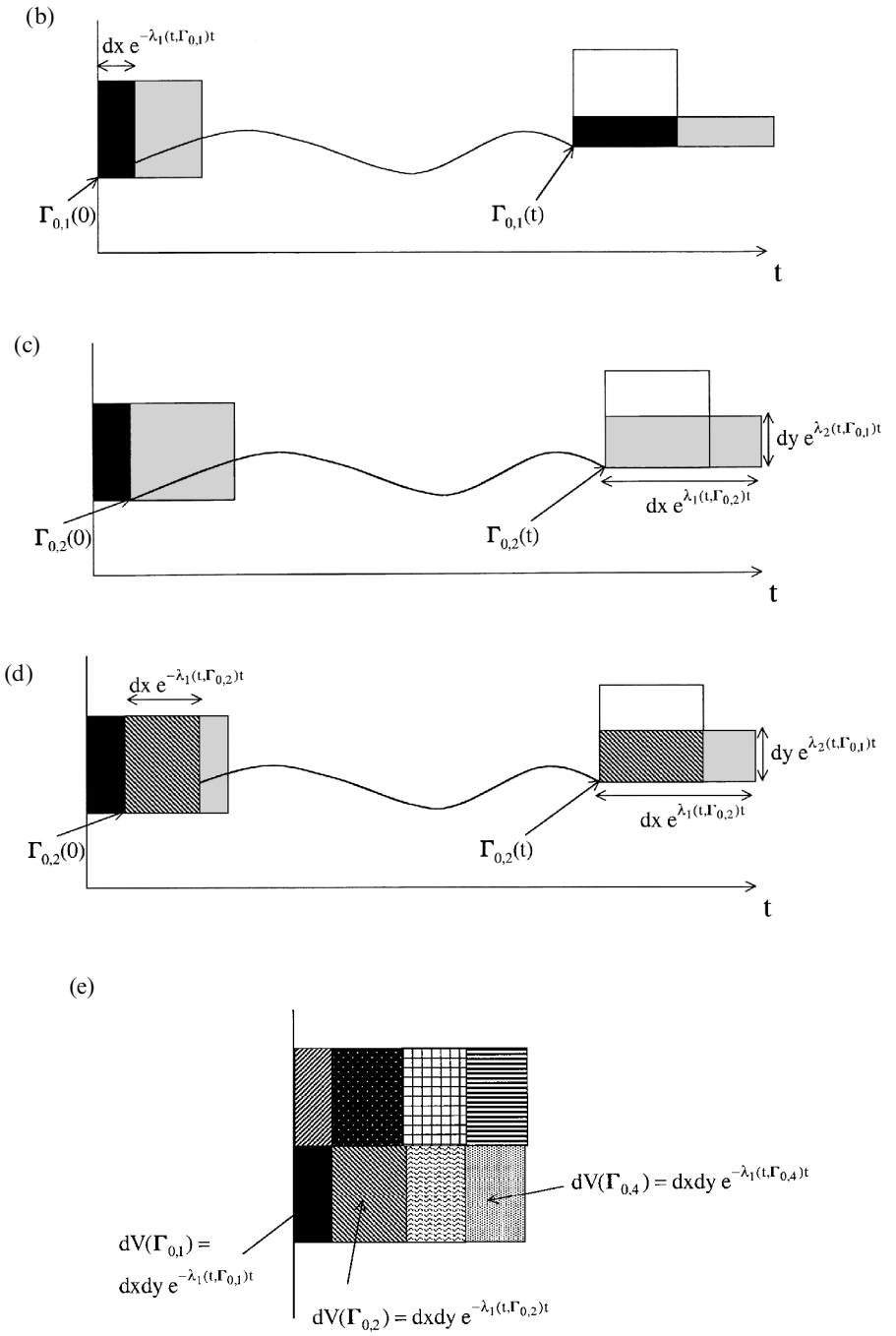


Figure 3.2 (*concluded*). A schematic diagram showing the construction of the partition, or mesh, used to determine phase space averages using Lyapunov weights. For convenience, we assume a two-dimensional ostensible phase space, and that there is one positive and one negative time-dependent local Lyapunov exponent for each region in the section of phase space shown. The size of phase volume elements is assumed to be sufficiently small that any curvature in the direction of the eigenvectors can be ignored.

$$\mathrm{d}\Gamma^{6N} \exp\left[-\sum_{\lambda_i>0} \lambda_i(t;\Gamma_0)t\right].$$

A second region is constructed in a similar manner, with a new mother phase point selected to be initially at a point on the corner of the first region, as shown in figure 3.2 (c), to ensure there is no overlap of regions in the partition. Again, the set of points that remain within the tube defined by $0 < \delta\Gamma_\alpha(t) < \mathrm{d}\Gamma$, $\forall\,\alpha$ are identified, and a second region in the partition is constructed. This is repeated until phase space is completely partitioned into regions of volume

$$\mathrm{d}\Gamma^{6N} \exp\left[-\sum_{\lambda_i>0} \lambda_i(t;\Gamma_0)t\right].$$

Because these volumes depend on the *time-dependent, local* Lyapunov exponents, the volume of each region may differ, and the partitioning will change as longer trajectories are considered. Note that because of the uniqueness of solutions, the time evolved mesh created using this partition never splits into sub-bundles, and one time evolved phase volume element never mixes with another.†

The partition can be formed as shown in figure 3.2. In figure 3.2 (a), a point $\Gamma_{0,1}$ is selected and the region $0 < \delta\Gamma_\alpha(0) < \mathrm{d}\Gamma$, $\forall\,\alpha$ is shaded grey. The location of this region at time t is also shown. In figure 3.2 (b), it is shown that the proportion of points that remain within the tube emanating from $\Gamma_{0,1}$ will be proportional to the Lyapunov weight,

$$\exp\left[-\sum_{\lambda_i>0} \lambda_i(t;\Gamma_0)t\right].$$

The origin of those points is shaded in black. The black region defines the first region of the partition. In figure 3.2 (c), a tube of equal cross-section to that in (a) is formed at a new origin, $\Gamma_{0,2}$, on the corner of $\Gamma_{0,1}$. Again the position of these points at time t is shown, and in figure 3.2 (d), the origin of the points that remain withinin the tube $0 < \delta\Gamma_\alpha(t) < \mathrm{d}\Gamma$, $\forall\,\alpha$ at time t are indicated by the hatching. The construction is repeated until phase space is covered and in figure 3.2 (e), we show the partitioning of a small region of phase space. To calculate phase averages, it is necessary to sum over all regions, with the weight of each region given by the volume of that region and the initial phase space distribution function for that region.

In the limit $\mathrm{d}\Gamma_i \to 0$, $\forall\, i$, we can compute $\langle \bar{A}_t \rangle$ and phase averages as,

$$\langle \bar{A}_t \rangle = \frac{\lim_{\mathrm{d}\Gamma\to 0} \sum_{\{\Gamma_0\}} \bar{A}_t(\Gamma_0) f(\Gamma_0(0),0) \exp\left[-\sum_{\lambda_i>0} \lambda_i(t;\Gamma_0)t\right]}{\lim_{\mathrm{d}\Gamma\to 0} \sum_{\{\Gamma_0\}} f(\Gamma_0(0),0) \exp\left[-\sum_{\lambda_i>0} \lambda_i(t;\Gamma_0)t\right]} \qquad (3.4\,a)$$

and

† In contrast, the tubes of size $\mathrm{d}\Gamma$, used to identify the regions associated with each mother phase point, will generally overlap, even at time zero, since they are of constant size but emanate from the irregularly spaced mesh of $\{\Gamma_0(0)\}$.

$$\langle A(s)\rangle = \frac{\lim_{\mathrm{d}\Gamma\to 0}\sum_{\{\Gamma_0\}} A(\Gamma_0(s))f(\Gamma_0(0),0)\exp\left[-\sum_{\lambda_i>0}\lambda_i(t;\Gamma_0)t\right]}{\lim_{\mathrm{d}\Gamma\to 0}\sum_{\{\Gamma_0\}} f(\Gamma_0(0),0)\exp\left[-\sum_{\lambda_i>0}\lambda_i(t;\Gamma_0)t\right]}, \tag{3.4b}$$

respectively, where we sum over the set of mother phase points $\{\Gamma_0\} = \{\Gamma_{0,i}; i = 1, N_{\Gamma_0}\}$. These equations simply mean that in order to obtain a phase space average, we sum over all regions in the partition, weighting each with its volume (determined from the Lyapunov weight given by equation (3.1) which is equivalent to the Sinai–Ruelle–Bowen (SRB) measure that is used to describe Anosov systems [14, 15]), and multiplying by the appropriate initial distribution function.†

We can describe in words what the Lyapunov weights appearing in equations (3.4 *a*,*b*) achieve. On the set of initial phases, our mesh places a greater density of initial mother phase points in those regions of greatest chaoticity—those regions with the greatest sums of positive local Lyapunov exponents. This is required because for strongly chaotic regions, trajectories diverge more quickly from the mother trajectory. In order to compute time averages correctly we need to weight the time-averaged properties along the mother trajectories, by the product of the initial distribution at the origin of the mother trajectory, and the measure of the initial hypervolume of those trajectories which do not escape from the mother trajectory. These volumes are proportional to the *negative* exponentials of the sums of positive local Lyapunov exponents.

An important consequence of equation (3.4) is that it can be used to show that if the dynamics of a system is not chaotic and its reverse dynamics is also not chaotic, no transport will occur. If the system is not chaotic there are no positive Lyapunov exponents, and if the anti-dynamics is also not chaotic, then due to the mapping given by equation (3.2), all Lyapunov exponents must be zero, and all the Lyapunov weights will be equal to unity (all the phase space volumes in the mesh will have equal measure). This means that there will be perfect Loschmidt pairing: the weight associated with the trajectory starting at Γ_0 will be identical to that associated starting at Γ_0^*; and the phase average of any function that is odd under time reversal, such as a dissipation function, will equal zero. This will apply to systems starting in any (equilibrium) ensemble since the initial distribution functions are even under time reversal.

We now apply these concepts to compute the ratio of conjugate averages of the dissipation function. The dissipation function that we consider is defined in equation (2.6). The ratio of corresponding probabilities is:

$$\frac{\Pr(\bar{\Omega}_t = A)}{\Pr(\bar{\Omega}_t = -A)} = \lim_{\mathrm{d}\Gamma\to 0}\frac{\sum_{\{\Gamma_0|\bar{\Omega}_t(\Gamma_0)=A\}} f(\Gamma_{0,i}(0),0)\exp\left[-\sum_{\lambda_j>0}\lambda_j(t,\Gamma_{0,i})t\right]}{\sum_{\{\Gamma_0|\bar{\Omega}_t(\Gamma_0)=-A\}} f(\Gamma_{0,i}(0),0)\exp\left[-\sum_{\lambda_j>0}\lambda_j(t,\Gamma_{0,i})t\right]}$$

† Although equation (3.4) provides an extremely useful theoretical expression, due to the difficulty of constructing the partition it does not currently provide a feasible route for numerical calculation of phase averages for many particle systems.

$$= \lim_{d\Gamma\to 0} \frac{\sum_{\{\Gamma_0|\bar{\Omega}_t(\Gamma_0)=A\}} f(\Gamma_{0,i}(0),0)\exp\left[-\sum_{\lambda_j>0}\lambda_j(t,\Gamma_{0,i})t\right]}{\sum_{\{\Gamma_0|\bar{\Omega}_t(\Gamma_0)=A\}} f(\Gamma^*_{0,i}(0),0)\exp\left[-\sum_{\lambda_j>0}\dot{\lambda}_j(t,\Gamma^*_{0,i})t\right]}$$

$$= \lim_{d\Gamma\to 0} \frac{\sum_{\{\Gamma_0|\bar{\Omega}_t(\Gamma_0)=A\}} f(\Gamma_{0.i}(0),0)\exp\left[-\sum_{\lambda_j>0}\lambda_j(t,\Gamma_{0.i})t\right]}{\sum_{\{\Gamma_0|\bar{\Omega}_t(\Gamma_0)=A\}} f(\Gamma_{0,i}(t),0)\exp\left[+\sum_{\lambda_j<0}\lambda_j(t,\Gamma_{0,i})t\right]}, \quad (3.5)$$

where we use the relationships between conjugate trajectories to express the numerator and denominator in terms of sums over $\{\Gamma_0|\bar{\Omega}_t(\Gamma_0)=A\}$. The notation $\sum_{\{\Gamma_0|\bar{\Omega}_t=A\}}\cdots$ is used to indicate that the sum is carried out over the set of regions in the mesh for which $\bar{\Omega}_t = A$. Using (2.6) to substitute for $f(\Gamma_{0,i}(t),0)$, we obtain,

$$\frac{\Pr(\bar{\Omega}_t=A)}{\Pr(\bar{\Omega}_t=-A)} = \lim_{d\Gamma\to 0} \frac{\sum_{\{\Gamma_0|\bar{\Omega}_t(\Gamma_0)=A\}} f(\Gamma_0(0),0)\exp\left[-\sum_{\lambda_i>0}\lambda_i(t,\Gamma_0)t\right]}{\sum_{\{\Gamma_0|\bar{\Omega}_t(\Gamma_0)=A\}} \exp[-\bar{\Omega}_t t] f(\Gamma_0(0),0)\exp[-\bar{\Lambda}_t t]\exp\left[+\sum_{\lambda_i<0}\lambda_i(t,\Gamma_0)t\right]}$$

$$= \lim_{d\Gamma\to 0} \frac{\sum_{\{\Gamma_0|\bar{\Omega}_t(\Gamma_0)=A\}} f(\Gamma_0(0),0)\exp\left[-\sum_{\lambda_i>0}\lambda_i(t,\Gamma_0)t\right]}{\sum_{\{\Gamma_0|\bar{\Omega}_t(\Gamma_0)=A\}} \exp[-\bar{\Omega}_t t] f(\Gamma_0(0),0)\exp\left[-\sum_{\lambda_i>0}\lambda_i(t,\Gamma_0)t\right]}$$

$$= \lim_{d\Gamma\to 0} \exp[At] \frac{\sum_{\{\Gamma_0|\bar{\Omega}_t(\Gamma_0)=A\}} f(\Gamma_0(0),0)\exp\left[-\sum_{\lambda_i>0}\lambda_i(t,\Gamma_0)t\right]}{\sum_{\{\Gamma_0|\bar{\Omega}_t(\Gamma_0)=A\}} f(\Gamma_0(0),0)\exp\left[-\sum_{\lambda_i>0}\lambda_i(t,\Gamma_0)t\right]}$$

$$= \exp[At]. \quad (3.6)$$

To obtain the second line we use the fact that the sum of *all* the local Lyapunov exponents is the time average of the phase space compression factor:

$$\bar{\Lambda}_t(\Gamma_0) = \sum_{\forall i} \lambda_i(t,\Gamma_0).$$

Of course equation (3.6) is identical to the ensemble independent TFT derived previously (2.8). Furthermore, the same arguments as those presented in section 2.2 can be applied to derive the SSFT (2.14). For an isoenergetic system, the SSFT derived from equation (3.6) is identical to that obtained previously for this system [14, 15, 23]. However, if the system is not microcanonical, the Lyapunov weights and associated SRB measure, do *not* dominate the weight that results from the non-uniformity of the initial distribution.

4. Applications

In sections 2 and 3 we have shown that a general form of the fluctuation theorem can be derived for various ergodically consistent combinations of ensemble and dynamics. Table 4.1 summarizes the TFT obtained for many of the systems of interest [6, 42, 43]. In the last row, the exact FT for an ensemble of steady state trajectory segments is also given. As shown there, this collapses to the usual asymptotic SSFT in the long time limit [42]. SSFT can be obtained for other ensembles in a similar manner.

Two classes of system can be considered:

(1) non-equilibrium steady states where the FT predicts the frequency of occurrence of Second Law violating antitrajectory segments [39, 42]—see sections 4.1 and 4.2;
(2) non-dissipative systems where the FT describes the free relaxation of systems towards, rather than further away from, equilibrium such as the free expansion of gases into a vacuum and mixing in a binary system [43]—see section 4.3.

In this section we discuss some of these systems in more detail. We also present in section 4.4, a generalized form of the FT that applies to any phase function that is odd under time-reversal symmetry and in section 4.5, the integrated form of the FT (1.3). We use reduced Lennard–Jones units throughout this section [16].

4.1. *Isothermal systems*

As an example, we consider the TFT for a system which is initially in the isokinetic ensemble and which undergoes isokinetic dynamics [42] with kinetic energy K_0. The isokinetic distribution function is,

$$f(\Gamma(0),0) = f_K(\Gamma(0),0) = \frac{\exp\left[-\beta H_0(\Gamma(0))\right]\delta(K(\Gamma(0)) - K_0)}{\int \mathrm{d}\Gamma \exp\left[-\beta H_0(\Gamma)\right]\delta(K(\Gamma) - K_0)}. \tag{4.1}$$

Substituting into equation (2.4) gives

$$\begin{aligned}\frac{p(\delta V_\Gamma(\Gamma(0),0))}{p(\delta V_\Gamma(\Gamma^*(0),0))} &= \frac{f_K(\Gamma(0),0)\delta V_\Gamma(\Gamma(0),0)}{f_K(\Gamma^*(0),0)\delta V_\Gamma(\Gamma^*(0),0)} \\ &= \frac{\exp\left[-\beta H_0(\Gamma(0))\right]}{\exp\left[-\beta H_0(\Gamma(t))\right]}\exp\left[-\int_0^t \Lambda(\Gamma(s))\,\mathrm{d}s\right] \\ &= \exp\left[\beta\int_0^t \dot{\Phi}(\Gamma(s))\,\mathrm{d}s\right]\exp\left[-\int_0^t \Lambda(\Gamma(s))\,\mathrm{d}s\right],\end{aligned} \tag{4.2}$$

where we have used the symmetry of the mapping, $H_0(\Gamma^*(0)) = H_0(\Gamma(t))$, $\delta V_\Gamma(\Gamma^*(0),0) = \delta V_\Gamma(\Gamma(t),t)$ and $K(\Gamma^*(0),0) = K(\Gamma(t),t)$ to obtain the second equality and $H_0(\Gamma(t)) = H_0(\Gamma(0)) + \int_0^t \dot{H}_0(\Gamma(s))\,\mathrm{d}s = H_0(\Gamma(0)) + \int_0^t \dot{\Phi}(\Gamma(s))\,\mathrm{d}s$ to obtain the final equality. We see that:

$$\begin{aligned}\Omega(\Gamma) &= \beta\dot{\Phi}(\Gamma) - \Lambda(\Gamma) \\ &= -\beta J(\Gamma) V F_\mathrm{e}\end{aligned} \tag{4.3}$$

Table 4.1. Transient fluctuation formula in various ergodically consistent ensembles.[a,b]

Isokinetic dynamics	$\ln \frac{p(\bar{J}_t = A)}{p(\bar{J}_t = -A)} = -AtF_e\beta V$
Isothermal-isobaric[c]	$\ln \frac{p(\overline{JV}_t = A)}{p(\overline{JV}_t = -A)} = -AtF_e\beta$
Isoenergetic	$\ln \frac{p(\bar{J}_t = A)}{p(\bar{J}_t = -A)} = -AtF_e\beta V$ or $\ln \frac{p(\bar{\Lambda}_t = A)}{p(\bar{\Lambda}_t = -A)} = -At$
Isoenergetic boundary driven flow	$\ln \frac{p(\bar{\Lambda}_t = A)}{p(\bar{\Lambda}_t = -A)} = -At$
Nosé-Hoover (canonical) dynamics	$\ln \frac{p(\bar{J}_t = A)}{p(\bar{J}_t = -A)} = -AtF_e\beta V$
Wall ergostatted field driven flow[c]	$\ln \frac{p(\overline{J\beta_{\text{wall}}}_t = A)}{p(\overline{J\beta_{\text{wall}}}_t = -A)} = -AtF_eV$ or $\ln \frac{p(\bar{\Lambda}_t = A)}{p(\bar{\Lambda}_t = -A)} = -At$
Wall thermostatted field driven flow[c]	$\ln \frac{p(\bar{J}_t = A)}{p(\bar{J}_t - A)} = -AtF_e\beta V - \ln\left(\langle \exp[\bar{\Lambda}_t t(1 - \beta_{\text{system}}/\beta_{\text{wall}})]\rangle_{\bar{J}_t = A}\right)$
Relaxation of a system with a non-homogeneous density profile imposed using a potential $\Phi_g(\mathbf{q})$; initial canonical distribution	$\ln \frac{p\left(\int_0^t \mathrm{d}s\Phi_g(s) = A\right)}{p\left(\int_0^t \mathrm{d}s\Phi_g(s) = -A\right)} = -A\beta$
Adiabatic response to a colour field	$\ln \frac{p(\bar{J}_t = A)}{p(\bar{J}_t = -A)} = -AtF_e\beta V$
Isoenergetic dynamics with a stochastic force[d]	$\ln \frac{p(\bar{J}_t = A)}{p(\bar{J}_t = -A)} = -AtF_e\beta V$ or $\ln \frac{p(\bar{\Lambda}_t = A)}{p(\bar{\Lambda}_t = -A)} = -At$
Steady state isoenergetic dynamics:[e] $\overline{\beta J}_t = \frac{1}{t}\int_{t_0}^{t_0+t} \beta(s)J(s)\,\mathrm{d}s$ where $t_0 \gg \tau_M$	$\ln \frac{p(\overline{J\beta}_t = A)}{p(\overline{J\beta}_t = -A)}$ $= -AtF_eV - \ln\left(\left\langle \exp\left[F_eV\left(\int_0^{t_0} J(s)\beta(s)\,\mathrm{d}s + \int_{t_0+t}^{2t_0+t} J(s)\beta(s)\,\mathrm{d}s\right)\right]\right\rangle_{\bar{J}\beta_t = A}\right)$ $\lim_{t\to\infty} \frac{1}{t}\ln \frac{p(\overline{J\beta}_t = A)}{p(\overline{J\beta}_t = -A)} = -AF_eV$

[a] It is assumed that the limit of a large system has been taken so that $O(1/N)$ effects can be neglected. Some of these relationships were presented in reference [33].

[b] In most cases considered here the dissipative flux, J, is defined by $-JF_eV = \frac{\mathrm{d}H_0^{\text{ad}}}{\mathrm{d}t}$ where H_0 is the equilibrium internal energy, however for the isothermal–isobaric case $-JF_eV = \frac{\mathrm{d}I_0^{\text{ad}}}{\mathrm{d}t}$ where I_0 is the equilibrium enthalpy.

[c] In these wall ergostatted/thermostatted systems, it is assumed that the energy/temperature of the full system (wall and fluid) is fixed.

[d] This result is valid for a class of stochastic systems where the stochastic force ensures that the system remains on the constant-energy zero-total momentum hypersurface. See reference [32] for further details.

[e] Similar steady state formulae can be obtained for other ensembles. τ_M is the Maxwell time that characterizes the time required for relaxation of the nonequilibrium system into a steady state.

and from equation (2.8) we therefore have,

$$\frac{p(\bar{J}_t = A)}{p(\bar{J}_t = -A)} = \exp\left[-AtF_e\beta V\right]. \tag{4.4}$$

The TFT given by equation (4.4) is true at all times for the isokinetic ensemble when all initial phases are sampled from an equilibrium isokinetic ensemble [42].

4.2. *Isothermal–isobaric systems*

We consider a system made up of N particles. These particles are identical except their colour: half the particles are one colour, say, red; whereas the other half are blue. The system is thermostatted and barostatted and the two coloured species are driven in opposite directions by an applied colour field. The system is closely related to electrical conduction but avoids the complications of long ranged electrostatic forces.

For an isobaric–isothermal ensemble the phase space trajectories are confined to constant hydrostatic pressure and constant peculiar kinetic energy hypersurfaces. The N-particle phase space distribution function is given by $f(\Gamma, V) \sim \delta(p - p_0)\delta(K - K_0)\exp\left[-\beta_0(H_0 + p_0 V)\right]$, where p is the hydrostatic pressure, V the system volume, p_0, K_0 are the fixed values of the pressure and kinetic energy and β_0 is the Boltzmann factor $\beta_0 = 1/(k_B T_0) = (2K_0)/(d_C N)$. We note that for isobaric systems the system volume is included as an additional coordinate [16].

The systems we examine are brought to a steady state using both a Gaussian thermostat and barostat. At time $t = 0$, a colour field is applied and the response of the system is monitored for a time, t, that is referred to as the length of the trajectory segment. The equations of motion used are [16],

$$\left.\begin{aligned} \dot{\mathbf{q}}_i &= \frac{\mathbf{p}_i}{m} + \dot{\varepsilon}\mathbf{q}_i \\ \dot{\mathbf{p}}_i &= \mathbf{F}_i - \mathbf{i}c_i F_c - \dot{\varepsilon}\mathbf{p}_i - \alpha\mathbf{p}_i \\ \dot{V} &= \mathrm{d}V\dot{\varepsilon}, \end{aligned}\right\} \tag{4.5}$$

where

$$\mathbf{F}_i = \frac{-[\partial\Phi(\mathbf{q})]}{[\partial\Phi(\mathbf{q}_i)]}, \quad \dot{\varepsilon} = -\left[\frac{1}{2m}\sum_{i\neq j}\mathbf{q}_{ij}\cdot\mathbf{p}_{ij}\left(\phi''_{ij} + \frac{\phi'_{ij}}{q_{ij}}\right)\right] \Big/ \left[\frac{1}{2}\sum_{i\neq j} q_{ij}^2\left(\phi''_{ij} + \frac{\phi'_{ij}}{q_{ij}}\right) + 9pV\right]$$

is the *dilation rate* [16] and

$$\alpha = -\dot{\varepsilon} + \left[\sum_{i=1}^{N}(\mathbf{F}_i - \mathbf{i}c_i F_c)\cdot\mathbf{p}_i\right] \Big/ \left[\sum_{i=1}^{N}\mathbf{p}_i\cdot\mathbf{p}_i\right]$$

is the thermostat multiplier. The particles have a colour 'charge' $c_i = (-1)^i$, so that they experience opposite forces from the colour field, F_c. For this system, the phase space compression factor is $\Lambda(t) = -d_C N\alpha$ and the dissipative flux, which is analogous to the electric current density, is defined as the time adiabatic time derivative of the enthalpy

$$\mathrm{d}(H_0 + p_0 V)/\mathrm{d}t|^{\mathrm{ad}} \equiv \dot{I}^{\mathrm{ad}} \equiv -J_c V F_c = -F_c\sum_{i=1}^{N} c_i p_{xi}$$

[16] where

$$J_c = \sum_{i=1}^{N} c_i p_{xi}/V$$

is the dissipative flux or the colour current density. Under constant pressure conditions the rate of change of the enthalpy is equal to the rate of change of the entropy multiplied by the absolute temperature.

Using equation (2.6), the phase space compression factor and the initial distribution function defined above, the dissipation function for this system is

$$\bar{\Omega}_t = -\frac{1}{t}\int_0^t \mathrm{d}s\, \beta_0 J_c(s) V(s) F_c = -\beta_0 \overline{[J_c V]}_t F_c. \tag{4.6}$$

Hence, equation (2.7) with this expression yields [44]

$$\frac{p(-\beta_0 \overline{[J_c V]}_t F_c = A)}{p(-\beta_0 \overline{[J_c V]}_t F_c = -A)} = \exp[At]. \tag{4.7}$$

It is straightforward to show that the same expression is obtained when Nosé–Hoover constraints [16] are applied to the pressure and temperature rather than Gaussian constraints; or if a combination of these types of thermostat is used.

4.3. *Free relaxation in Hamiltonian systems*

We now consider the free relaxation of a colour density modulation. Firstly we need to construct an ensemble of systems with a colour density modulation. Without loss of generality, consider a system of N particles that for $t < 0$ is subject to a colour Hamiltonian,

$$H_c = H_0 + F_c \sum_{i=1}^{N} c_i \sin(k x_i), \tag{4.8}$$

where $c_i = (-1)^i$ is the colour charge of particle i, $k = 2\pi/L$ where L is the boxlength, and

$$H_0 \equiv \sum_i p_i^2/2m + \sum_{i<j} \phi(q_{ij})$$

is a colour blind interaction Hamiltonian (like the potential energy ϕ, H_0 does not refer to the colour charges). The colour density modulation can be measured by averaging the appropriate Fourier component

$$\rho_c(k) \equiv \sum_{i=1}^{N} c_i \sin(k x_i). \tag{4.9}$$

We assume that for $t < 0$, the system is in contact with a heat bath. Since the system is at thermal equilibrium for $t < 0$, the colour field induces a colour density wave,

$$\langle \rho_c(k,0)\rangle_{F_c} = \frac{\int d\Gamma\ \rho_c(k)\exp\left[-\beta(H_0 + F_c\rho_c(k))\right]}{\int d\Gamma\ \exp\left[-\beta(H_0 + F_c\rho_c(k))\right]}$$

$$\stackrel{F_c\to 0}{=} -\beta F_c\langle \rho_c(k)^2\rangle_{F_c=0}. \tag{4.10}$$

From the last line of (4.10) it is clear that in the weak field limit $\lim_{F_c\to 0}\langle \rho_c(k,0)\rangle_{F_c} < 0$. So at $t = 0$ the system is initially modulated with a colour density wave. We wish to consider the behaviour of the system for $t > 0$, when the external colour field is 'turned off' and the contact between the system and the heat bath is broken. The system then relaxes freely, under the colour blind interaction Hamiltonian H_0.

For $t > 0$, no work is done on the system and there is no dissipation. The system evolves with constant energy; $E = H_0(\Gamma)$. No heat is exchanged with the surroundings and thus there is no change in the *thermodynamic* entropy of either the system or the surroundings.† Nevertheless, according to Le Chatelier's Principle [12], the colour density modulation should decay rather than grow as the system becomes homogeneous with respect to colour.

The 'dissipation function', $\Omega(\Gamma)$, can be determined using equation (2.6). For $t > 0$ there is no phase space compression since the dynamics is Newtonian and there is no applied field. Furthermore, energy is conserved. Therefore, the dissipation function becomes:

$$t\bar{\Omega}_t = \beta[H_c(t) - H_c(0)]$$

$$= \beta F_c\int_0^t ds\,\dot{\rho}_c(k,s) = \beta F_c[\rho_c(k,t) - \rho_c(k,0)]. \tag{4.11}$$

The dissipation function thus gives a direct measurement of the change in the colour density modulation order parameter. Applying the TFT (2.8) to this system gives,

$$\frac{p[\rho_c(k,t) - \rho_c(k,0) = A]}{p[\rho_c(k,t) - \rho_c(k,0) = -A]} = \exp\left[\beta F_c A\right], \tag{4.12}$$

where β is the reciprocal temperature of the initial ensemble.

In order to test this equation, we considered a system of 32 particles in two Cartesian dimensions. The particles interact via a WCA [45] potential and the equations of motion at $t < 0$ are

$$\left.\begin{aligned}\dot{\mathbf{q}}_i &= \frac{\mathbf{p}_i}{m}\\ \dot{\mathbf{p}}_i &= \mathbf{F}_i - \mathbf{i}c_i kF_c\cos kx_i - \zeta\mathbf{p}_i\\ \dot{\zeta} &= \frac{1}{Q}\left(\sum_{i=1}^{N}\frac{\mathbf{p}_i^2}{m} - d_C N k_B T\right),\end{aligned}\right\} \tag{4.13}$$

where ζ is the Nosé–Hoover thermostat multiplier [16]. At $t = 0$, the field and the thermostat are switched off and the system is allowed to relax to equilibrium.

† Any *thermodynamic* change in the entropy must be measurable using calorimetry: $dS \equiv dQ_{rev}/T$ where dQ_{rev} is the reversible heat exchanged with the surroundings.

Figure 4.1 shows the modulation in the colour density of the particles at $t < 0$ and the mixing that occurs as predicted by the Le Chatelier's Principle† when the field is switched off. The FT for this system would predict that although mixing would be the most likely outcome, for small systems and short periods of time, the colour modulation could in fact become *stronger*. This de-mixing violates Le Chatelier's Principle. Figure 4.2 shows a histogram of $p(t\bar{\Omega}_t)$ and figure 4.3 shows that the FT is satisfied for this system.

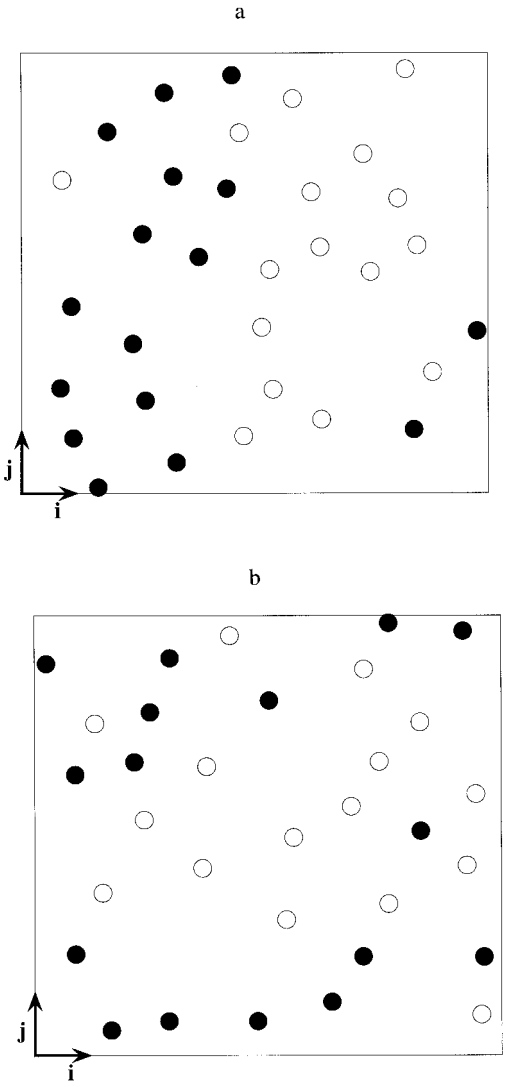


Figure 4.1. Snapshots from a molecular dynamics simulation showing the phase separation of red and blue particles at $t < 0$ (with field on) in (*a*), and their relaxation to equilibrium (at $t = 32$) in (*b*). Here $T = 1.0$, $n = 0.4$ and $F_e = 2.0$.

† If a system is in stable equilibrium, then any spontaneous change in its parameters must bring about processes which tend to restore the system to equilibrium [12].

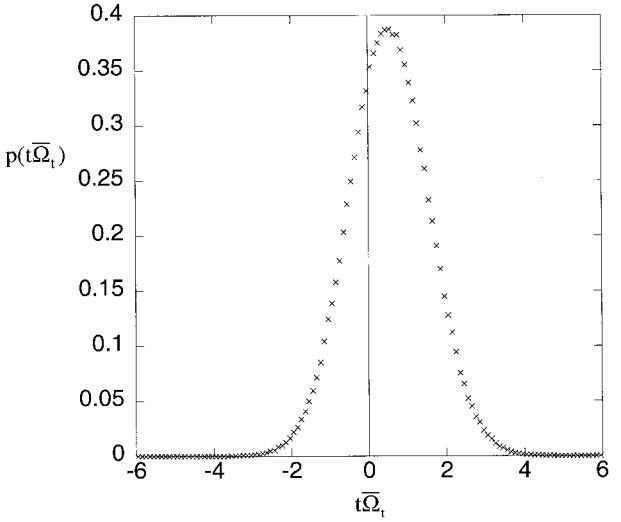


Figure 4.2. A histogram of the distribution of the dissipation function for a system containing a colour separated binary system that is relaxing to equilibrium. Here $T = 1.0$, $n = 0.4$, $F_c = 2.0$ and $t = 0.4$.

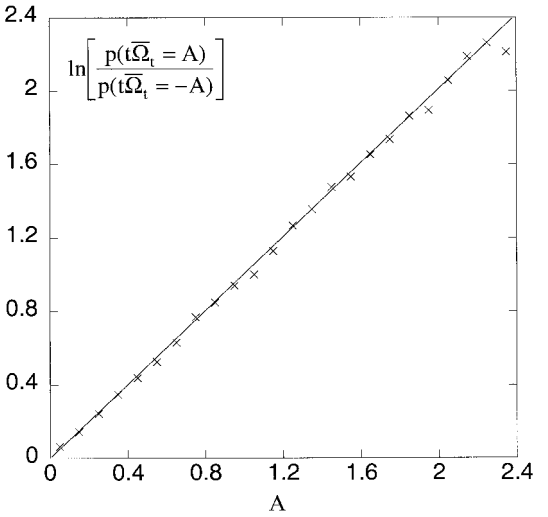


Figure 4.3. A test of the FT given by equation (4.12) for a system containing a colour separated binary system that is relaxing to equilibrium. Here $T = 1.0$, $n = 0.4$, $F_c = 2.0$ and $t = 0.4$.

4.4. *FT for arbitrary phase functions*

The FTs derived above predict the ratio of the probabilities of observing conjugate values of the dissipation function. As given above, these theorems give no information on the probability ratios for any functions other than the dissipation function (2.6). In this section we describe how the FT can be extended to apply to arbitrary phase functions which have a specific parity under time reversal symmetry [46].

Let $\phi(\Gamma)$ be an arbitrary phase function and define the time average

$$\bar{\phi}_{i,t} = \frac{1}{t}\int_0^t \mathrm{d}s\, \phi(\Gamma_i(s)) \tag{4.14}$$

for a phase space trajectory: $\Gamma_i(s)$. At $t = 0$ the phase space volume occupied by a contiguous bundle of trajectories for which $\{\Gamma_i | A < \bar{\phi}_{i,t} < A + \delta A\}$ is given by $\delta V_\Gamma(\Gamma(0), 0)$ and at time t these phase points will occupy a volume $\delta V_\Gamma(\Gamma(t), t) = \delta V_\Gamma(\Gamma(0), 0) \exp[\bar{\Lambda}(t)t]$ where $\bar{\Lambda}(t)$ is the time-averaged phase space compression factor along these trajectories. We denote $\bar{\phi}(t) = \langle \bar{\phi}_{i,t} \rangle_{\{i\}}$, that is the average value of $\bar{\phi}_{i,t}$ over the set of contiguous trajectories, $\{\Gamma_i\}$.

If the dynamics is reversible, there will be a contiguous set of initial phases $\{\Gamma_i^*(0)\}$, given by $\Gamma_i^*(0) = M^{\mathrm{T}}(\Gamma_i(t))$, that will occupy a volume $\delta V_\Gamma(\Gamma^*(0), 0) = \delta V_\Gamma(\Gamma(t), t) = \delta V_\Gamma(\Gamma(0), 0) \exp[\bar{\Lambda}(t)t]$ along which the time-averaged value of the phase function is $\bar{\phi}_{i^*,t} = M^{\mathrm{T}}(\bar{\phi}_{i,t})$. For any $\phi_i(\Gamma)$ that is odd under time reversal, $\bar{\phi}_{i^*,t} = -\bar{\phi}_{i,t}$.

The probability ratio of observing trajectories originating in an initial phase volume and its conjugate phase volume will be related to the initial phase space distribution function and the size of the volume elements by equation (2.4). Therefore, from the definition of the dissipation function in (2.6) we obtain,

$$\frac{p(\delta V_\Gamma(\Gamma(0), 0))}{p(\delta V_\Gamma(\Gamma^*(0), 0))} = \exp[\bar{\Omega}_t t]. \tag{4.15}$$

It is possible that there are non-contiguous bundles of trajectories for which $\{\Gamma_i | A < \bar{\phi}_{i,t} < A + \delta A\}$, and since these bundles may have different values $\bar{\Omega}_t$ the probability ratio (4.15) may differ for each bundle. The probability of observing a trajectory for which $A < \bar{\phi}_t < A + \delta A$, is obtained by summing over the probabilities of observing these $m = 1, M$ non-contiguous volume elements, $\delta V_{\Gamma,m}(\Gamma(0), 0)$. If the phase function is odd under time reversal symmetry, then the ratio of the probability of observing trajectories for which $A < \bar{\phi}_t < A + \delta A$ to the probability of observing conjugate trajectories, for which $-A < \bar{\phi}_t < -A + \delta A$ is,

$$\begin{aligned}
\frac{p(\bar{\phi}_t = A)}{p(\bar{\phi}_t = -A)} &= \frac{\sum_{m=1}^{M} p(\delta V_{\Gamma,m}(\Gamma(0), 0)}{\sum_{m=1}^{M} p(\delta V_{\Gamma,m}(\Gamma^*(0), 0))} \\
&= \frac{\sum_{m=1}^{M} p(\delta V_{\Gamma,m}(\Gamma(0), 0))}{\sum_{m=1}^{M} p(\delta V_{\Gamma,m}(\Gamma(0), 0) \exp(-\bar{\Omega}_t t)} \\
&= \langle \exp(-\bar{\Omega}_t t) \rangle^{-1}_{\bar{\phi}_t = A},
\end{aligned} \tag{4.16}$$

where the notation $\langle \ldots \rangle_{\bar{\phi}_t = A}$ refers to the ensemble average over (possibly) non-contiguous trajectory bundles for which $\bar{\phi}_t = A$. Equation (4.16) gives the ratio of the measure of those phase space trajectories for which $\bar{\phi}_t = A$ to the measure of those trajectories for which $\bar{\phi}_t = -A$. This is the Generalised Transient Fluctuation Theorem (GTFT) for any phase variable $\bar{\phi}_t$ that is odd under time reversal. Provided

it has a definite parity under time reversal symmetry, the actual form of $\bar{\phi}_t$ is quite arbitrary. If the phase variable is even, the we obtain the trivial relationship

$$\langle \exp(-\bar{\Omega}_t t)\rangle^{-1}_{\bar{\phi}_t=A} = \frac{p(\bar{\phi}_t = A)}{p(\bar{\phi}_t = A)} = 1. \tag{4.17}$$

For isoenergetic dynamics initiated from a microcanonical ensemble [42, 46],

$$\frac{p(\bar{\phi}_t = A)}{p(\bar{\phi}_t = -A)} = \langle \exp(\bar{\Lambda}_t t)\rangle^{-1}_{\bar{\phi}_t=A} = \langle \exp(V F_e \overline{\beta J}_t t)\rangle^{-1}_{\bar{\phi}_t=A} \tag{4.18}$$

while for isokinetic or Nosé–Hoover dynamics initiated from a canonical ensemble [42, 46],

$$\frac{p(\bar{\phi}_t = A)}{p(\bar{\phi}_t = -A)} = \langle \exp(V F_e \beta \bar{J}_t t)\rangle^{-1}_{\bar{\phi}_t=A}. \tag{4.19}$$

Formulas for other ergodically consistent ensembles can be obtained in a similar manner [42, 46]. Results for various phase functions for a system undergoing isoenergetic shear flow have been presented in reference [46].

4.5. *Integrated FT*

The Fluctuation Theorem quantifies the probability of observing time-averaged dissipation functions with complimentary values. The Second Law of Thermodynamics only states that the dissipation should be positive rather than negative. Therefore, it is of interest to construct a fluctuation theorem which predicts the probability ratio that the dissipation function is either positive or negative. When the statistical error is large and the ensemble sample sizes are small, it is useful to be able to predict the probability that the entropy production will be positive. The Integrated form of the FT (IFT) gives a relationship that quantifies the probability of observing Second Law violations in small systems observed for a short time.

The TFT can be written as

$$\frac{p(\bar{\Omega}_t = -A)}{p(\bar{\Omega}_t = A)} = \exp(-At). \tag{4.20}$$

We wish to give the probability ratio of observing trajectories with positive and negative values of $\bar{\Omega}_t$ and so we consider:

$$p_+(t) \equiv p(\bar{\Omega}_t > 0), \quad p_-(t) \equiv p(\bar{\Omega}_t < 0). \tag{4.21}$$

Now

$$\frac{p_-(t)}{p_+(t)} = \frac{\int_0^\infty \mathrm{d}A\, p(\bar{\Omega}_t = -A)}{\int_0^\infty \mathrm{d}A\, p(\bar{\Omega}_t = A)}. \tag{4.22}$$

Using (4.20):

$$\frac{p_-(t)}{p_+(t)} = \frac{\int_0^\infty \mathrm{d}A\, p(\bar{\Omega}_t = -A)}{\int_0^\infty \mathrm{d}A\, p(\bar{\Omega}_t = A)} = \frac{\int_0^\infty \mathrm{d}A\, \exp(-At) p(\bar{\Omega}_t = A)}{\int_0^\infty \mathrm{d}A\, p(\bar{\Omega}_t = A)}. \tag{4.23}$$

The right hand side of this equation is just the ensemble average of $\exp(-\bar{\Omega}_t t)$ evaluated over trajectories which have a positive value of A:

$$\frac{p_-(t)}{p_+(t)} = \langle \exp(-\bar{\Omega}_t t) \rangle_{\bar{\Omega}_t > 0}. \tag{4.24}$$

From (4.24) we can also obtain the reciprocal relationship:

$$\frac{p_+(t)}{p_-(t)} = \frac{1}{\langle \exp(-\bar{\Omega}_t t) \rangle_{\bar{\Omega}_t > 0}}. \tag{4.25}$$

Similarly, it can be shown that

$$\frac{p_+(t)}{p_-(t)} = \langle \exp(-\bar{\Omega}_t t) \rangle_{\bar{\Omega}_t < 0}. \tag{4.26}$$

We note that in actual experiments, where $\langle \Omega \rangle > 0$, equations (4.24) and (4.25) have much smaller statistical uncertainties than (4.26), because rarely observed trajectory segments with highly negative values of $\bar{\Omega}_t$ will have a large influence on the ensemble average in (4.26). Consequently (4.26) should be avoided in numerical calculations or experiments.

Finally we note that equation (4.25) can be used to show that

$$p_-(t) = \frac{\langle \exp(-\bar{\Omega}_t t) \rangle_{\bar{\Omega}_t > 0}}{[1 + \langle \exp(-\bar{\Omega}_t t) \rangle_{\bar{\Omega}_t > 0}]}, \quad p_+(t) = \frac{1}{[1 + \langle \exp(-\bar{\Omega}_t t) \rangle_{\bar{\Omega}_t > 0}]}. \tag{4.27}$$

Thus far all our equations refer to transient experiments. When t is large, corresponding asymptotic expressions can be determined for steady state averages [47].

5. Green–Kubo relations

The Green–Kubo formulae relate the macroscopic, linear transport coefficients of a system to its microscopic equilibrium fluctuations. It has been shown that the Green–Kubo relations (GK) can be derived from the SSFT and the assumption that the distribution of time-averaged dissipative flux is Gaussian [23, 32, 48]. We summarize those arguments here.

For simplicity, we firstly consider the isokinetic case. In this case β is a constant of the motion: $\overline{\beta J}_t = \beta_0 \bar{J}_t$, and the SSFT (4.4) states,

$$\lim_{t \to \infty} \frac{1}{t} \ln \left(\frac{p(\overline{J}_t = A)}{p(\overline{J}_t = -A)} \right) = -A \beta_0 V F_e. \tag{5.1}$$

As the averaging time, t, becomes large compared with the Maxwell time, τ_M, which characterizes serial correlations in the dissipative flux, contributions to the trajectory segment averages of the dissipative flux, $\{\bar{J}_t\}$, become statistically independent and therefore according to the Central Limit Theorem, (CLT), *near its mean*, the distribution should *approach* a Gaussian. We also note that the SSFT requires averaging times that are longer than the Maxwell time. In order to obtain GK relations, we require that the distribution can be approximated by a Gaussian for values of $\bar{J}_t \approx \pm \langle J \rangle$, where $t \approx O(\tau_M)$, and within a few standard deviations of these values [48]. As the averaging time becomes longer, the variance of the distribution of time-averaged dissipative fluxes becomes ever smaller and it is less and less likely that it will be well approximated by a Gaussian at $\bar{J}_t \approx \langle J \rangle$ and $\bar{J}_t \approx -\langle J \rangle$. Furthermore, as the field is increased, $|\langle J \rangle|$ increases, and therefore this condition will be violated

at earlier times. If the condition is violated at times that are less than several τ_M, then the field-dependent GK relations will not be valid. The distribution can only be Gaussian for complementary values of the dissipative flux $\pm\langle J\rangle$ and within a few standard deviations of this value in the zero field limit [48]. Only in the linear regime will the field-dependent GK relations provide a good estimation of the field-dependent transport coefficient [48].

If the distribution is Gaussian, it is easy to show that,

$$\lim_{(t\to\infty)} \frac{1}{t} \ln\left(\frac{p(\overline{\beta J}_t = \beta A)}{p(\overline{\beta J}_t = -\beta A)}\right) = \lim_{(t\to\infty)} \frac{1}{t} \ln\left(\frac{p(\overline{J}_t = A)}{p(\overline{J}_t = -A)}\right) = \lim_{(t\to\infty)} \frac{2A\langle J\rangle_{F_e}}{t\sigma^2_{\bar{J}_t}(t;F_e)}, \tag{5.2}$$

where the averaging time is t, the applied field is F_e and $\sigma^2_{\bar{J}_t}(t;F_e)$ is the variance of the distribution of $\{\bar{J}_t\}$:

$$\sigma^2_{\bar{J}_t}(t;F_e) = \langle(\overline{J}(t) - \langle J\rangle_{F_e})^2\rangle_{F_e}. \tag{5.3}$$

From equations (5.1) and (5.2) we see that *if* the t-averaged dissipative fluxes are Gaussian near $\bar{J}_t \approx \pm\langle J\rangle$ (i.e. in the zero field limit), then the limiting zero field transport coefficient is given as,†

$$L(0) = \lim_{F_e\to 0} L(F_e) = \lim_{F_e\to 0} \frac{-\langle J\rangle_{F_e}}{F_e} = \lim_{(t\to\infty)} \tfrac{1}{2}\beta_0 V t\sigma^2_{\bar{J}_t}(t;F_e=0). \tag{5.4}$$

This equation constitutes an Einstein relation for the linear transport coefficient, $L(0)$. Except for the case of colour conductivity where (5.4) is equivalent to the standard Einstein expression for the self diffusion coefficient [49], these zero field Einstein relations are not well known [48, 50].

If

$$\tilde{L}_J(s;F_e) \equiv \beta_0 V \int_0^\infty \mathrm{d}t \exp(-st)\langle(J(0) - \langle J\rangle_{F_e})(J(t) - \langle J\rangle_{F_e})\rangle_{F_e} \tag{5.5}$$

is the frequency- and field-dependent Green–Kubo transform (GK) of the dissipative flux, then (see reference [32], and the appendix of reference [48]),

$$\lim_{(t\to\infty)} t\sigma^2_{\bar{J}_t}(t;F_e) = \frac{2\tilde{L}_J(0;F_e)}{\beta_0 V} + \lim_{(t\to\infty)} \frac{2\tilde{L}'_J(0;F_e)}{\beta_0 V t} = \frac{2\tilde{L}_J(0;F_e)}{\beta_0 V}, \tag{5.6}$$

where

$$\tilde{L}'_J(s;F_e) \equiv \frac{\mathrm{d}\tilde{L}_J(s;F_e)}{\mathrm{d}s}. \tag{5.7}$$

† Note that a Gaussian distribution does not imply a FT. The FT is a much stronger statement: it specifies the relationship between the mean and the standard deviation if the distribution is Gaussian, i.e. $\sigma(2/\Omega_t) = 2\langle\Omega\rangle/t$. We note that the FT of course, does not require the distribution to be Gaussian.

Combining equations (5.4) and (5.6), shows that *if* the t-averaged dissipative fluxes are Gaussian near $\bar{J}_t \approx \pm\langle J\rangle$ (i.e. in the zero field limit), then the linear transport coefficient, $L(0)$, is given by the zero frequency Green–Kubo transform of the dissipative flux,

$$L(0) = \lim_{F_e \to 0} L(F_e) = \tilde{L}_J(0; F_e = 0) = \beta_0 V \int_0^\infty \mathrm{d}t\, \langle J(0)J(t)\rangle_{F_e=0}. \tag{5.8}$$

This is the well known Green–Kubo expression for the linear transport coefficient, $L(0)$. The relationship between the FT and GK expressions in the linear regime has been considered previously [7, 32, 51–53].

In the isoenergetic case a similar analysis can be applied. *If* the distribution is Gaussian near $\bar{J}_t \approx \pm\langle J\rangle$, we have,

$$\lim_{(t\to\infty)} \frac{1}{t} \ln\left(\frac{p(\overline{[\beta J]}_t = B)}{p(\overline{[\beta J]}_t = -B)}\right) = \lim_{(t\to\infty)} \frac{2B\langle [\beta J]\rangle_{F_e}}{t\sigma^2_{\overline{[\beta J]}_t}} = \lim_{(t\to\infty)} \frac{-2B\langle [\beta J]\rangle_{F_e}}{t\sigma^2_{\overline{[\beta J]}_t t}}. \tag{5.9}$$

Combining this equation with the fluctuation system for this system (see table 4.1) shows that the Einstein relation for linear transport coefficients of isoenergetic systems is,

$$\langle\beta\rangle_{F_e} L(F_e = 0) \equiv \lim_{F_e\to 0} \frac{-\langle [\beta J]\rangle_{F_e}}{F_e} = \lim_{(t\to\infty)} \tfrac{1}{2} V t \sigma^2_{\overline{[\beta J]}_t}(t; F_e = 0), \tag{5.10}$$

while the corresponding Green–Kubo relation for isoenergetic systems is,

$$L(0) = \lim_{F_e\to 0} L(F_e) = \tilde{L}_J(0;0) \equiv V\langle\beta\rangle^{-1}_{F_e=0} \int_0^\infty \mathrm{d}t\, \langle [\beta J](0)[\beta J](t)\rangle_{F_e=0}. \tag{5.11}$$

Numerical tests verify the correctness of this relationship [48].

6. Causality

6.1. *Introduction*

The Transient Fluctuation Theorem and time dependent response theory are meant to model the following types of experiment. One *begins* an experiment with an ensemble of systems characterized by some *initial* (often equilibrium) distribution function. One *then* does something to the system (applies or turns off a field as the case may be) and one tries to predict what *subsequently* happens to the system. It is completely natural that we assume that the probability of subsequent events can be predicted from the probabilities of finding initial phases and a knowledge of preceding changes in the applied field and environment of the system. The future state of the system is computed solely from the probabilities of events in the past. This is called the Axiom of *Causality*.

It is logically possible to compute the probability of occurrence of present states from the probabilities of future events, but this seems totally unnatural. Will the electric light be on now, because at some time in the (near) *future*, we *will* throw a switch which applies the necessary voltage? A major problem with this approach is that at any given instant, the future states are generally not known! In spite of these philosophical and practical difficulties, we will explore the logical consequences of the (unphysical) Axiom of *Anticausality*.

Mechanics is indifferent to the direction of time—Hamilton's Action Principle shows this with great clarity. However, mechanics on its own does not give us

enough information to predict experimental results. We need to know initial or logically, final conditions. When we model laboratory experiments we require initial conditions because this is precisely how the experiments are conducted and because initially, the final state of the system is generally not known. Although we can mimic the effects of time flowing backwards (time decrementing) by applying a time reversal mapping to a set of phases, time nevertheless evolves in a positive sense. Indeed the effectiveness of the time reversal mapping relies on the fact that time only increases.

We now show that if we derive Green–Kubo relations for the transport coefficients defined by *anticausal* constitutive relations, firstly, these anti-transport coefficients have the opposite sign to their causal counterparts and secondly, it becomes overwhelmingly more likely to observe Second Law violating anticausal non-equilibrium steady states [40]. This argument shows that in an anticausal world it becomes overwhelmingly probable to observe *final* equilibrium microstates that evolved from Second Law violating non-equilibrium steady states. Although this behaviour is *not* seen in the macroscopic world, *anticausal* behaviour is permitted by the solution of the time reversible laws of dynamics and we demonstrate, using computer simulation, how to find phase space trajectories which exhibit *anticausal* behaviour.

6.2. *Causal and anticausal constitutive relations*

Consider the component of the linear response at time t_1, $\mathrm{d}B(t_1)$, of a system characterized by a response function $L(t_1, t_2)$. The response is due to the application of an external force F, acting for an infinitesimal time $\mathrm{d}t_2(>0)$, at time t_2, could be written as,

$$\delta B(t_1) = L(t_1, t_2)F(t_2)\delta t_2. \tag{6.1}$$

This is the most general linear, scalar relation between the response and the force components. If the response of the system is independent of the time at which the experiment is undertaken (i.e. if the same response is generated when both times appearing in (6.1) are translated by an amount t: $t_2 \to t_2 + t$, $t_1 \to t_1 + t$), then the response function $L(t_1, t_2)$ is solely a function of the difference between the times at which the force is applied and the response is monitored,

$$\delta B(t_1) = L(t_1 - t_2)F(t_2)\delta t_2. \tag{6.2}$$

The invariance of the response to time translation is called the *assumption of stationarity*. Equation (6.2) does not in fact describe the results of actual experiments because it allows the response at time t_1 to be influenced not only by forces in the past, $F(t_2)$, where $t_2 < t_1$, but also by forces that have not yet been applied $t_2 > t_1$ [54]. We therefore distinguish between the causal and anticausal response components,

$$\delta B_{\mathrm{C}}(t_1) \equiv +L_{\mathrm{C}}(t_1 - t_2)F(t_2)\delta t_2, \quad t_1 > t_2 \tag{6.3a}$$

$$\delta B_{\mathrm{A}}(t_1) \equiv -L_{\mathrm{A}}(t_1 - t_2)F(t_2)\delta t_2, \quad t_1 < t_2. \tag{6.3b}$$

Later, we will prove that $L_{\mathrm{C}}(t) = L_{\mathrm{A}}(-t)$.

Considering the response at time t to be a linear superposition of influences due to the external field at all possible previous (or future) times gives,

$$B_{\mathrm{C}}(t) = \int_{-\infty}^{t} L_{\mathrm{C}}(t - t_1)F(t_1)\,\mathrm{d}t_1 \tag{6.4a}$$

for the causal response and,

$$B_{\mathrm{A}}(t) = -\int_{t}^{+\infty} L_{\mathrm{A}}(t - t_1)F(t_1)\,\mathrm{d}t_1 \tag{6.4b}$$

for the anticausal response.

6.3. *Green–Kubo relations for the causal and anticausal linear response functions*

To make this discussion more concrete we will discuss Green–Kubo relations for shear viscosity [16]. Analogous results can be derived for each of the Navier–Stokes transport coefficients. We assume that the regression of fluctuations in a system at equilibrium, whose constituent particles obey Newton's equations of motion, are governed by the Navier–Stokes equations. We consider the wave vector dependent transverse momentum density,

$$J_{\perp}(k_y, t) \equiv \sum_i p_{xi}(t) \exp\left[ik_y y_i(t)\right], \tag{6.5}$$

where p_{xi} is the x-component of the momentum of particle i, y_i is the y-coordinate of particle i and k_y is the y-component of the wave vector. The (Newtonian) equations of motion can be used to calculate the rate of change of the transverse momentum density. They give,

$$\begin{aligned}\dot{J}_{\perp} &= ik_y\left[\sum_i p_{xi}p_{yi}\exp\left(ik_y y_i\right) + \frac{1}{2}\sum_{i,j} y_{ij}F_{xij}\,\frac{1-\exp\left(ik_y y_{ij}\right)}{ik_y y_{ij}}\exp\left(ik_y y_i\right)\right] \\ &\equiv ik_y P_{yx}(k_y, t).\end{aligned} \tag{6.6}$$

In this equation F_{xij} is the x-component of the force exerted on particle i by particle j, $y_{ij} \equiv y_j - y_i$ and P_{xy} is the xy-component of the pressure tensor.

We now consider the response of the pressure tensor to a strain rate, γ, applied to the fluid for $t > 0$ in the causal system and for $t < 0$ in the anticausal system. In the causal system the strain rate is turned on at $t = 0$ while in the anticausal system the strain rate is turned off at $t = 0$. Since the pressure tensor is related to the time derivative of the transverse momentum current by (6.6) and the strain rate is related to the Fourier transform of the transverse momentum density by $\gamma(k_y, t) = -ik_y J_{\perp}(k_y, t)/\rho$, the most general linear, stationary and causal constitutive relation can be written as,

$$\dot{J}_{\perp}(k_y, t) = \frac{-k_y^2}{\rho}\int_0^t \eta_{\mathrm{C}}(k_y, t - s)J_{\perp}(k_y, s)\,\mathrm{d}s, \quad t > 0, \tag{6.7}$$

where η_{C} is the causal response function (or memory function) and ρ is the density. The corresponding anticausal relation is,

$$\dot{J}_{\perp}(k_y, t) = \frac{k_y^2}{\rho}\int_t^0 \eta_{\mathrm{A}}(k_y, t - s)J_{\perp}(k_y, s)\,\mathrm{d}s, \quad t < 0, \tag{6.8}$$

where η_{A} is the anticausal 'response' function. Note that because $t < 0$, we find that the argument $(t - s)$ in (6.8) is less than zero, and we are indeed exploring the response of the system at times less than zero, which is prior to the changes in the strain rate that occur at times greater than zero!

It is straightforward to use standard techniques to evaluate the Green–Kubo relations for the causal and anticausal shear viscosity coefficients. In the anticausal case it is important to remember that the usual Laplace transform,

$$\tilde{F}(s) \equiv \int_0^{+\infty} F(t) \exp(-st)\, \mathrm{d}t, \quad t \geq 0 \tag{6.9}$$

is inappropriate and needs to be replaced by an anti-Laplace transform,

$$\hat{F}(s) \equiv \int_{-\infty}^{0} F(t) \exp(st)\, \mathrm{d}t, \quad t \leq 0. \tag{6.10}$$

[Note: $\hat{F}(s) = \int_0^{\infty} F(-t) \exp(-st)\, \mathrm{d}t = \tilde{F}'(s)$, $t \geq 0$, where $F'(t) \equiv F(-t)$.] We note that the anti-Laplace transform of a time derivative is $\hat{\dot{F}}(s) = F(0) - s\hat{F}(s)$ and that the anti-Laplace transform of a convolution is the product of the anti-Laplace transforms of the convolutes. By multiplying both sides of equations (6.7) and (6.8) by $J_\perp(-k_y, 0)$ and taking an (equilibrium) ensemble average, one can easily derive the following relations for the shear viscosity and the anticausal shear viscosity,

$$\left.\begin{aligned} \tilde{C}(k_y, s) &= \frac{C(k_y, 0)}{s + \dfrac{k_y^2 \tilde{\eta}_{\mathrm{C}}(k_y, s)}{\rho}}, \\ \hat{C}(k_y, s) &= \frac{C(k_y, 0)}{s + \dfrac{k_y^2 \hat{\eta}_{\mathrm{A}}(k_y, s)}{\rho}}, \end{aligned}\right\} \tag{6.11}$$

where

$$C(k_y, t) \equiv \langle J_\perp(k_y, t) J_\perp(-k_y, 0) \rangle, \quad \forall\, t. \tag{6.12}$$

More useful relations for the viscosity coefficients, especially at $k = 0$, can be obtained by utilising the equilibrium stress autocorrelation function,

$$N(k_y, t) \equiv \frac{1}{V k_{\mathrm{B}} T} \langle P_{yx}(k_y, t) P_{yx}(-k_y, 0) \rangle, \quad \forall\, t. \tag{6.13}$$

Using the fact that $\hat{N} = -\hat{\dot{\tilde{C}}}/k_y^2 V k_{\mathrm{B}} T$, one can show [16, 55],

$$\left.\begin{aligned} \tilde{\eta}_{\mathrm{C}}(k_y, s) &= \frac{\tilde{N}(k_y, s)}{1 - k_y^2 \tilde{N}(k_y, s)/\rho s}, \\ \hat{\eta}_{\mathrm{A}}(k_y, s) &= \frac{\hat{N}(k_y, s)}{1 - k_y^2 \hat{N}(k_y, s)/\rho s}. \end{aligned}\right\} \tag{6.14}$$

At zero wave vector, we find that the causal and anticausal memory functions are both given by the equilibrium autocorrelation function of the pressure tensor,

$$\left.\begin{aligned} \eta_{\mathrm{C}}(t) &= \eta_{\mathrm{A}}(-t), \quad \text{where } t > 0 \\ &\equiv \eta(t), \quad \forall\, t \\ &= \frac{V}{k_{\mathrm{B}} T} \langle P_{yx}(t) P_{yx}(0) \rangle, \end{aligned}\right\} \tag{6.15}$$

where we have used $P_{yx}(t) V = \lim_{k \to 0} P_{yx}(k_y, t)$. Since equilibrium autocorrelation functions are symmetric in time, one does not have to distinguish between the

positive and negative time domains. This proves our assertion made in section 6.2 that $L_C(t) = L_A(-t)$.

Using equations (6.6)–(6.8) and taking the zero wave vector limit, we obtain the causal response of the xy-component of the pressure tensor,

$$P_{yxC}(t) = -\int_0^t \eta(t-s)\gamma(s)\,\mathrm{d}s \quad t > 0 \tag{6.16}$$

and the anticausal response is,

$$P_{yxA}(t) = \int_t^0 \eta(t-s)\gamma(s)\,\mathrm{d}s \quad t < 0. \tag{6.17}$$

In the linear regime close to equilibrium the entropy production per unit time, $\mathrm{d}\Sigma/\mathrm{d}t$, is given by,

$$\frac{\mathrm{d}\Sigma}{\mathrm{d}t} = -P_{yx}(t)\gamma(t)V/T, \tag{6.18}$$

where $\gamma(t)$ is the time-dependent strain rate. From equations (6.16) and (6.17), it is easy to see that if we conduct two shearing experiments, one on a causal system with a strain rate history $\gamma_C(t)$ and one on an anticausal system with $\gamma_A(t) = \pm\gamma_C(-t)$, then

$$\left.\frac{\mathrm{d}\Sigma(t)}{\mathrm{d}t}\right|_A = \left.\frac{-\mathrm{d}\Sigma(-t)}{\mathrm{d}t}\right|_C. \tag{6.19}$$

This proves that if the causal system satisfies the Second Law of Thermodynamics then the anticausal system must violate that Law and *vice versa*.

6.4. *Example: the Maxwell model of viscosity*

In this section we examine the consequences of the causal and anticausal response by considering the Maxwell model for linear viscoelastic behaviour [16]. If we consider the causal response of a system to a two step strain rate ramp:

$$\left.\begin{aligned}\gamma_C(t) &= a \quad 0 < t < t_1\\ \gamma_C(t) &= b \quad t_1 < t < t_2\end{aligned}\right\} \tag{6.20}$$

then use the Maxwell memory kernel,

$$\eta(t) = G_\infty \exp(-|t|/\tau_M), \quad \forall\, t \tag{6.21}$$

in equation (6.16) and the fact that the causal, η_C, and anticausal, η_A, Maxwell shear viscosities in the zero frequency limit are

$$\eta_C = \eta_A = G_\infty \tau_M = \eta, \tag{6.22}$$

we find that the causal response is:

$$\begin{aligned}P_{xyC}(t) &= -a\eta[1 - \exp(-t/\tau_M)], \quad 0 < t < t_1\\ P_{xyC}(t) &= -a\eta\{\exp[-(t-t_1)/\tau_M] - \exp(-t/\tau_M)\}\\ &\quad - b\eta\{(1 - \exp[-(t-t_2)/\tau_M]\}, \quad t_1 < t < t_2.\end{aligned} \tag{6.23}$$

If we now consider the corresponding anticausal experiment with strain rate histories given by:

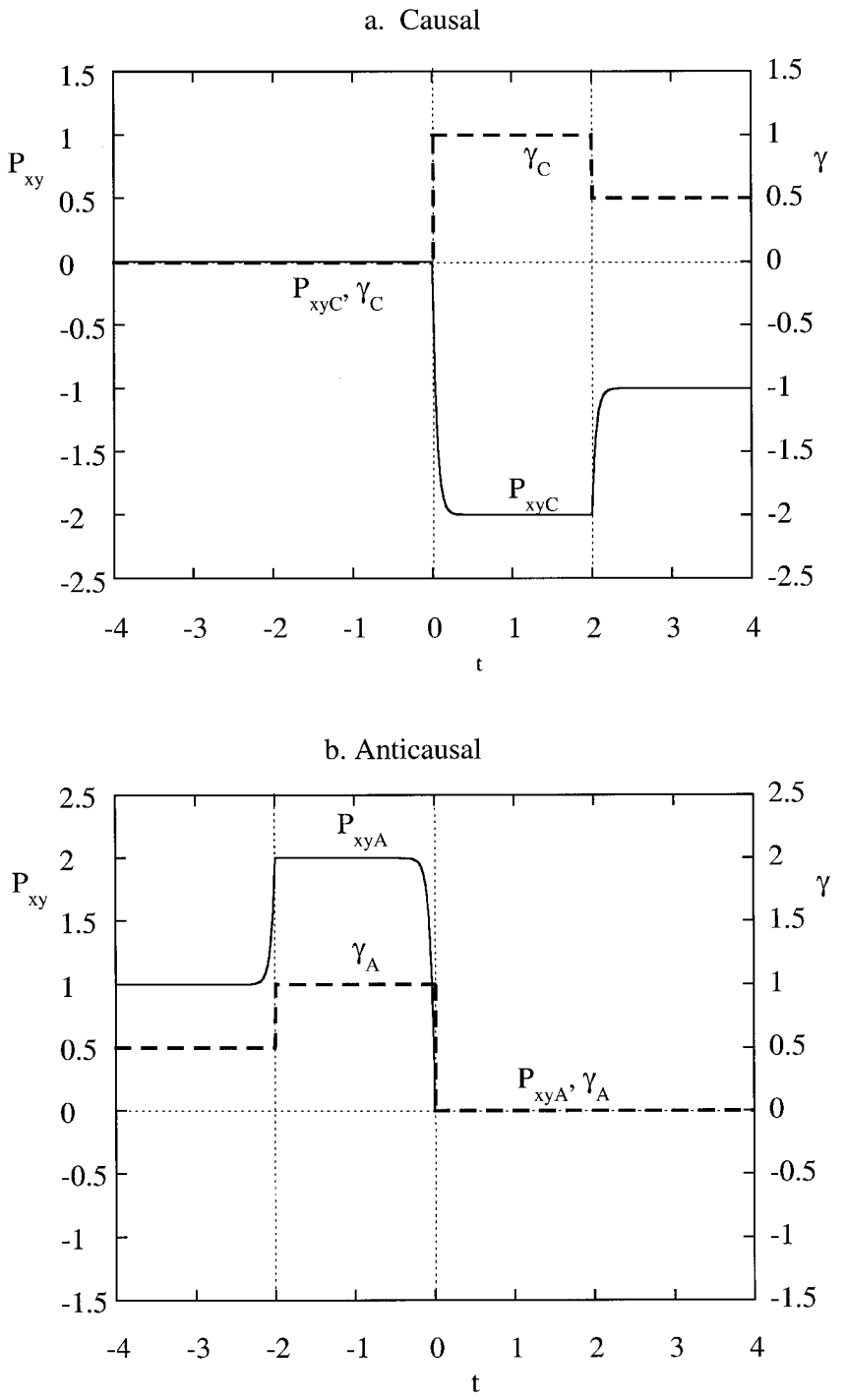


Figure 6.1. A schematic diagram of the (*a*) causal and (*b*) anticausal response of P_{xy} to a two-step strain rate ramp determined using the Maxwell model for linear viscoelastic behaviour with $G_\infty = 40$ and $\tau_M = 0.05$ (solid line). In both cases the time dependence of the strain rate is shown as a dashed line.

$$\left.\begin{array}{ll} \gamma_A(t) = a & -t_1 < t < 0 \\ \gamma_{\mathrm{A}}(t) = b & -t_2 < t < -t_1 \end{array}\right\} \tag{6.24}$$

we find that the anticausal response is:

$$\begin{aligned} P_{xy\mathrm{A}}(t) &= a\eta[1 - \exp(t/\tau_{\mathrm{M}})], \quad -t_1 < t < 0 \\ P_{xy\mathrm{A}}(t) &= a\eta\{\exp[(t+t_1)/\tau_{\mathrm{M}}] - \exp(t/\tau_{\mathrm{M}})\} \\ &\quad + b\eta\{1 - \exp[(t+t_2)/\tau_{\mathrm{M}}]\}, \quad -t_1 < t < -t_2. \end{aligned} \tag{6.25}$$

From equations (6.23) and (6.25) it is clear that,

$$P_{xy\mathrm{A}}(t) = -P_{xy\mathrm{C}}(-t). \tag{6.26}$$

These response functions are shown graphically in figure 6.1. A two-step strain rate ramp with $a = 1.0$, $b = 0.5$, $t_1 = 2$ and $t_2 = 4$ was considered. Equations (6.23) and (6.25) were used to predict the causal and anticausal responses, respectively, of the xy-component of the pressure tensor. Values of $G_\infty = 40.0$ and $t = 0.05$ were used in the model. These values were obtained from approximate fits to computer simulation data (see section 6.5).

The data in figure 6.1 show that for the causal response, P_{xy} is zero at equilibrium ($t \leq 0$) and decreases when the field is applied until the steady state value is obtained. It remains at the steady value until $t = 2$, at which time the strain rate is reduced. Since this system is causal, no change in P_{xy} occurs until *after* the strain rate is reduced, when it increases until the system reaches a new steady state. We display the anticausal response from $t = -4$ where it is in an antisteady state. Just *before* the strain rate is increased (at $t = -2$), P_{xy} increases to a new antisteady state value. Using equation (6.18) we see that the the causal response is entropy increasing and Second Law satisfying, whereas the anticausal response is entropy decreasing and Second Law violating.

6.5. *Phase space trajectories for ergostatted shear flow*

We now examine the causal and anticausal response on a microscopic scale and we consider the relative probability of observing Second Law satisfying and Second Law violating trajectories by studying a ergostatted system of N particles under shear.

The ergostatted SLLOD equations of motion (1.11), (1.12) are time reversible [16]. Therefore for every i-segment $\mathbf{\Gamma}_{(i)}(t)$, $(0 < t < \tau)$, there exists a conjugate trajectory segment $\mathbf{\Gamma}_{(i^{(\mathrm{K})})}(t)$, $(0 < t < \tau)$ with the property that, $P_{xy}(\mathbf{\Gamma}_{(i^{(\mathrm{K})})}(t)) = -P_{xy}(\mathbf{\Gamma}_{(i)}(-t))$, $(0 < t < \tau)$. Thus, the t-averaged shear stress $\bar{P}_{xy,i,t} \equiv 1/t \int_0^t \mathrm{d}s\, P_{xy}(\mathbf{\Gamma}_i(s))$ for segment i is equal and opposite to that for its conjugate: $\bar{P}_{xy,i^{\mathrm{K}},t} = -\bar{P}_{xy,i,t}$. We note that since the solution of the equations of motion is a unique function of the initial conditions the conjugate segment is also unique.

We have previously shown that for shear flow conjugate segments may be generated by using a phase space mapping known as a Kawasaki- or K-map [16]. A K-map of a phase, $\mathbf{\Gamma}$, is defined as a time-reversal map which is followed by a y-reflection. In the case of shear flow the K-map leaves the strain rate unchanged but changes the sign of the shear stress, that is $M^{\mathrm{K}}\mathbf{\Gamma} = M^{\mathrm{K}}(x, y, z, p_x, p_y, p_z, \gamma) = (x, -y, z, -p_x, p_y, -p_z, \gamma) \equiv \mathbf{\Gamma}^{(\mathrm{K})}$ [16]. It is straightforward to show that the Liouville operator for the system simulated by equations (1.11) and (1.12), $iL(\mathbf{\Gamma}, \gamma) \equiv \sum [\dot{\mathbf{q}}_i(\mathbf{\Gamma}, \gamma) \cdot \partial/\partial \mathbf{q}_i + \dot{\mathbf{p}}_i(\mathbf{\Gamma}, \gamma) \cdot \partial/\partial \mathbf{p}_i]$, has the property that under a K-

map, $M^{\mathrm{K}} iL(\Gamma,\gamma) = iL(\Gamma^{(\mathrm{K})},\gamma^{(\mathrm{K})}) = -iL(\Gamma,\gamma)$. If we assume a strain rate history such that, $\gamma_{\mathrm{K}}(-t) = \gamma(t)\ \forall t$, then it follows that if a K-map is carried out on an arbitrary phase Γ at $t = 0$, then evolution *forward* in time from $\Gamma^{(\mathrm{K})}$ under a strain rate $\gamma_{\mathrm{K}}(t)$ is equivalent to time evolution *backwards* in time from Γ under the strain rate history $\gamma(t)$, $(t < 0)$,

$$P_{xy}(-t,\Gamma,\gamma(-t)) = \exp\left[-iL(\Gamma,\gamma(-t))t\right]P_{xy}(\Gamma) = -P_{xy}(t,\Gamma^{(\mathrm{K})},\gamma_{\mathrm{K}}(t)). \qquad (6.27)$$

We note that if we do not assume that $\gamma_{\mathrm{K}}(-t) = \gamma(t)\ \forall t$, then there is no general method for generating conjugate trajectory segments. This is because propagators with different strain rates do not commute, and the inverse propagator must therefore retrace the strain rate history of the conjugate propagator but in inverse historical order.

We will now indicate in more detail, how to construct the conjugate segment, $i^{(\mathrm{K})}$, from an arbitrary phase space trajectory segment i [32]. The construction is illustrated in figure 6.2 for the case where the strain rate remains the same for the duration of the trajectory. A trajectory of length τ is generated by solving the equations of motion. The conjugate segment is then constructed by applying a K-map to the phase at the midpoint of the segment $(t = \tau/2)$, $M^{\mathrm{K}}\Gamma_{(2)} = \Gamma_{(5)}$. We then advance in time from the point $(\Gamma_{(5)})$, to $t = \tau$, by solving the equations of motion and also go backwards in time from the K-mapped point, $t = \tau/2$, to $t = 0$. A

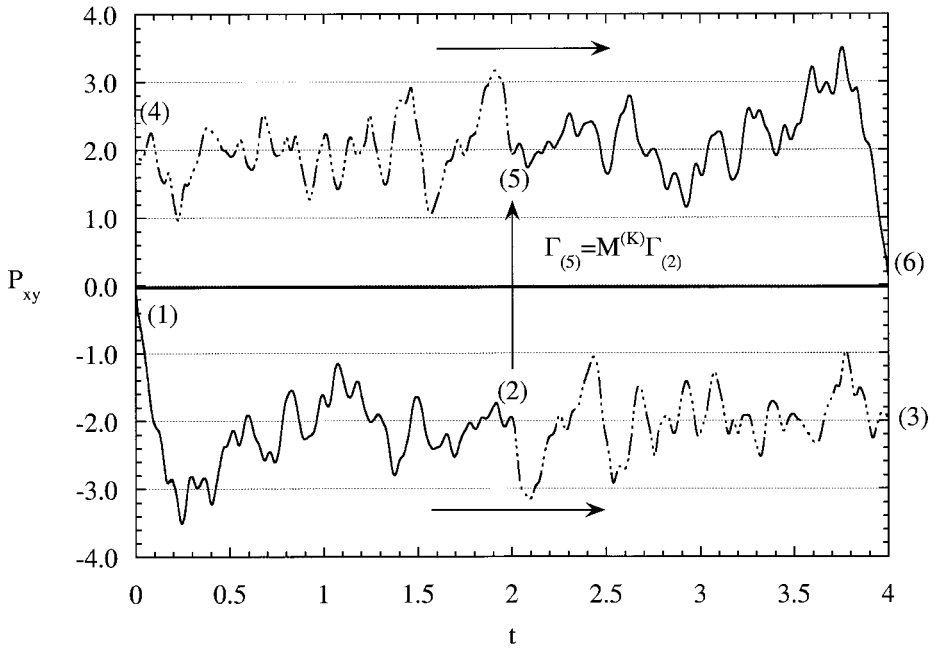


Figure 6.2. P_{xy} for trajectory segments from a simulation of 200 discs at $T = 1.0$ and $n = 0.8$. A constant strain rate of $\gamma = 1.0$ is applied at $t = 0$. The trajectory segment $\Gamma_{(1,3)}$ was obtained from a forward time simulation. At $t = 2$, a K-map was applied to $\Gamma_{(2)}$ to give $\Gamma_{(5)}$. Forward and reverse time simulations from this point give the trajectory segments $\Gamma_{(5,6)}$ and $\Gamma_{(5,4)}$, respectively. If one inverts P_{xy} in $P_{xy} = 0$ and inverts time about $t = 2$, one transforms the $P_{xy}(t)$ values for the antisegments $\Gamma_{(4,6)}$ into those for the conjugate segment $\Gamma_{(1,3)}$.

conjugate trajectory of length τ is thereby produced. This construction has previously been described in more detail [32].

Clearly, the mapped trajectory is a solution of the equations of motion for the system, and therefore it would eventually be observed from the ensemble of starting states. When the K-map is carried out at $t = 0$, the shear stress is inverted and equation (6.27) shows that $P_{xy}(\tau/2 + t, \Gamma) = -P_{xy}(\tau/2 - t, \Gamma^{(K)})$ and similarly $P_{xy}(\tau/2 - t, \Gamma) = -P_{xy}(\tau/2 + t, \Gamma^{(K)})$, therefore for every point on the original trajectory there is a unique point on the mapped trajectory with opposite shear stress. The τ-averaged shear stress of the conjugate trajectory is opposite to that of the original trajectory, that is $\bar{P}_{xy,i^K}(\tau) = -\bar{P}_{xy,i}(\tau)$. Thus, if the original segment was a Second Law satisfying segment then the conjugate segment is a Second Law violating segment, and *vice versa*.

In a *causal* world, which is described by causal macroscopic constitutive relations such as (6.4), observed segments are overwhelmingly likely to be Second Law satisfying. It is a simple matter to apply the arguments of section 2.1 for the special case of ergostatted shear flow where a simple time reversal map cannot be used, and must be replaced by the K-map (see footnote† on page 1542). The condition of ergodic consistency has to be modified slightly to require:

$$f(\Gamma^K(t), 0) \neq 0, \ \forall \Gamma(0). \tag{6.28}$$

The result is the TFT given in (2.10).

6.6. *Simulation results*

We can demonstrate the relationships between the conjugate pairs of trajectories, the Second Law of Thermodynamics and causal and anticausal response using numerical simulations of the system described by equations (1.11) and (1.12). Figure 6.2 shows the response of P_{xy} for a trajectory and its conjugate when a constant strain rate is applied. The response was determined using non-equilibrium molecular dynamics simulations of 200 discs in two Cartesian dimensions. The discs interact via the WCA potential [45],

$$\phi(r) = \begin{cases} 4(r^{-12} - r^{-6}) + 1 & r < 2^{1/6} \\ 0 & r > 2^{1/6}. \end{cases} \tag{6.29}$$

Shearing periodic boundary conditions were used to minimize boundary effects [16]. The system was maintained at a constant kinetic temperature of $T = 1.0$ and the particle density was $n = N/V = 0.8$. An initial phase was selected from an equilibrium distribution and a strain rate of $\gamma = 1.0$ was applied to the system at $t = 0$. A trajectory segment was generated by simulating forward in time to $t = 4$. The conjugate trajectory was constructed using the scheme describe above. Examination of the trajectories shows that $P_{xy}(\tau + t)$ for the Second Law satisfying trajectory is equal in magnitude but opposite in sign to $P_{xy}(\tau - t)$ for the Second Law violating trajectory, where t is the time at which the K-map is applied ($\tau = 2$). These results therefore confirm the relationship between P_{xy} of Second Law satisfying trajectories and Second Law violating conjugate trajectories given by equation (6.27).

The causality of the response is more clearly demonstrated in figure 6.3 where the response of P_{xy} to a strain rate ramp is shown. $P_{xy}(t)$ was averaged over 100 individual trajectories to reduce the fluctuations in the steady state and giving a partially ensemble averaged response $\overbrace{P_{xy}(t)}$. In these simulations 56 discs were used.

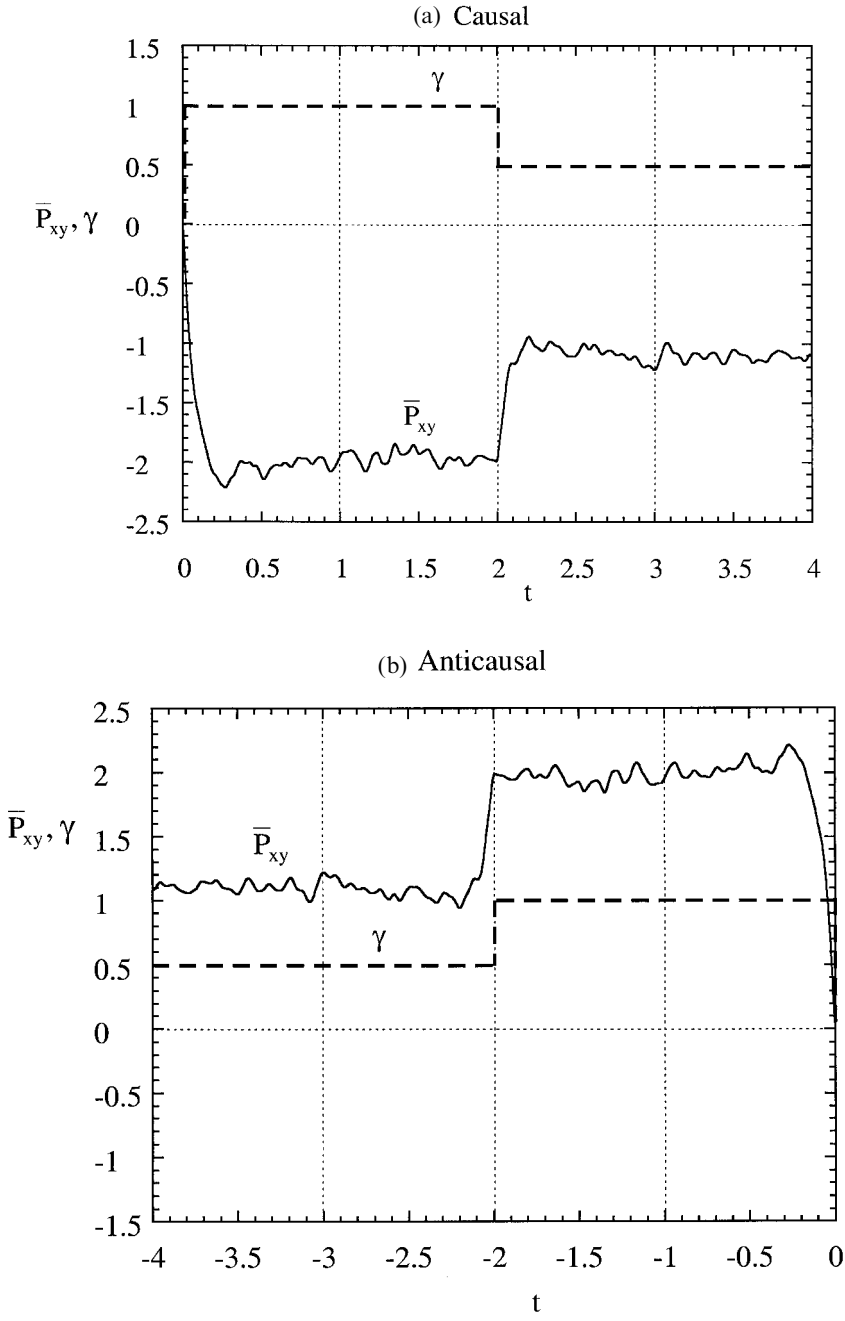


Figure 6.3. $\bar{P}_{xy}$ (solid line) from non-equilibrium molecular dynamics simulations of 56 particles at $T = 1.0$ and $n = 0.8$ undergoing shear flow. The dashed line gives the time-dependence of the strain rate. In (a) $\bar{P}_{xy}$ was determined using 1000 trajectories whose initial phases were selected from an equilibrium distribution, and to which a two step strain rate was applied. (b) shows $\bar{P}_{xy}$ for their conjugate trajectories. The conjugate trajectories were obtained by applying a K-map to the phase of the trajectory at $t = 2$, simulating forward and backward in time from this point and translating in time so that the conjugate trajectory ends at $t = 0$. Note that the strain rate history of the conjugate trajectory is reversed.

The initial phases of the trajectories shown in figure 6.3 were sampled from the equilibrium distribution at $t = 0$. $\widehat{P_{xy}}$ is close to zero at equilibrium and decreases to near a steady state value after the field is applied. After the strain rate is reduced, $\widehat{P_{xy}}$ increases towards a new steady state value.

The conjugate trajectories are shown in figure 6.3. They were constructed as described above and translated in time to begin at $t = -4$. At this time, the system is in an antisteady state and $\widehat{P_{xy}}$ remains near its antisteady state value until just *before* the the strain rate is changed, when it increases towards a new antisteady state value.

In accord with the TFT, these response curves demonstrate that most initial phases (here all 100 randomly selected initial phases) satisfy the Second Law and most phases (again all 100 initial random phases) exhibit response curves that we would describe as having 'causal' characteristics (i.e. the stress responds to prior rather then future, changes in the strain rate). Second Law violating conjugate trajectories respond to the step in the strain rate *before* it is made, so they are *anticausal*. Close inspection of the graph reveals that at all points along pairs of conjugate trajectories, $P_{xy}(t)_{\text{trajectory}} = -P_{xy}(-t)_{\text{conjugate trajectory}}$ which follows from (6.27).

The system used in the simulations corresponds to that examined using the Maxwell model described in section 6.4. Figure 6.3 shows the response, determined by non-equilibrium molecular dynamics simulation, to the same two step strain rate ramp which was used to model the response shown in figure 6.1. Comparison of these response curves indicates that the system is reasonably well represented by the Maxwell model.

7. Experimental confirmation

The importance of experimental verification of theoretical predictions is self-evident, and a number of numerical simulations have been carried out to test various proposed FTs [6, 17, 32, 39–41, 42–44, 46–48, 52, 53, 56–59]. Even if we have rigorous proofs of mathematical theorems (as in the case of the Fluctuation Theorems derived here), the *applicability* of the conditions that are necessary for the construction of such proofs, to real experimental systems can never be taken for granted. When we speak of experimental confirmation of a rigorous theorem, we are really testing the applicability, to natural systems, of the conditions required by the theorem. Experiments however serve a second purpose. They tell us the magnitude of predicted effects—theory is often silent on this issue. For asymptotic theorems, experiments tell how large a variable needs to be before the asymptotic theoretical prediction is accurate. Here we describe experiments which confirm the Fluctuation Theorem and show that for micron sized latex particles trapped by radiation pressure in an optical trap, the Second Law can be violated for macroscopic times, namely three seconds, or so.

The Fluctuation Theorem has been tested experimentally in two different studies [60, 61]. The test by Ciliberto and Laroche considers temperature fluctuations in a fluid undergoing Rayleigh–Benard convection. The study is somewhat inconclusive because they are unable to measure the entropy production directly, and they assumed proportionality between the entropy production and temperature fluctuations. If this assumption is valid, it would still be necessary to know the proportionality constant fully to test the fluctuation theorem. However, this proportionality constant was not known.

A more satisfactory test has recently been carried out that measures the fluctuation in the position of a latex bead in water when it is trapped by laser tweezers [61]. In this experiment, a laser beam forms an optical trap that is used to move a micron sized latex bead with respect to its surrounding solvent, water. The sample cell is on a movable stage, driven in the xy-direction by two piezoelectric crystals, and the particle can be viewed through a microscope that is also mounted on the stage. The trajectory of the particle can be displayed on a computer using a CCD camera, and the particle position is recorded using a quadrant photodiode. The photodiode provides sufficient resolution to determine forces on the particle to 2 femtonewtons. This experiment enables the position of the particle to be measured to a resolution of approximately 15 nm. Details of the experimental design are given in reference [61].

When the laser trap is applied, the latex bead is trapped by a potential well with a force constant that can be modified by adjusting the laser power. The bead then fluctuates about the minimum of the potential energy field formed by the laser trap. Once the bead is trapped (at $t = 0$), the movable stage is given a constant translational velocity that results in a convective field and which imposes, on average, a drag force on the particle. After some time, the system will reach a steady state where the average steady-state position is given by the balance of the average optical trap and drag forces. The average position will from now on be displaced from the minimum of the optical trap, in the direction of the convective field. Since we are interested in investigating the transient response of the particle, the experiment is designed for the time required to reach the steady state position to be within measurable time limits.

These experiments can also be directly simulated using non-equilibrium molecular dynamics simulations. In the simulation, one particle in a fluid is distinguished from the other particles by giving it a colour charge. This colour charge does not alter its interaction with other particles, but allows it to be influenced by an applied colour field. To model the optical trap, a harmonic colour potential is applied to this particle (designated as particle 1),

$$\Phi_{\text{trap}}(\mathbf{q}_1, t) = \tfrac{1}{2}k(\mathbf{q}_1(t) - \mathbf{q}_0(t))^2, \tag{7.1}$$

where k is the force constant and $\mathbf{q}_0(t)$ is the position of the optical trap.

The movement of the stage is simulated by shifting the focus of the laser in the $\pm x$-direction with a constant speed, so $x_0(t) = x_0(0) \pm \mathrm{v}_{\text{opt}}t$, where v_{opt} is the speed of the optical trap. The fluid is contained between walls that run parallel to the direction of the constant force, and the wall particles only are thermostatted using a Nosé–Hoover thermostat. The wall particles are forced to oscillate about their initial positions by a harmonic potential and the box is periodic in the x- and y-directions. For this system, the $t = 0$ distribution function is

$$f(\Gamma, \mathbf{q}_0, \dot{\mathbf{q}}_0, \zeta; 0) \sim \exp\{-\beta[K(\mathbf{p}) + \Phi(\mathbf{q}) + \Phi_{\text{trap}}(\mathbf{q}_1, 0) + \tfrac{1}{2}Q\zeta^2]\}\delta(|\dot{\mathbf{q}}_0|), \tag{7.2}$$

where β is the Boltzmann factor, $\beta = 1/(k_{\text{B}}T)$, T the wall temperature, Q is the effective mass of the thermostat [16] and ζ is the Nosé–Hoover thermostat multiplier (defined below). For $t > 0$, the equations of motion are,

$$\begin{aligned}
\dot{\mathbf{q}}_0 &= \mathbf{i}v_{\text{opt}} \\
\dot{\mathbf{q}}_i &= \frac{\mathbf{p}_i}{m},\ i > 0 \\
\dot{\mathbf{p}}_i &= \mathbf{F}_i - k(\mathbf{q}_i - \mathbf{q}_0(t))\delta_{1,i} + S_i(\mathbf{F}_{\text{w}i} - \zeta \mathbf{p}_i),\ i > 0 \\
\dot{\zeta} &= \frac{1}{Q}\left(\sum_{i=1}^{N_{\text{W}}} \frac{\mathbf{p}_i^2}{m} - d_{\text{C}} N_{\text{W}} k_{\text{B}} T\right),
\end{aligned} \tag{7.3}$$

where $\mathbf{F}_i$ is the force on the ith particle due to interparticle interactions, $\delta_{1,i}$ is the Kronecker delta, N_{W} is the number of wall particles, $S_i = 1$ for wall particles and $S_i = 0$ otherwise and $\mathbf{F}_{\text{w}i}$ is the constraint force on wall particles.

Using equations (2.6), (7.2) and (7.3), the dissipation function that appears in the TFT (2.8) and the ITFT, is

$$\Omega(\mathbf{\Gamma}, t) = -\beta \mathbf{v}_{\text{opt}} \cdot\ k(\mathbf{q}_1(t) - \mathbf{q}_0(t)). \tag{7.4}$$

A histogram of $\bar{\Omega}_t$ obtained from the numerical simulations is shown in figure 7.1, and tests of the FT for this system are presented in figures 7.2 (a,b). Clearly, the TFT and transient IFT are both satisfied.

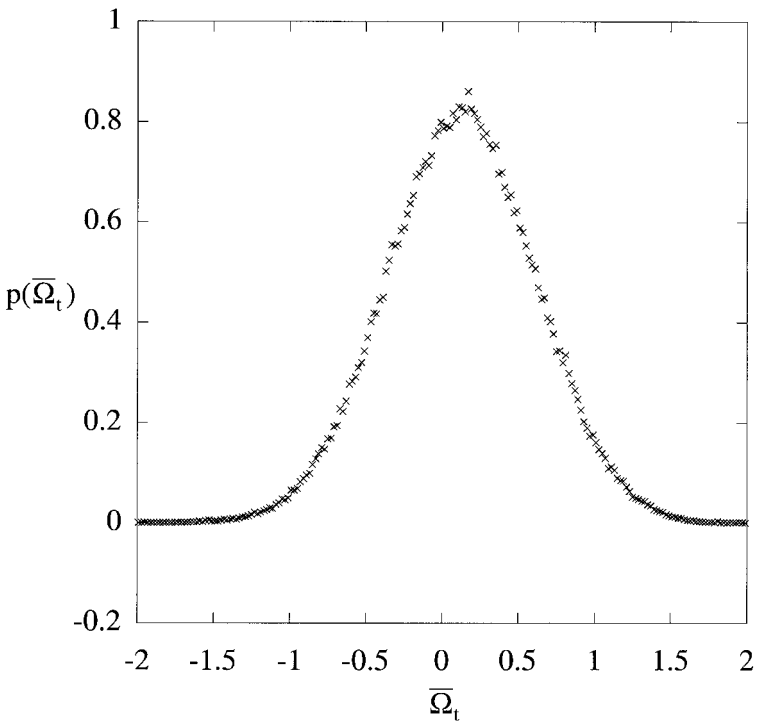


Figure 7.1. A histogram of $\bar{\Omega}_t$ obtained from a simulation of the transient response of a trapped particle to the onset of motion of the stage. A fluid of 32 particles was confined between parallel walls that were thermostatted using a Nosé–Hoover thermostat. The temperature of the system was $T = 1.0$, the particle density of the fluid $n = 0.3$ and the length of the trajectory $t = 1.0$. The optical trap moved with a velocity of $v_{\text{opt}} = 0.5$ and a force constant of 1.0 was used in the harmonic potential that models the optical trap.

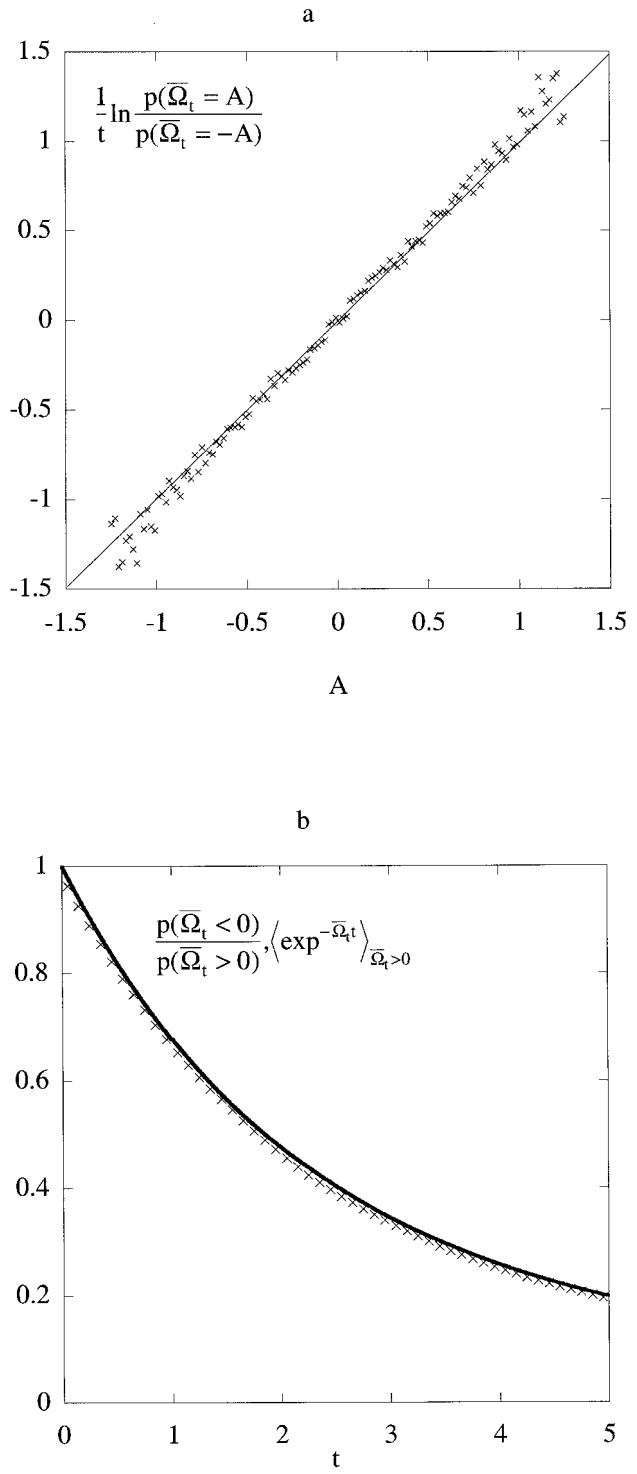


Figure 7.2. (*a*) A test of the TFT, using numerical simulations of the system described in figure 7.1. (*b*) A test of the transient IFT, using numerical simulations of the system described in figure 7.1.

Due to the relatively large statistical error in the experimental results—where 540 trajectories are considered, compared with $\sim 3 \times 10^5$ trajectories in the numerical simulations presented in figures 7.1 and 7.2 (a)—the results were only used to test the transient IFT. In figure 7.3, the experimental results for a test of the transient IFT are presented. The latex beads were 6.3 μm in diameter; the trapping constant was $k \sim 0.1\,\mathrm{pN\,mm^{-1}}$; the optical trap speed was, $v_{opt} = 1.25\,\mu\mathrm{m\,s^{-1}}$ and the reservoir temperature was 300 K. The results shown in figure 7.3 were taken over an ensemble of 540 transient trajectories. As can be seen in figure 7.3 the results are in excellent agreement with the transient Integrated Fluctuation Theorem. It is worth commenting that we see a significant number of ensemble members which violate the Second Law of Thermodynamics for times ~2–3 seconds! Experimental sensitivity limits the maximum time for which violations can be observed.

It is also worth pointing out that the fictitious Nosé–Hoover thermostat used in part to derive the form for the dissipation function (7.4), tested in the experiment, does not actually occur in the experiment. However, since in both the theory and the experiment, the thermostatting occurs only in a region that is remote from the optical trap (the walls), the dissipation function derived in the theory must also be valid for the laboratory experiment. There is simply no way that the molecules near the (experimental or theoretical) optical trap can 'know' how the system is thermostatted in the remote wall regions. Furthermore, the dissipation function appearing in the FT (7.4), involves variables that are not Nosé–Hoover specific quantities. The dissipation function refers only to the trap force and velocity, and the temperature of the thermal reservoir.

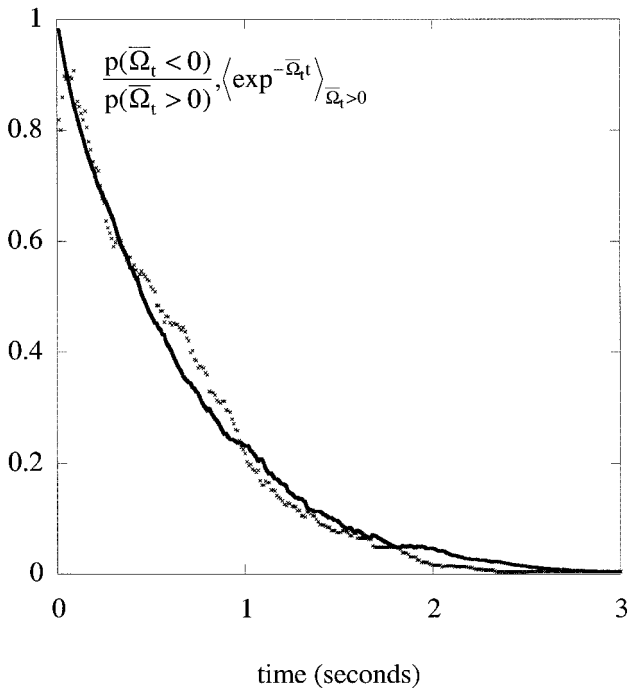


Figure 7.3. A test of the transient IFT for an optical tweezers experiment.

8. Conclusion

In this Review we have derived a family of relationships that have come to be known collectively as the Fluctuation Theorem. Because the FTs deal with *fluctuations*, the FTs take on different forms for different combinations of initial ensemble and dynamics (constant energy, temperature, pressure, etc). Broadly speaking, the FTs give mathematical expressions for the ratio of probabilities that the time average of the *dissipation function*, (2.6), takes on complementary values ($\pm A$).

When the mathematical form of the initial ensemble is known, one can derive what are known as Transient Fluctuation Theorems, which are exact for time-averaging periods of arbitrary duration. Another family of FTs give asymptotic expressions for the ratio of probabilities that in an ensemble of non-equilibrium steady states the time-averaged *entropy absorbed by the thermostat*, takes on complementary values ($\pm A$). If the non-equilibrium steady state is ergodic then the distribution of time averaged entropy absorption can be taken from a single very long phase space dynamical trajectory.

Among the TFTs, if the system starts from some known initial ensemble and evolves in time in contact with a thermostat or ergostat, then the resulting TFTs describe the probability ratio that the time averaged entropy absorbed by the thermostat, takes on complementary values. For such systems the TFT or SSFT gives a proof of the Second Law of Thermodynamics. This is because the dissipation function referred to in such FTs—see (4.3), (4.6)—can be recognized as the rate at which entropy is absorbed by the thermostat.

However, if the system is not thermostatted (this excludes all SSFTs of course), the dissipation function referred to in the relevant TFT is not usually recognizable as an entropy production or absorption. For example, in section 4.3 we dealt with the adiabatic relaxation of a modulation in colour density in a fluid of otherwise identical particles. The relaxation takes place under a colour-blind Hamiltonian—the natural Hamiltonian for a system of identical but interacting particles. From the initiation of the experiment, the motion is conservative and Hamiltonian. The colour labels that we imposed on the fluid particles initially, have no thermodynamic relevance—there is no *thermodynamic* entropy production. In spite of this, the resulting TFT enables us to prove that with overwhelming likelihood (exponential in time and system size) the initial colour modulation will relax (rather than be amplified) and at long times the system will be homogeneous with respect to colour—in complete accord with Le Chatelier's Principle [12].

The FTs are quite general and they are not restricted to the linear response regime close to equilibrium. However, as one moves further away from equilibrium, either the system size or the observation time must be shortened in order actually to observe significant numbers of fluctuations that violate either the Second Law or Le Chatelier's Principle. In the linear response regime close to equilibrium, the SSFT can be used to derive the famous Green–Kubo or Einstein relations for linear transport coefficients. Outside the linear response regime the nonlinear Green–Kubo and Einstein relations, unlike the SSFT, are not valid. In this regime the proof of the Green–Kubo relations from the SSFT breaks down because the Central Limit Theorem is insufficiently strong: it does not apply sufficiently many standard deviations from the mean.

Various generalizations of the FTs are possible. In section 4.4 we derived FTs for time averages of arbitrary functions which have a definite parity under time reversal

symmetry. Thus, these Generalized FTs can be applied to variables other than the dissipation function. For unthermostatted systems where the dissipation function is not (generally) an entropy production, the Generalized FT can be applied to the entropy production in order to prove the correctness of the Second Law for these systems [43].

In section 4.5 we developed an integrated FT which gives the probability ratio that the time-averaged dissipation function is either positive or negative. When applied to a thermostatted system such as a particle in a thermostatted optical trap, the relevant IFT gives the probability ratio that the entropy absorbed by the heat bath is either positive or negative. This expression has been tested numerically and, more significantly, in an actual laboratory experiment; see section 7. In the Optical Tweezers experiment, micron sized latex spheres were observed to move both towards and against a piconewton sized optical force for periods of approximately 3 seconds. The experimentally observed probability ratios are in statistical agreement with the prediction of the IFT.

There is an important practical application of the FTs. In recent years there has been much talk about nanotechnology. People believe that one can scale down machines, devices and engines to nanometer sizes for a wide range of biological, electronic and technological purposes. The FTs point out that there is a fundamental limit to this scaling down process, that small engines are *not* simply rescaled versions of their larger counterparts. If the work performed during the duty cycle of any machine is comparable to the thermal energy per degree of freedom in the system (i.e. $k_B T$), the FTs say that there is a significant probability that the machine will actually run backwards. In violation of the Second Law, heat energy from the surroundings can be converted into useful work to provide sufficient energy for the machine to run in reverse. This is an inescapable property of Nature which places a fundamental constraint on the operation of nanomachines. As you scale down machines they inescapably run in a mode: 'two steps forward and one step back'. The ratio of forward to backward steps is given by the FT. This must also be the way that living sub-cellular organelles (machines) operate.

We have a few remarks regarding the reversibility paradox. If every microscopic law or axiom of Nature is symmetric under time reversal symmetry, then obviously one cannot derive time asymmetric theorems such as the FTs. Somewhere in the derivations of the FT given in section 2.3, we must have introduced a time asymmetric assumption. That assumption was the Axiom of Causality. We computed the probability of subsequent events (time averages of dissipation functions) from the probabilities of initially observing those states from which the subsequent phase space trajectories evolved (6.3 *a*), (6.4 *a*).

It is logically possible that the Axiom of Causality could be replaced by its time conjugate: the Axiom of Anti-Causality. In section 6.3 we showed how an assumption of anticausality would lead to a response with anticausal characteristics and we also showed that an assumption of anticausality (6.4 *b*) leads to anti-Green–Kubo transport coefficients which not only lead to negative entropy production but inescapably lead to characteristic anticausal responses—dissipative fluxes responding in advance of future changes in the applied fields (see figure 6.1)! We have thus shown that there is a deep connection between Causality and the Second Law of Thermodynamics. We cannot violate the Second Law for long and still satisfy causality.

Had we employed the fictitious Axiom of Anticausality (6.4 *b*), then we would have derived an *anti-Fluctuation Theorem*. Instead of equation (2.8) we would have in its place

$$[p(\bar{\Omega}_t = A)]/[p(\bar{\Omega}_t = -A)] = \exp[-At]!$$

We remark that for dilute gases an analogous state of affairs exists with regard to the calculation of transport coefficients from the Boltzmann equation. The Boltzmann equation is time irreversible and leads to Second Law satisfying transport coefficients. This is analogous to the Second Law satisfying Green–Kubo relations for linear transport coefficients in fluids of arbitrary density—see section 3. In 1960 Cohen and Berlin [62] showed that if the *molecular chaos* assumption of Boltzmann is assumed to apply to *post*-collisional distributions rather than (as Boltzmann assumed) to *pre*-collisional distributions, then one can derive an *anti-Boltzmann* Equation (our terminology). This use of Boltzmann's molecular chaos assumption for pre- and post-collisional distributions is analogous to our use of Boltzmann's ansatz before and after the strain rate ramps—see section 5. The anti-Boltzmann equation derived by Cohen and Berlin obeys an anti-H theorem [62] which violates the Second Law. Consequently, the anti-Boltzmann equation also leads to negative values for the Navier–Stokes transport coefficients. So in our view macroscopic irreversibility ultimately derives from the time reversible laws of motion and the time asymmetric, Axiom of Causality, (6.4 *a*).

We remark, as an aside, that from a practical point of view an Axiom of Anticausality would be difficult to use in actual calculations since the required information about the future states of a system usually does not exist. In an anticausal Universe, knowledge of its present state would enable us to predict the past but not the future.

One can explore the connection between the Second Law and reversibility directly. The proof of the TFT given in section 2.1 assumed that the dissipative field took the form of a Heaviside step function in time. However, this assumption is unnecessary. The proof also applies to systems with time dependent external fields of definite parity under time reversal mappings:

$$M^{\mathrm{T}}[F_{\mathrm{e}}(t)] = \pm F_{\mathrm{e}}(T - t). \tag{8.1}$$

The initial distribution should be even under time reversal (section 2.1), and the initial ensemble and dynamics must be ergodically consistent (2.5). When these conditions are met, the proof given in section 2.1 is still valid and the Fluctuation Theorem still holds.

We now examine the sub-ensemble averaged transient time dependent responses of the dissipation function for complementary values of the time averaged dissipation $\langle \Omega(t) \rangle_{\bar{\Omega}_t = A}$, $\langle \Omega(t) \rangle_{\bar{\Omega}_t = -A}$. According to our proof of the TFT, one would expect that

$$\langle \Omega(t) \rangle_{\bar{\Omega}_t = A} = M^{\mathrm{T}} \langle \Omega(t) \rangle_{\bar{\Omega}_t = -A} = -\langle \Omega(T - t) \rangle_{\bar{\Omega}_t = -A}. \tag{8.2}$$

This equation is a direct test of our standard proof, in section 2, of the TFT. Given the time reversibility of the equations of motion equation (8.2) must be correct. However, in an actual experiment it is by no means obvious that it can be verified in practice. In a complex many-particle phase space, we expect that there are many non-contiguous, *distinct* phase space trajectory bundles that have the specified time-integrated values of the dissipation function, $\bar{\Omega}_t = \pm A$. Each of these distinct

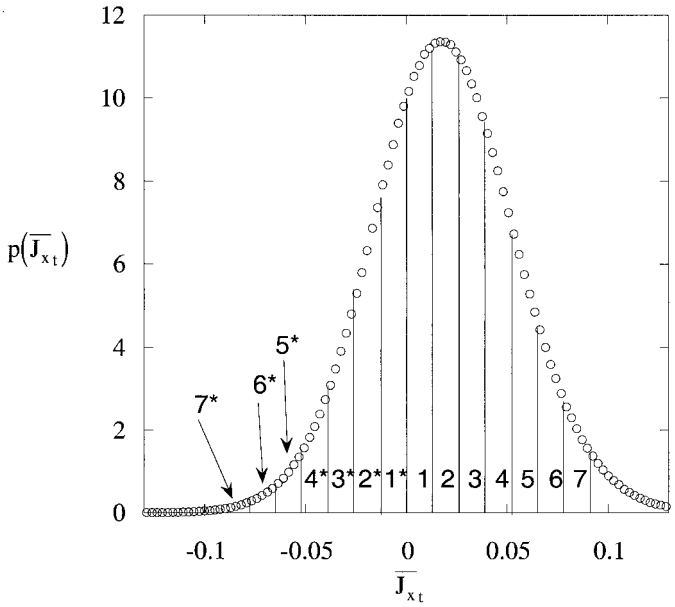


Figure 8.1 Histogram of the distribution of the time-averaged dissipative flux $\overline{J_{x_t}}$.

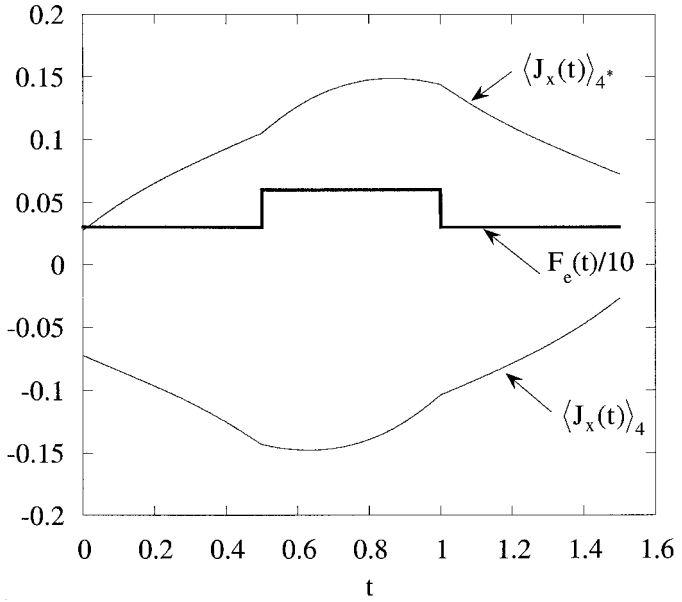


Figure 8.2. The dissipative flux as a function of time for trajectories with conjugate values (bins 4,4* in figure 8.1) of the dissipative flux.

trajectory bundles i, could have very different time dependent dissipation functions, $\Omega_i(t)$. If this is the case it could be very difficult to sample conjugate trajectory bundle pairs, i, i^*, for which $\Omega_i(t) = -\Omega_{i^*}(T - t)$. This would make experimental confirmation of equation (8.2), very difficult.

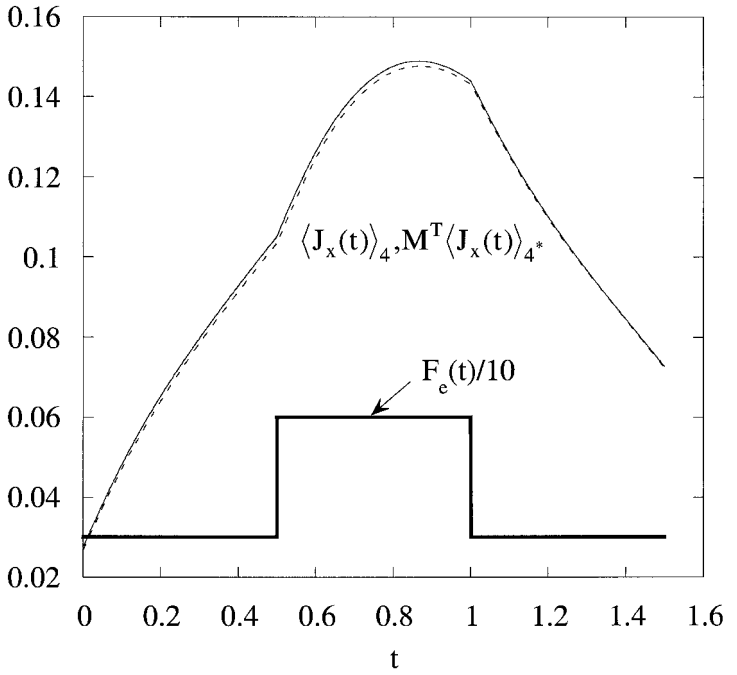


Figure 8.3. The current traces for conjugate histogram bins are related by a time reversal mapping. The full line is $\langle J_x(t)\rangle_4$ and the dotted line is $M^T\langle J_x(t)\rangle_{4^*}$.

We checked equation (8.2) by performing non-equilibrium molecular dynamics simulations [63] of an 8-particle, binary 50 : 50 mixture of coloured WCA particles. The state point studied was $T = 1.0$, $n = 0.4$. The temperature was controlled by a Nosé–Hoover thermostat. In figure 8.1 we show the histogram of the dissipative flux.

The system was subject to a three-step colour field, as shown in figure 8.2. Figure 8.2 also shows the typical response from conjugate histogram bins to the time dependent colour field shown in the figure. The actual bins shown are 4 and 4*, of figure 8.1. However, other conjugate bins show similar results. The subensemble-averaged current traces each show mixed causal and anticausal characteristics. The current for bin 4 clearly begins to decrease before the colour field is decreased at $t = 1$. Overall, comparing bins 4 and its conjugate 4*, we would say that bin 4 exhibits stronger causal characteristics than does bin 4*.

In figure 8.3 we show that within statistical uncertainties the dissipative fluxes for conjugate values of the time-integrated entropy production are, within statistical uncertainties, time-reversed maps of each other—verifying equation (8.2).

The famous problem, the creation of anti-events from events, has no solution. Although, using simple instructions, the [solution] may be put into words: reverse the instantaneous velocities of all of the atoms in the Universe—Loschmidt, 1876 [5]. For an ensemble of experiments we see that we can observe conjugate pairs of time-reversed responses without intervening and reversing particle velocities. All one has to do is to sort the ensemble of responses on the basis of their time-integrated dissipation functions (entropy production in thermostatted systems), and to compare those responses with complementary values of total dissipation. These responses will

be time-reversed mappings of each other. The ratio of probabilities of observing these complementary time-integrated values of dissipation are given by the Fluctuation Theorem, with Second Law satisfying responses being exponentially dominant.

Acknowledgements

We wish to acknowledge the long term support, encouragement and debate from E. G. D. Cohen. During the preparation of this Review, O. Jepps, E. Mittag, E. Sevick and Genmaio Wang provided preprints of unpublished work. We thank E. Mittag who provided translations of Loschmidt papers and L. Rondoni for his useful comments. We also thank the Australian Research Council and the Australian Partnership for Advanced Computing for their support of this work.

References

[1] De Groot, S. R., and Mazur, P., 1984, *Non-equilibrium Thermodynamics* (New York: Dover).
[2] Evans, D. J., and Rondoni, L., 2002, *J. stat. Phys.* (in the press).
[3] Ehrenfest, P., and Ehrenfest, T., 1912, *Enzycl. d Math.* Wiss IV, 2, II Heft 6; English translation by Moravcsik, (1959), *The Conceptual Foundations of the Statistical Approach in Mechanics* (Ithaca, NY: Cornell University Press).
[4] Lebowitz, J. L., 1993, *Phys. Today*, **46**, 32.
[5] Loschmidt, J., 1876, *J. Sitzungsber. der kais. Akad. d. W. math. naturw.*, II, **73,** 128.
[6] Searles, D. J., and Evans, D. J., 1999, *Phys. Rev.* E, **60,** 159.
[7] Lebowitz, J. L., and Spohn, H., 1999, *J. stat. Phys.*, **95,** 333.
[8] Maes, C., 1999, *J. stat. Phys.*, **95,** 367.
[9] Maes, C., Redig, F., and van Moffaert, A., 2000, *J. math. Phys.*, **41,** 1528.
[10] Maes, C., and Redig, F., 2000, *J. stat. Phys.*, **101,** 3.
[11] Kurchan, J., 1998, *J. Phys.* A, **31,** 3719.
[12] Le Chatelier, H. L., 1888, *Ann. Min.*, **13,** 157.
[13] Partington, J. R., 1989, *A Short History of Chemistry* (New York: Dover).
[14] Gallavotti, G., and Cohen, E. G. D., 1995, *Phys. Rev. Lett.*, **74,** 2694.
[15] Gallavotti, G., and Cohen, E. G. D., 1995, *J. stat. Phys.*, **80,** 931.
[16] Evans, D. J., and Morriss, G. P., 1990, *Statistical Mechanics of Nonequilibrium Fluids*, (London: Academic Press).
[17] Ayton, G., Evans, D. J., and Searles, D. J., 2001, *J. chem. Phys.*, **115,** 2035.
[18] Tolman, R. C., 1979, *The Principles of Statistical Mechanics* (New York: Dover).
[19] Evans, D. J., and Morriss, G. P., 1984, *Phys. Rev.* A, **30,** 1528.
[20] Brown, D., and Clarke, J. H. R., 1986, *Phys. Rev.* A, **34,** 2093.
[21] Evans, D. J., Cohen, E. G. D., and Morriss, G. P., 1990, *Phys. Rev.* A, **42,** 5990.
[22] Gaspard, P., and Nicolis, G., 1990, *Phys. Rev. Lett.*, **65,** 1693.
[23] Evans, D. J., Cohen, E. G. D., and Morriss, G. P., 1993, *Phys. Rev. Lett.*, **71,** 2401.
[24] Eckmann, J.-P., and Ruelle, D., 1985, *Rev. mod. Phys.*, **57,** 617.
[25] Benettin, G., Galgani, L., Giorgilli, A., and Strelcyn, J.-M., 1980, *Meccanica*, 9.
[26] Benettin, G., Galgani, L., Giorgilli, A., and Strelcyn, J.-M., 1980, *Meccanica*, 21.
[27] Hoover, W. G., and Posch, H. A., 1985, *Phys. Lett.*, **113A,** 82.
[28] Goldhirsch, I., Sulem, P. L., and Orszag, S. A., 1987, *Physica D*, **27,** 311.
[29] Shimada, I., and Nagashima, T., 1979, *Prog. theor. Phys.*, **61,** 1605.
[30] Geist, K., Parlitz, U., and Lauterborn, W., 1990, *Prog. theor. Phys.*, **83,** 875.
[31] Evans, D. J., Cohen, E. G. D., Searles, D. J., and Bonetto, F., 2000, *J. stat. Phys.*, **101,** 17.
[32] Evans, D. J., and Searles, D. J., 1995, *Phys. Rev.* E, **52,** 5839.
[33] Jarzynski, C., 1997, *Phys. Rev. Lett.*, **78,** 2690.
[34] Crooks, G. E., 1999, *Phys. Rev.* E, **60,** 2721.
[35] Crooks, G. E., 2000, *Phys. Rev.* E, **61,** 2361.

[36] Hatano, T., and Sasa, S., 2001, *Phys. Rev. Lett.*, **86,** 3463.
[37] Hatano, T., preprint, archived in xxx.lanl.gov cond-mat #9905012.
[38] Hummer, G., 2002, *Mol. Sim.*, **28,** 81.
[39] Evans, D. J., and Searles, D. J., 1994, *Phys. Rev.* E, **50,** 1645.
[40] Evans, D. J., and Searles, D. J., 1996, *Phys. Rev.* E, **53,** 5808.
[41] Jepps, O., Evans, D. J., and Searles, D. J. (in preparation).
[42] Searles, D. J., and Evans, D. J., 2000, *J. chem. Phys.*, **113,** 3503.
[43] Evans, D. J., Searles, D. J., and Mittag, E., 2001, *Phys. Rev.* E, **63,** 051105.
[44] Mittag, E., Searles, D. J., and Evans, D. J., 2002, *J. chem. Phys.*, **116,** 6875.
[45] Weeks, J. D., Chandler, D., and Andersen, H. C., 1971, *J. chem. Phys.*, **54,** 5237.
[46] Searles, D. J., Ayton, G., and Evans, D. J., 2000, *AIP Conf. Ser.*, **519,** 271.
[47] Ayton, G., and Evans, D. J., 1999, *J. stat. Phys.*, **87,** 811.
[48] Searles, D. J., and Evans, D. J., 2000, *J. chem. Phys.*, **112,** 9727.
[49] Evans, D. J., and Morriss, G. P., 1985, *Phys. Rev.* A, **31,** 3817.
[50] Hess, S., and Evans, D. J., 2001, *Phys. Rev.* E, **64,** 011207.
[51] Gallavotti, G., 1996, *Phys. Rev. Lett.*, **77,** 4334.
[52] Bonetto, F., Gallavotti, G., and Garrido, P. L., 1997, *Physica* D, **105,** 226.
[53] Bonetto, F., Chernov, N. I., and Lebowitz, J. L., 1998, *Chaos*, **8,** 823.
[54] Pippard, A. B., 1985, *Response and Stability: An Introduction to the Physical Theory* (Cambridge: Cambridge University Press).
[55] Boon, J. P., and Yip, S., 1980, *Molecular Hydrodynamics* (New York: McGraw-Hill).
[56] Biferale, L., Pierotti, D., and Vulpiani, A., 1998, *J. Phys.* A, **31,** 21.
[57] Bonetto, F., and Lebowitz, J. L., 2001, *Phys. Rev.* E, **64,** 056129.
[58] Sasa, S., preprint, archived in xxx.lanl.gov nlin #0010026.
[59] Lepri, S., Livi, R., and Politi, A., 1997, *Phys. Rev. Lett.*, **78,** 1896.
[60] Ciliberto, S., and Laroche, 1998, C., *J. Phys. IV Fr.*, **8,** 215.
[61] Wang, G. M., Sevick, E., Mittag, E., Searles, D. J., and Evans, D. J., 2002, *Phys. Rev. Lett.*, **89,** 050601.
[62] Cohen, E. G. D., and Berlin, T. H., 1960, *Physica*, **26,** 717.
[63] Mittag, E., and Evans, D. J., 2002 (in preparation).