

## Chapter 3

### The Evans-Searles Fluctuation Theorem

Hence the Second Law of thermodynamics is continually being violated and that to a considerable extent in any sufficiently small group of molecules belonging to any real body. *As the number of molecules in the group is increased*, the deviations from the mean of the whole become smaller and less frequent; and when the number is increased till the group includes a sensible portion of the body, the probability of a measurable variation from the mean occurring *in a finite number of years* becomes so small that it may be regarded as practically an impossibility.

J.C. Maxwell, Nature, **17**, 278 (1878) (our italics)

#### 3.1 The Transient Fluctuation Theorem

The first proof (1994) of any fluctuation theorem was for a special case of what is now known as the Evans-Searles Transient Fluctuation Theorem (ESFT). Here we give a very general proof. Consider the response of a system, initially in some known but arbitrary distribution,

$$f(\mathbf{\Gamma}, 0) = \frac{\exp[-F(\mathbf{\Gamma})]}{\int_D d\mathbf{\Gamma} \exp[-F(\mathbf{\Gamma})]}, \quad (3.1.1)$$

where  $F(\mathbf{\Gamma})$  is some arbitrary single valued real function for which

$f(\mathbf{\Gamma}, 0) = f(M^T \mathbf{\Gamma}, 0)$  (*i.e.* the initial distribution is an even function of the momenta), defined over some specified phase space domain  $D$ .  $\mathbf{\Gamma}$  is the extended phase space vector which includes the phase space vector and may include additional dynamical variables such as the volume or the thermostat multiplier associated with a possible Nosé-Hoover thermostat.

Consider any system whose dynamics is described by continuous, deterministic, time-reversible equations of motion. The equations of motion may have an applied dissipative field or the field may be zero. If the field is zero then in

order to see anything interesting, the initial distribution should not be preserved by the equations of motion (if it is preserved, then the ESFT is completely trivial). On the other hand if a dissipative field is applied then it is frequently useful to consider the case where the initial distribution is the equilibrium distribution for the field free dynamics.

We assume the unthermostatted equations of motion satisfy the AIT condition. A thermostat may be added (as in (2.2.5), for example), but again this is not absolutely essential. The equations of motion *must* however be time reversal symmetric.

**Definition:**

The *time averaged dissipation*  $\bar{\Omega}_t(\Gamma)$  along a trajectory originating at phase  $\Gamma$  and averaged for a time  $t$ , is defined as [12,13]:

$$\begin{aligned} \int_0^t ds \Omega(S^s \Gamma) &\equiv \ln \left( \frac{f(\Gamma; 0)}{f(M^T S^t \Gamma; 0)} \right) - \int_0^t \Lambda(S^s \Gamma) ds \\ &\equiv \bar{\Omega}_t(\Gamma) t \equiv \Omega_t(\Gamma) \end{aligned} \quad (3.1.2)$$

It is useful to define  $\Gamma^* \equiv M^T S^t \Gamma$ . From (2.1.9) we know that this phase space vector is the origin of the conjugate anti trajectory to that trajectory starting at  $\Gamma$ . Going forward in time with the natural propagator from  $\Gamma^*$  is like going backwards in time from  $S^t \Gamma$  except that the velocities are reversed – see (2.1.9).

**Definition:**

A system is said to be *ergodically consistent* over a phase space domain  $D$ , if

$$\begin{aligned} \forall \Gamma \in D, \quad st \quad f(\Gamma; 0) \neq 0, \\ M^T S^t \Gamma \in D, \quad and \quad f(M^T S^t \Gamma; 0) \neq 0, \forall t \end{aligned} \quad (3.1.3)$$

In order for the dissipation function to be well defined over the phase space domain  $D$  the system must be *ergodically consistent* over  $D$ . There are systems that fail to

satisfy this condition. For example, if we let the initial distribution be microcanonical and if the dynamics does not preserve the energy (there may be a dissipative field but no ergostat etc), then ergodic consistency obviously breaks down.

Ergodic consistency also implies that for almost all trajectories that start at a phase vector  $\mathbf{\Gamma}$  inside the domain  $D$ , the conjugate antitrajectory which starts at  $M^T S' \mathbf{\Gamma}$  is also inside  $D$ . We say “almost all” because if there is a zero measure set of trajectories that have missing antitrajectories, this will not violate (3.1.3). Ergodic consistency is concerned with phase space density not with zero measure objects (*e.g.* individual phase space trajectories). As mentioned in Chapter 1, almost all Loschmidt’s antitrajectories exist in the initial distribution of states.

Physicists usually down play the importance of specifying the phase space domain but this specification can be very important. If  $N$ -particles are physically constrained to be located in a physical region (by impenetrable walls or so) then the specification of the domain can be very important.

We can re-write the definition of the dissipation function so that it directly gives the ratio of the probabilities,  $p$ , at time zero, of observing sets of phase space trajectories originating inside infinitesimal volumes of phase space  $\delta V_{\mathbf{\Gamma}}$  and

$$\delta V_{\mathbf{\Gamma}}(\mathbf{\Gamma}^*) \equiv \delta V_{\mathbf{\Gamma}}(M^T S' \mathbf{\Gamma})$$

$$\begin{aligned} \frac{p(\delta V_{\mathbf{\Gamma}}(\mathbf{\Gamma}, 0))}{p(\delta V_{\mathbf{\Gamma}}(\mathbf{\Gamma}^*, 0))} &= \frac{f(\mathbf{\Gamma}; 0) \delta V_{\mathbf{\Gamma}}(\mathbf{\Gamma})}{f(\mathbf{\Gamma}^*; 0) \delta V_{\mathbf{\Gamma}}(\mathbf{\Gamma}^*)} \\ &\xrightarrow{\delta V_{\mathbf{\Gamma}} \rightarrow 0} \frac{f(\mathbf{\Gamma}; 0)}{f(\mathbf{\Gamma}^*; 0)} \exp \left[ - \int_0^t ds \Lambda(S^s \mathbf{\Gamma}) \right] \\ &= \exp \left[ \bar{\Omega}_t(\mathbf{\Gamma}) t \right] \end{aligned} \quad (3.1.4)$$

We have also used (2.4.11) for  $\delta V_{\mathbf{\Gamma}}(\mathbf{\Gamma})/\delta V_{\mathbf{\Gamma}}(S' \mathbf{\Gamma})$  together with the observation that the Jacobian for the time reversal map is unity  $\delta V_{\mathbf{\Gamma}}(M^T \mathbf{\Gamma}^*)/\delta V_{\mathbf{\Gamma}}(\mathbf{\Gamma}^*) = 1$ . The third line follows by a trivial use of the definition of (3.1.2).

In all the work in this book we will assume that the initial distribution is invariant under the time reversal mapping,  $M^T$  and  $f(M^T \mathbf{\Gamma}; 0) = f(\mathbf{\Gamma}; 0)$ .

Thus the dissipation function gives the logarithm of the probability ratio of observing at time zero, an infinitesimal set of trajectories relative to the conjugate set of anti-trajectories. Thus one way to think of the dissipation function is as a measure of the temporal asymmetry inherent in sets of trajectories originating from an initial distribution of states. As we will see the dissipation function has an eclectic set of properties.

What is not so obvious is that this definition of the time averaged dissipation function even applies to some non-autonomous systems. If a time dependent external field has a definite parity under time reversal the conjugate sets of trajectories and anti-trajectories still exist and the time averaged dissipation can still be calculated using (3.1.4).

We have not said anything about how we could choose,  $\delta V_{\Gamma}$ . Now suppose we choose the volume element  $\delta V_{\Gamma}$  to be that set of volume elements in  $D$ , within which all trajectories originating at time zero from within that volume have the time integrated dissipation function,  $\bar{\Omega}_t(\Gamma) = (A \pm \delta A)$ . Then we have

$$\begin{aligned} \lim_{\delta A \rightarrow 0} \frac{p(\delta V_{\Gamma}(\Gamma, 0))}{p(\delta V_{\Gamma}(\Gamma^*, 0))} &= \lim_{\delta A \rightarrow 0} \frac{f(\Gamma; 0) \delta V_{\Gamma}(\Gamma)}{f(\Gamma^*; 0) \delta V_{\Gamma}(\Gamma^*)} \\ &= \exp[\bar{\Omega}_t(\Gamma)t] \\ &= \exp[At] \end{aligned} \tag{3.1.5}$$

Using the time reversal symmetry of the equations of motion, all trajectories originating within  $\delta V_{\Gamma^*}$  must have the property that  $\bar{\Omega}_t(\Gamma^*) = -(A \pm \delta A)t$  and therefore we see that:

$$\frac{p(\bar{\Omega}_t = A)}{p(\bar{\Omega}_t = -A)} = \exp[At], \tag{3.1.6}$$

This is the Evans-Searles Transient Fluctuation Theorem (ESFT).

It is clearly asymmetric in time. The integrated dissipation function itself is odd under time reversal. In (3.1.6) if  $A$  is positive then (3.1.6) says it is exponentially

more likely to observe positive rather than negative dissipation. If on the other hand  $A$  is negative, then it is exponentially more *unlikely* to observe negative rather than positive dissipation. Regardless of the sign of  $A$ , the implication is the same. Positive dissipation is more likely than its complementary negative counterpart.

What is not so obvious is that for a given system there may be multiple *noncontiguous* phase space subvolumes which each have a time averaged dissipation equal to  $A \pm dA$ . However because the system is ergodically consistent every such subvolume has its own conjugate phase space subvolume that contains the phase space vectors for the time reversed conjugate antitrajectories. Every such subvolume has a time averaged dissipation of  $-A \pm dA$ .

We need to stress again the conditions required for the validity of (3.1.6):

- the initial distribution should be an even function of the momenta
- we assume the system is ergodically consistent over the relevant domain,
- the dynamics must of course be time reversal symmetric, and
- the dynamics should be smooth.
- any time dependent external fields must have a definite parity under time reversal.

Since the time integrated dissipation function itself is extensive in the integration time and in the number of degrees of freedom we see that for macroscopic systems observed for macroscopic times the probability of observing negative dissipation “becomes so small that it may be regarded as practically an impossibility” – Maxwell, 1878. It is interesting to note that Maxwell, recognized the importance of *both* time and system size in relation to observing violations of the Second “Law”. The quote reveals that Maxwell would not be surprised at the qualitative implications of the ESFT. However the ESFT gives a precise quantification of the matter.

It should be noted that the ESFT gives a relation between probabilities of time integrals of the dissipation function. These time integrals start at the time when the dissipation function is defined (3.1.2). The dissipation function is a functional of *both* the dynamical equations of motion that determine  $S'\Gamma = \exp[-iL(\Gamma)t]\Gamma$  from the initial phase  $\Gamma$  and also the initial distribution  $f(\Gamma;0)$ . These “initial” times need to be one and the same time.

The instantaneous dissipation function can be determined by differentiation of (3.1.2) as

$$\begin{aligned}
\frac{\partial}{\partial t} \int_0^t ds \Omega(S^s \Gamma) &= \Omega(S^t \Gamma) \\
&= \frac{\partial}{\partial t} [\ln(f(\Gamma; 0) - \ln(f(e^{iL(\Gamma)t} \Gamma; 0) - \int_0^t ds \Lambda(e^{iL(\Gamma)s} \Gamma))] \\
&= -\frac{1}{f(e^{iL(\Gamma)t} \Gamma, 0)} \frac{\partial f(e^{iL(\Gamma)t} \Gamma; 0)}{\partial t} - \Lambda(e^{iL(\Gamma)t} \Gamma) \\
&= -\frac{1}{f(e^{iL(\Gamma)t} \Gamma, 0)} \frac{\partial e^{iL(\Gamma)t} \Gamma}{\partial t} \cdot \frac{\partial f(e^{iL(\Gamma)t} \Gamma; 0)}{\partial \Gamma} - \Lambda(e^{iL(\Gamma)t} \Gamma) \quad (3.1.7) \\
&= -\frac{1}{f(e^{iL(\Gamma)t} \Gamma, 0)} iL(\Gamma) e^{iL(\Gamma)t} \Gamma \cdot \frac{\partial f(e^{iL(\Gamma)t} \Gamma; 0)}{\partial \Gamma} - \Lambda(e^{iL(\Gamma)t} \Gamma) \\
&= -\frac{1}{f(e^{iL(\Gamma)t} \Gamma, 0)} e^{iL(\Gamma)t} \dot{\Gamma}(\Gamma) \cdot \frac{\partial f(e^{iL(\Gamma)t} \Gamma; 0)}{\partial \Gamma} - \Lambda(e^{iL(\Gamma)t} \Gamma) \\
&= -\frac{1}{f(S^t \Gamma, 0)} \dot{\Gamma}(S^t) \cdot \frac{\partial f(S^t \Gamma; 0)}{\partial \Gamma} - \Lambda(S^t \Gamma)
\end{aligned}$$

The derivative on the left hand side of (3.1.7) is to be computed at a fixed point in phase space. If we now set  $t = 0$  we obtain an expression for the instantaneous dissipation function:

$$\Omega(\Gamma) = -\frac{1}{f(\Gamma, 0)} \dot{\Gamma}(\Gamma) \cdot \frac{\partial f(\Gamma; 0)}{\partial \Gamma} - \Lambda(\Gamma). \quad (3.1.8)$$

The ESFT has generated much interest, as it shows how irreversibility emerges from the deterministic, reversible equations of motion [16]. Its proof is extremely simple and uses almost nothing but the time reversibility of the underlying

dynamics. Because its proof relies on so few assumptions, the ESFT is extremely general. It is valid arbitrarily far from equilibrium. It provides a generalized form of the 2<sup>nd</sup> “Law” of Thermodynamics that can be applied to small systems observed for short periods of time. It also resolves the longstanding Loschmidt Paradox. The ESFT has been verified experimentally [17-24]. (See Section 9)

The form of the ESFT (3.1.6) applies to any valid ensemble/dynamics combination. However the precise expression for  $\bar{\Omega}_t$  given in (3.1.2) is dependent on both the initial distribution *and* the dynamics.

### 3.2 Second Law Inequality

We are now in a position to use the ESFT to derive a number of simple inequalities. The derivation of the Second Law Inequality (SLI) from the ESFT provides what amounts to a proof of the Second “Law” of Thermodynamics. The SLI shows that *time averages* (rather than instantaneous values) of the ensemble averaged dissipation are nonnegative. This Second Law Inequality is valid for the appropriately time averaged dissipation but the ensemble averaged instantaneous dissipation may be negative for intermediate of times.

The Second Law Inequality states that[72]:

$$\langle \Omega_t \rangle \geq 0, \forall t > 0 \quad (3.2.1)$$

The proof is almost trivial and is obtained by integration of (3.1.6):

$$\begin{aligned} \langle \Omega_t \rangle &= \int_{-\infty}^{+\infty} dB \, p(\Omega_t = B) B \\ &= \int_0^{+\infty} dB \, p(\Omega_t = B) B + \int_{-\infty}^0 dB \, p(\Omega_t = B) B \\ &= \int_0^{+\infty} dB \, p(\Omega_t = B) B - \int_0^{+\infty} dB \, p(\Omega_t = -B) B \\ &= \int_0^{+\infty} dB \, p(\Omega_t = B) B (1 - \exp[-B]) \geq 0. \end{aligned} \quad (3.2.2)$$

In linear irreversible thermodynamics it is asserted that the quantity called the spontaneous entropy production cannot be negative. Close to equilibrium the ensemble averaged dissipation is equal to the ensemble averaged entropy production (see §x.x). In an electric circuit close to equilibrium both quantities are equal to the product of the electric current times the voltage divided by the ambient temperature. If the circuit has a complex impedance there will necessarily be a phase lag between the applied voltage and the current. This means that for an AC sinusoidal electric circuit, there will always be intervals within a cycle, within which the entropy production is negative. This presents serious difficulties for linear irreversible thermodynamics, but



the Second Law Inequality is not presented with any difficulties by this matter. The Second Law Inequality only asserts that the time integrated, average dissipation is positive. The time integral begins at the initial time when the dissipation function itself was defined. The Second Law Inequality does *not* state that the ensemble averaged instantaneous dissipation cannot be negative.

Now let us look at (3.2.2) in more detail. For *every* value of  $B > 0$  if  $p(\bar{\Omega}_t = B) > 0$ , ergodic consistency implies  $p(\bar{\Omega}_t = -B) > 0$ . This is because for every set of trajectories, the conjugate set of antitrajectories exists within the ostensible phase space domain. Furthermore as can be seen from (3.2.2), the fluctuation theorem shows  $p(\bar{\Omega}_t = B) > p(\bar{\Omega}_t = -B)$ . This in turn means that if the time integrated dissipation is nonzero for some infinitesimal set of initial points in phase space near  $\Gamma$ :

$$\bar{\Omega}_t(\Gamma) \neq 0, \Gamma \in D, t > 0 \Rightarrow \langle \bar{\Omega}_t \rangle > 1. \quad (3.2.3)$$

### Definition

A nonzero value for the time-averaged dissipation of an infinitesimal set of phase space trajectories anywhere in the ostensible phase space domain, implies a *Strict Second Law Inequality* (3.2.3).

We note that the Second Law Inequality (both strict and otherwise) has *macroscopic* consequences for the Fluctuation Theorem. The Second Law Inequality has important consequences in widely varied applications such as atmospheric physics and aerodynamics.

### 3.3 Nonequilibrium Partition Identity

This Identity (also referred to as the Kawasaki identity, Kawasaki normalisation factor, Kawasaki function and the integral fluctuation theorem) was first implied for Hamiltonian systems by Yamada and Kawasaki in 1967, and for thermostatted systems driven by an external field, was explicitly noted by Morriss and Evans in 1984 [11,74,75]. The Nonequilibrium Partition Identity (NPI) is stated as:

$$\langle \exp[-\Omega_t] \rangle = 1 \quad (3.3.1)$$

A very simple proof can be obtained using the ESFT given in eq. (3.1.6):

$$\begin{aligned} \langle \exp[-\Omega_t] \rangle &= \int_{-\infty}^{+\infty} dB \, p(\Omega_t = B) \exp[-B] \\ &= \int_{-\infty}^{+\infty} dB \, p(\Omega_t = -B) \\ &= \int_{-\infty}^{+\infty} dB' \, p(\Omega_t = B') = 1 \end{aligned} \quad (3.3.2)$$

It is quite extraordinary that although the Second Law Inequality says the exponent of the NPI is negative on average, the rare instances when the dissipation function has a negative time average occur with such frequency that their exponentially enhanced effect ensures the average of the exponential is always unity. Trivially we observe that the NPI is still valid even in the case where  $\Omega_t(\Gamma) = 0, \forall \Gamma \in D$ .

We note that in order to *observe* the NPI in real experimental data we must be able to observe the antitrajectories that are conjugate to the most probable trajectories. In macroscopic systems this may be (as Maxwell already noted) impossible due to the extremely low probability of observing these events.

Although the ESFT implies the NPI, the converse is not true [75].

### 3.4 Integrated Fluctuation Theorem

The Fluctuation Theorem quantifies the probability of observing time-averaged dissipation functions having 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 that predicts the probability ratio that the dissipation function is either positive or negative.

In experimental situations where the statistical error is large and the ensemble sample sizes are small, it is useful to be able to predict the probability that the time averaged dissipation is negative. 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 ESFT, (3.1.6), can be written as

$$\frac{p(\bar{\Omega}_t = -A)}{p(\bar{\Omega}_t = A)} = \exp(-At). \quad (3.4.1)$$

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

$$p_+(t) \equiv p(\bar{\Omega}_t > 0), \quad p_-(t) \equiv p(\bar{\Omega}_t < 0). \quad (3.4.2)$$

Now

$$\frac{p_-(t)}{p_+(t)} = \frac{\int_0^\infty dA \, p(\bar{\Omega}_t = -A)}{\int_0^\infty dA \, p(\bar{\Omega}_t = A)}. \quad (3.4.3)$$

Using (3.1.6):

$$\frac{p_-(t)}{p_+(t)} = \frac{\int_0^\infty dA \exp(-At) p(\bar{\Omega}_t = A)}{\int_0^\infty dA \, p(\bar{\Omega}_t = A)}. \quad (3.4.4)$$

The right hand side of this equation is just the ensemble average of  $\exp(-\bar{\Omega}_t t)$  evaluated over that subset of trajectories for which the time-averaged dissipation is positive.

Again if we look at (3.4.3,4) in detail, we see that on the right hand side for *every* value of  $A > 0$   $p(\bar{\Omega}_t = -A) < p(\bar{\Omega}_t = A)$ . This in turn means that if the time integrated dissipation is non zero for *any* value of  $A$ :

$$p(\bar{\Omega}_t(\Gamma) = A) \neq 0, \Rightarrow \frac{p_-(t)}{p_+(t)} = \left\langle \exp(-\bar{\Omega}_t t) \right\rangle_{\bar{\Omega}_t > 0} < 1. \quad (3.4.5)$$

From (3.4.4) we can also obtain the reciprocal relationship:

$$\frac{p_+(t)}{p_-(t)} = \frac{1}{\left\langle \exp(-\bar{\Omega}_t t) \right\rangle_{\bar{\Omega}_t > 0}} \geq 1. \quad (3.4.6)$$

where the equality only holds if  $p(\bar{\Omega}_t = A) = 0, \forall A$ .

Similarly, it can be shown that

$$\frac{p_+(t)}{p_-(t)} = \left\langle \exp(-\bar{\Omega}_t t) \right\rangle_{\bar{\Omega}_t < 0} \geq 1, \quad (3.4.7)$$

where the equality only holds if  $p(\bar{\Omega}_t = A) = 0, \forall A$ .

We note that in actual experiments, where  $\langle \bar{\Omega}_t \rangle > 0$ , equations (3.4.5) and (3.4.6) have much smaller statistical uncertainties than (3.4.7), because rarely observed trajectory segments with highly negative values of  $\bar{\Omega}_t$  will have a large influence on the ensemble average in (3.4.7). Consequently (3.4.7) should be avoided in numerical calculations or experiments.

Finally, we note that equation (3.4.6) can be used to show that

$$p_-(t) = \frac{\langle \exp(-\bar{\Omega}_t t) \rangle_{\bar{\Omega}_t > 0}}{\left(1 + \langle \exp(-\bar{\Omega}_t t) \rangle_{\bar{\Omega}_t > 0}\right)}, \quad p_+(t) = \frac{1}{\left(1 + \langle \exp(-\bar{\Omega}_t t) \rangle_{\bar{\Omega}_t > 0}\right)}. \quad (3.4.8)$$

Obviously  $p_-(t) + p_+(t) = 1, \forall t$  and again  $p_-(t) \leq p_+(t), \forall t$ . Furthermore one can only have equality if there is zero dissipation. Nonzero dissipation anywhere in the relevant phase space implies  $p_-(t) < p_+(t), \forall t$  !

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

### 3.5

### Functional Transient Fluctuation Theorem – leave for Debra

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 (3.1.2). In this section we describe how the FT can be extended to apply to arbitrary phase functions that have an odd parity under time reversal [46].

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

$$\bar{\phi}_{i,t} = \frac{1}{t} \int_0^t ds \phi(S^s \Gamma_i), \quad (3.5.1)$$

for a phase space trajectory:  $S^s \Gamma_i$ . 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

$\delta V_\Gamma(\Gamma, 0)$  and at time  $t$  these phase points will occupy a volume

$\delta V_\Gamma(S^t \Gamma, t) = \delta V_\Gamma(\Gamma, 0) e^{\bar{\Lambda}_t}$  where  $\bar{\Lambda}_t$  is the time-averaged phase space compression

factor along these trajectories – see (2.4.11). We denote  $\bar{\phi}(t) = \langle \bar{\phi}_{i,t} \rangle_{\{\Gamma_i\}}$ , 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

$\{\Gamma_i^*\}$ , given by  $\Gamma_i^* = M^T(S^t \Gamma_i)$ , that will occupy a volume  $\delta V_\Gamma(\Gamma^*) = \delta V_\Gamma(S^t \Gamma) =$

$\delta V_\Gamma(\Gamma) e^{\bar{\Lambda}_t}$  along which the time-averaged value of the phase function is

$\bar{\phi}_{i^*,t} = M^T(\bar{\phi}_{i,t})$ . 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 measure of the volume elements by equation (2.1.5).

Therefore, from the definition of the dissipation function in (2.1.7) we obtain,

$$\lim_{\delta V_\Gamma \rightarrow 0} \frac{p(\delta V_\Gamma(\Gamma, 0))}{p(\delta V_\Gamma(\Gamma^*, 0))} = \exp[\bar{\Omega}_t(\Gamma)t]. \quad (3.5.2)$$

It is possible that there are non-contiguous bundles of trajectories for which

$\{\Gamma | A < \bar{\phi}_t(S^s \Gamma < A + \delta A)\}$ , and since these bundles may have different values  $\bar{\Omega}_t$ , the probability ratio (equation (3.5.2)) 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,

$$\begin{aligned}
 \frac{p(\bar{\phi}_t = A)}{p(\bar{\phi}_t = -A)} &= \frac{\int_{\bar{\phi}_t(\Gamma)=A} d\Gamma f(\Gamma; 0)}{\int_{\bar{\phi}_t(\Gamma^*)=-A} d\Gamma^* f(\Gamma^*; 0)} \\
 &= \frac{\int_{\bar{\phi}_t(\Gamma)=A} d\Gamma f(\Gamma; 0)}{\int_{\bar{\phi}_t(\Gamma)=A} d\Gamma f(\Gamma; 0) e^{-\bar{\Omega}_t(\Gamma)t}} \\
 &= \left\langle e^{-\bar{\Omega}_t t} \right\rangle_{\bar{\phi}_t=A}^{-1}
 \end{aligned} \tag{3.5.3}$$

where the notation  $\langle \dots \rangle_{\bar{\phi}_t=A}$  refers to the ensemble average over (possibly) non-contiguous trajectory bundles for which  $\bar{\phi}_t = A$ . Equation (3.5.3) 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 Functional Transient Fluctuation Theorem (FTFT) 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, then we obtain the trivial relationship

$$\left\langle e^{-\bar{\Omega}_t t} \right\rangle_{\bar{\phi}_t=A}^{-1} = \frac{p(\bar{\phi}_t = A)}{p(\bar{\phi}_t = A)} = 1 \tag{3.5.4}$$

### 3.6 The Covariant Dissipation Function

As we have seen already the dissipation function is a rather important function in statistical mechanics. In later chapters we will see that it plays a key role in almost all aspects of nonequilibrium statistical mechanics – in response theory and in understanding the process of relaxation towards equilibrium. It is defined in terms of the initial distribution of states and also by the dynamical equations of motion.

What happens to the dissipation function if we redefine the dissipation function in terms of the time evolving N-particle phase space distribution function rather than the initial distribution? The time covariant dissipation function could be written as,

$$\Omega_\tau(S^{t_1}\Gamma, t_1) \equiv \ln\left(\frac{f(S^{t_1}\Gamma, t_1)}{f(M^T S^{t_1+\tau}\Gamma, t_1)}\right) - \int_{t_1}^{t_1+\tau} \Lambda(S^s\Gamma) ds \quad (3.6.1)$$

where dissipation function is integrated for a time  $\tau$  but defined with respect to the phase space density at time  $t_1$  rather than at the usual time zero. By constructing the Evans-Searles FT at time  $t_1$  and allowing this time to increase without bound, we could construct an exact steady state FT for thermostatted driven systems that evolve to nonequilibrium steady states. This steady state FT would not be asymptotic unlike the FT discussed in §6.9??.

However, there is a serious problem posed by this scenario. The time integrated dissipation function is related to a number of important physical properties. We have only met a small number of these properties thus far in this book. This means that there must be some kind of invariance properties satisfied by the dissipation function. So you really can't constantly redefine the quantity.

From the definition (3.6.1) we see that,

$$\begin{aligned} \lim_{\delta V_\Gamma \rightarrow 0} \frac{p[\delta V_\Gamma(S^{t_1}\Gamma, t_1)]}{p[\delta V_\Gamma(M^T S^{t_1+\tau}\Gamma, t_1)]} &= \lim_{\delta V_\Gamma \rightarrow 0} \frac{f(S^{t_1}\Gamma, t_1)\delta V_\Gamma(S^{t_1}\Gamma)}{f(M^T S^{t_1+\tau}\Gamma, t_1)\delta V_\Gamma(M^T S^{t_1+\tau}\Gamma)} \\ &= \exp[\Omega_\tau(S^{t_1}\Gamma, t_1)] \end{aligned} \quad (3.6.2)$$



Now all the trajectories that arrive in  $\delta V_{\mathbf{r}}(S^{t_1}\mathbf{\Gamma})$  at time  $t_1$  started out within  $\delta V_{\mathbf{r}}(\mathbf{\Gamma})$  at time zero. All the trajectories that arrive at  $\delta V_{\mathbf{r}}(S^{t_1+\tau}\mathbf{\Gamma})$  at time  $t_1 + \tau$  would have continued on to  $\delta V_{\mathbf{r}}(S^{2t_1+\tau}\mathbf{\Gamma})$  at time  $2t_1 + \tau$ . Furthermore all trajectories in  $\delta V_{\mathbf{r}}(S^{2t_1+\tau}\mathbf{\Gamma})$  at time  $2t_1 + \tau$  started within  $\delta V_{\mathbf{r}}(\mathbf{\Gamma})$  at time zero. So in fact,

$$\lim_{\delta V_{\mathbf{r}} \rightarrow 0} \frac{p[\delta V_{\mathbf{r}}(S^{t_1}\mathbf{\Gamma}), t_1]}{p[\delta V_{\mathbf{r}}(M^T S^{t_1+\tau}\mathbf{\Gamma}), t_1]} = \lim_{\delta V_{\mathbf{r}} \rightarrow 0} \frac{p[\delta V_{\mathbf{r}}(\mathbf{\Gamma}), 0]}{p[\delta V_{\mathbf{r}}(M^T S^{2t_1+\tau}\mathbf{\Gamma}), 0]} . \quad (3.6.3)$$

For the antitrajectories, going backwards in time from  $t_1$  to zero is like going forward in time an additional amount  $t_1$  from time  $t_1 + \tau$ , and therefore,

$$\Omega_{\tau}(S^{t_1}\mathbf{\Gamma}, t_1) = \Omega_{2t_1+\tau}(\mathbf{\Gamma}, 0) \quad (3.6.4)$$

The antitrajectories at time zero to those within  $\delta V_{\mathbf{r}}(\mathbf{\Gamma})$  are the time reversal mapped phases to those  $\delta V_{\mathbf{r}}(S^{2t_1+\tau}\mathbf{\Gamma})$ .

So there is no new information contained within the time covariant dissipation function. It does imply an important result however. There is *no* time-local non-asymptotic ESFT for steady states with time reversible deterministic dynamics.

### 3.7 The Definition of Equilibrium Distributions

“If a system is very weakly coupled to a heat bath at a given ‘temperature’, if the coupling is indefinite or not known precisely, if the coupling has been on for a long time, and if all the ‘fast’ things have happened and all the ‘slow’ things not, the system is said to be in *thermal equilibrium*.”

R. P. Feynman, Statistical Mechanics, a series of lectures, 1972.

One of the aims of this book is to understand the true nature of thermal equilibrium. We will return to discuss the nature of equilibrium many times in this book; each time with a little more knowledge than before, until at the end of Chapter 5 we will be able to demonstrate that the definition we are about to introduce, indeed encompasses each aspect of the qualitative notion of equilibrium given by Feynman in the quote above.

It may seem somewhat odd that we should introduce a definition of equilibrium in a chapter that is mostly devoted to discussing nonequilibrium systems. However you cannot really understand equilibrium without first knowing how nonequilibrium systems relax towards equilibrium.

#### **Definition:**

An *equilibrium system* is characterized by a  $N$ -particle phase space distribution and a dynamics for which, over the phase space domain  $D$ , the time integrated dissipation function is identically zero:

$$\begin{aligned}\bar{\Omega}_{eq,t}(\Gamma) &= 0, \forall \Gamma \in D, \forall t > 0, \\ \Rightarrow \left\langle \left( \bar{\Omega}_{eq,t} \right) \right\rangle &= 0, \forall t > 0 \\ \Rightarrow p_{eq,+}(t) &= p_{eq,-}(t), \forall t > 0\end{aligned}\tag{3.7.1}$$

Although this is a convenient definition of equilibrium we do not yet know whether equilibrium systems exist or whether such systems are stable. It turns out that the answer to both these questions is yes but these answers will only be given in the next chapter.

We have already seen that the only way  $\langle \Omega_t \rangle = 0$  is if the instantaneous dissipation and the time averaged dissipation are both zero everywhere (3.7.1), Consequently the ensemble averaged time integrated dissipation is zero if and only if the time integrated dissipation is zero everywhere in the ostensible phase space:

$$\langle \Omega_t \rangle_{eq} = 0 \Leftrightarrow \Omega_t(\Gamma) = 0, \forall \Gamma \in D, \forall t > 0 . \quad (3.7.2)$$

### Definition

This equation is called the *Second Law Equality*.

A number of corollaries follow immediately. From (3.7.1) we observe that for equilibrium systems that are ergodically consistent over  $D$ , the probability of observing *every* infinitesimal set of phase space trajectories is equal to the probability of observing, at time zero, the conjugate set of anti-trajectories:

$$\frac{p_{eq}(\delta V_\Gamma(\Gamma, 0))}{p_{eq}(\delta V_\Gamma(M^T S' \Gamma, 0))} = 1, \forall \Gamma \in D, \forall t \quad (3.7.3)$$

The equilibrium state is therefore time reversal symmetric.

For instance, if we compute the Lyapunov spectrum  $\{\lambda_i; i = 1, \dots, d; \Gamma(0)\}$  for a system with time reversible dynamics, for a trajectory originating at  $\Gamma(0)$ , we know that for any *steady* system (nonequilibrium steady state or an equilibrium state) the spectra have the property that if we reverse the direction of time, the largest most positive exponent will be -1 times the smallest most negative exponent of the original system. If we denote the exponents of the time reversed system as,  $\{\lambda_i^*; i = 1, \dots, d; \Gamma^*(0)\}$ , we will have,

$$\lambda_i^*(\Gamma^*(0)) = -\lambda_{d-i}(\Gamma(0)), \forall \Gamma(0) \in D . \quad (3.7.4)$$

(Note:  $d$  is the number of nonzero Lyapunov exponents in the system.)

Now, if we further assume that the system is an ergodic equilibrium system we see that the spectrum must be independent of the initial phase  $\Gamma(0)$  or  $\Gamma^*(0)$ . This means that the spectrum for the trajectory must be the same as the spectrum of the anti-trajectories. At equilibrium therefore the time reversal map transforms the spectrum into itself. This means that at equilibrium

$$\lambda_{eq,i}(\Gamma(0)) = -\lambda_{eq,d-i}(\Gamma(0)), \forall \Gamma(0), i \quad (3.7.5)$$

This is termed the Conjugate Pairing Rule for equilibrium systems.

All ergodic equilibrium systems have Lyapunov spectra that, apart from any unpaired zero exponents, consist of conjugate pairs of exponents that each sum to zero. The conjugate paired exponents define sets of 2-dimensional areas that are each preserved in measure, by the natural dynamics. The Kaplan-Yorke dimension of an ergodic equilibrium system is equal to the number of Lyapunov exponents (including unpaired zero exponents) which is also the ostensible dimension of phase space. For these systems the ostensible phase space volume is at least on average, preserved by the natural dynamics.

Thus far we have only discussed equilibrium systems in the context of time integrated dissipation. Later in Chapter 4 we will talk about equilibrium in the context of instantaneous dissipation. At the moment we do not know whether if  $\Omega_t(\Gamma) = 0, \forall \Gamma \in D \Rightarrow \Omega_{t+\tau}(\Gamma) = 0, \forall \Gamma \in D, t, \tau > 0$ . These questions and others will be answered in the next chapter.

### 3.8 Conclusion

One often sees in the historical and even in the recent literature, statements that imply the irreversibility results from the special nature of the initial state. For example:

“it is in any case impossible on the basis of *present* theory to carry out a mechanical derivation of the second law without specializing the initial state” E. Zermelo translated S G Brush Kinetic theory Vol 2 Irreversible processes pp 194 – 202 Pergamon Press (1966).

or,

“I have called it one of the most brilliant confirmations of the mechanical view of Nature that it provides an extraordinarily good picture of the dissipation of energy, as long as one assumes that the world began in an initial state satisfying certain conditions. I have called this state an improbable state.” L Boltzmann “A word from mathematics to energism” (1896).

“The time-asymmetry comes merely from the fact that the system has been *started off* in a very special (i.e. low entropy) state” p408 “The Emperor’s new mind”, Roger Penrose Oxford University Press 1989.

With respect to the Fluctuation Theorem the initial state need not be a state of particularly low probability. Equation (3.1.1) is a rather general distribution function and the FT holds for all distributions subject to the rather mild assumptions given above. Nor does it matter whether the initial distribution of states is what is called high or low entropy, the probability of positive dissipation is exponentially more likely (in time and in the number of degrees of freedom) than the probability of complementary negative dissipation (3.1.6). Of course if there are no dissipative fields and the initial distribution is an equilibrium distribution that the probability ratio predicted by the FT is unity for all averaging times.

What was never realized until the proof of the Fluctuation Theorem was a rather simple fact. Loschmidt’s objection (that for every trajectory there exists a conjugate anti trajectory and that summing over all such conjugate pairs implies that

irreversibility is impossible), is simply wrong. One must instead, consider not individual phase space trajectories but the probabilities of infinitesimal *sets* of trajectories having specified properties within some tolerance. It is this probability ratio that gives the dissipation function its meaning. It makes no mathematical sense to think that individual conjugate trajectory pairs somehow cancel each other out. Only when the system is at equilibrium do the probabilities of observing *sets* of trajectories and their conjugate antitrajectories, become equal.