

On the fluctuation theorem for the dissipation function and its connection with response theory

Denis J. Evans

Research School of Chemistry, Australian National University, Canberra, Australian Capital Territory 0200, Australia

Debra J. Searles^{a)}

Nanoscale Science and Technology Centre, School of Biomolecular and Physical Sciences, Griffith University, Brisbane, Queensland 4111, Australia

Stephen R. Williams

Research School of Chemistry, Australian National University, Canberra, Australian Capital Territory 0200, Australia

(Received 12 July 2007; accepted 23 October 2007; published online 3 January 2008)

Recently, there has been considerable interest in the fluctuation theorem (FT). The Evans-Searles FT shows how time reversible microscopic dynamics leads to irreversible macroscopic behavior as the system size or observation time increases. We show that the argument of this FT, the dissipation function, plays a central role in nonlinear response theory and derive the dissipation theorem, giving exact relations for nonlinear response of classical N -body systems that are more widely applicable than previous expressions. These expressions should be verifiable experimentally. When linearized they reduce to the well-known Green-Kubo expressions for linear response. © 2008 American Institute of Physics. [DOI: [10.1063/1.2812241](https://doi.org/10.1063/1.2812241)]

I. INTRODUCTION

Recently, there has been considerable interest in a group of theorems known collectively as the fluctuation theorem (FT). Various forms of this theorem have been confirmed experimentally.¹ The FT is remarkable in that it represents one of the few exact results that apply to nonequilibrium systems far from equilibrium. It has been shown that in the weak field limit, the Evans-Searles FT can be used to prove the well-known Green-Kubo expressions for the linear response of thermostatted systems to an applied dissipative field.² For technical reasons, the connection between the response theory and the FT for deterministic systems has been restricted to the linear regime close to equilibrium. Crooks³ has given a description of the connection between various stochastic versions of the fluctuation theorem and the so-called Kawasaki expression for the nonlinear response. However, even in the stochastic case, this connection did not extend to the more useful transient time correlation function (TTCF) formalism for the nonlinear response.⁴ In the present paper, we give the first proof that thermostatted nonlinear response theory (both Kawasaki and TTCF) can be derived directly using the Evans-Searles fluctuation function for classical systems satisfying time reversible deterministic dynamics. In fact, the new derivation shows that the formal expressions for the unnatural homogeneously thermostatted nonlinear response that have been derived in the past do also formally apply much more generally, including to naturally thermostatted systems as occur in real experimental systems.

In Sec. II, we will present general equations of motion

for the deterministic and reversible nonequilibrium, thermostatted dynamics and discuss the use of thermostats. We also show how the N -particle distribution function and phase variables evolve under these dynamics. In Sec. III, we present the Evans-Searles transient FT (TFT) and discuss some of its important implications. Section IV presents the main results of this paper: the dissipation theorem, and compares various representations of the nonequilibrium distribution function. Section V summarizes our findings.

II. THERMOSTATTED, NONEQUILIBRIUM DYNAMICS

Consider a classical system of N interacting particles in a volume V . The microscopic state of the system is represented by a phase space vector of the coordinates and momenta of all the particles, in phase space— $\{\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_N, \mathbf{p}_1, \dots, \mathbf{p}_N\} \equiv (\mathbf{q}, \mathbf{p}) \equiv \Gamma$, where $\mathbf{q}_i, \mathbf{p}_i$ are the position and momentum of particle i . Initially (at $t=0$), the microstates of the system are distributed according to a normalized probability distribution function $f(\Gamma, 0)$. While the results in this paper are generally applicable to demonstrate its application to realistic systems, we separate the N -particle system into a system of interest and a reservoir region containing N_w particles. We shall assume that the reservoir region contains many more particles than the system of interest, $N_w \gg (N - N_w)$, and we write the equations of motion for the composite N -particle system as

$$\dot{\mathbf{q}}_i = \mathbf{p}_i/m + \mathbf{C}_i(\Gamma) \cdot \mathbf{F}_e, \quad (1)$$

$$\dot{\mathbf{p}}_i = \mathbf{F}_i(\mathbf{q}) + \mathbf{D}_i(\Gamma) \cdot \mathbf{F}_e - S_i \alpha(\Gamma) \mathbf{p}_i,$$

where $\mathbf{F}_e$ is the dissipative external field that couples to the system via the phase functions $\mathbf{C}_i(\Gamma)$ and $\mathbf{D}_i(\Gamma)$,

^{a)}Electronic mail: d.bernhard@griffith.edu.au

$\mathbf{F}_i(\mathbf{q}) = -\partial\Phi(\mathbf{q})/\partial\mathbf{q}_i$ is the interatomic force on particle i , $\Phi(\mathbf{q})$ is the interparticle potential energy, and the last term $-S_i\alpha(\mathbf{\Gamma})\mathbf{p}_i$ is a deterministic time reversible thermostat used to add or remove heat from the particles in the reservoir region.⁵⁻⁷ We assume that in the absence of the thermostating terms, the adiabatic equations of motion preserve the phase space volume, $(\partial/\partial\mathbf{\Gamma})\cdot\dot{\mathbf{\Gamma}}^{\text{ad}}=0$: a condition known as the adiabatic incompressibility of phase space, or AIP.⁵ The thermostat multiplier can be chosen in a number of ways, such as using Gauss' principle of least constraint,⁵⁻⁷ 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, $S_i=0, 1 \leq i \leq (N-N_{\text{therm}})$; $S_i=1, (N-N_{\text{therm}}+1) \leq i \leq N, N_{\text{therm}} \leq N_w$. The equations of motion for the particles in the system of interest are quite natural. The reservoir region is assumed to not interact with the dissipative field $\mathbf{C}_i, \mathbf{D}_i=0$; $(N-N_w+1) \leq i \leq N$ and the equations of motion for the more distant reservoir particles, $N_{\text{therm}} \leq N_w$, are supplemented with the unnatural thermostat term. It is worth pointing out that as described, Eq. (1) is time reversible and heat can be either absorbed or given out by the thermostat. This construction has been applied in various studies (see, for example, Ref. 8–10). Of course, if $S_i=1$ for all i , we obtain a homogeneously thermostatted system that has been studied in detail in the past.⁵

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 [Eq. (1)], it is assumed that the reservoir momenta $\mathbf{p}_i$ are peculiar (i.e., measured relative to the local streaming velocity of the fluid or wall). When a Gaussian thermostat is used the thermostat multiplier is chosen to fix the peculiar kinetic energy of the wall particles,

$$\mathbf{K}_{\text{therm}} \equiv \sum_{S_i=1} \mathbf{p}_i^2/2m = (d_c N_{\text{therm}} - 1)k_B T_w/2, \quad (2)$$

with $N_{\text{therm}} = \sum S_i$. 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. It is assumed that $N_w, N_{\text{therm}} \gg (N - N_{\text{therm}}) \gg (N - N_w)$. This means that the entire wall region can be assumed to be arbitrarily close to equilibrium at the thermodynamic temperature T_w .

One might object that our analysis is compromised by our use of these artificial (time reversible) thermostats. However, the artificial thermostat region can be made arbitrarily remote from the system of interest by ensuring that the particles with $S_i=1$ are far from the system of interest.^{8,9,11} If this is the case, the system cannot “know” the precise details of how heat 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). This mathematical thermostat may be unrealistic, however, in the final analysis, it is very convenient but ultimately irrelevant device.⁹

As Tolman¹² pointed out, in a purely Hamiltonian system, the neglect of “irrelevant” degrees of freedom (as in

thermostats or by neglecting solvent degrees of freedom in a colloidal or Brownian system) will inevitably result in a non-conservation of phase space volume for the remaining “relevant” degrees of freedom. For a mathematical proof that when the thermostating region has a much larger number of degrees of freedom than the unthermostatted system of interest, the fluctuation theorem is independent of the mathematical details of how the thermostating see reference 9. This proof is given for an infinite family of so-called μ thermostats.

The exact equation of motion for the N -particle distribution function is the time reversible Liouville equation,⁵

$$\frac{\partial f(\mathbf{\Gamma}, t)}{\partial t} = -\frac{\partial}{\partial\mathbf{\Gamma}} \cdot [\dot{\mathbf{\Gamma}} f(\mathbf{\Gamma}, t)] \equiv -iL(\mathbf{\Gamma})f(\mathbf{\Gamma}, t), \quad (3)$$

where $iL(\mathbf{\Gamma})$ is the distribution function (or f -) Liouvillean and appears in the propagator for the phase space distribution function [$f(\mathbf{\Gamma}, t) = \exp(-iL(\mathbf{\Gamma})t)f(\mathbf{\Gamma}, 0)$]. The Liouville equation can also be written in Lagrangian form,¹³

$$\frac{df(\mathbf{\Gamma}, t)}{dt} = -f(\mathbf{\Gamma}, t) \frac{d}{d\mathbf{\Gamma}} \cdot \dot{\mathbf{\Gamma}} \equiv -\Lambda(\mathbf{\Gamma})f(\mathbf{\Gamma}, t). \quad (4)$$

The presence of the thermostat is reflected in the phase space expansion factor, $\Lambda(\mathbf{\Gamma}) \equiv \partial/\partial\mathbf{\Gamma} \cdot \dot{\mathbf{\Gamma}}$, which is, assuming AIP, to first order in N_{therm} , $\Lambda = -d_c N_{\text{therm}} \alpha$. The equation of motion for an arbitrary phase function $B(\mathbf{\Gamma})$ is⁵

$$\dot{B}(\mathbf{\Gamma}) = \dot{\mathbf{\Gamma}} \cdot \frac{dB}{d\mathbf{\Gamma}} \equiv iL(\mathbf{\Gamma})B(\mathbf{\Gamma}), \quad (5)$$

where $iL(\mathbf{\Gamma})$ is the phase variable (or p -) Liouvillean and appears in the propagator for phase variables [$B(\mathbf{\Gamma}(t)) = \exp(iL(\mathbf{\Gamma})t)B(\mathbf{\Gamma}(0))$]. The difference between the f -Liouvillean and the p -Liouvillean is $iL(\mathbf{\Gamma}) - iL(\mathbf{\Gamma}) = \Lambda(\mathbf{\Gamma})$. The time-reversibility condition implies that there exists a time reversal mapping M^T such that $\mathbf{\Gamma} = M^T \exp(iL t) M^T \exp(iL t) \mathbf{\Gamma}$, and this imposes some conditions on $\mathbf{C}_i(\mathbf{\Gamma})$ and $\mathbf{D}_i(\mathbf{\Gamma})$ in Eq. (1) that depend on M^T .

III. THE TRANSIENT FLUCTUATION THEOREM

Derivation of the Evans-Searles TFT considers the response of a system that is initially at equilibrium to application of a field. The initial distribution can be written in a quite arbitrary form

$$f(\mathbf{\Gamma}, 0) = \frac{\exp[-F(\mathbf{\Gamma})]}{\int d\mathbf{\Gamma} \exp[-F(\mathbf{\Gamma})]}, \quad (6)$$

where $F(\mathbf{\Gamma})$ is some arbitrary single valued real function for which $f(\mathbf{\Gamma}, 0) = f(M^T \mathbf{\Gamma}, 0)$. The Evans-Searles TFT (Refs. 13–16) states that provided the system satisfies the condition of ergodic consistency,¹⁵ the time-averaged dissipation function $\bar{\Omega}_t$, defined as^{15,16}

$$\int_0^t ds \Omega(\mathbf{\Gamma}(s)) \equiv \ln \left(\frac{f(\mathbf{\Gamma}(0), 0)}{f(\mathbf{\Gamma}(t), 0)} \right) - \int_0^t \Lambda(\mathbf{\Gamma}(s)) ds \equiv \bar{\Omega}_t t \quad (7)$$

satisfies the following time reversal symmetry.^{13–15}

$$\frac{p(\bar{\Omega}_t = A)}{p(\bar{\Omega}_t = -A)} = \exp[At]. \quad (8)$$

The derivation of the TFT is straightforward and it has been given in the past (see, for example, Ref. 15) and, therefore, will not be repeated here. The instantaneous dissipation function can be determined by differentiation of Eq. (7) as

$$\begin{aligned} \Omega(\Gamma) &= -\frac{\partial}{\partial \Gamma} \cdot \dot{\Gamma}(\Gamma) - \frac{\dot{\Gamma}(\Gamma)}{f(\Gamma, 0)} \cdot \frac{\partial}{\partial \Gamma} f(\Gamma, 0) \\ &= -\frac{\partial}{\partial \Gamma} \cdot \dot{\Gamma}(\Gamma) - \dot{\Gamma}(\Gamma) \cdot \frac{\partial}{\partial \Gamma} \ln f(\Gamma, 0) \end{aligned} \quad (9)$$

and, therefore, $\Omega(\Gamma)f(\Gamma, 0) = -(\partial/\partial \Gamma) \cdot (\dot{\Gamma}(\Gamma)f(\Gamma, 0))$, the divergence of $\dot{\Gamma}(\Gamma)f(\Gamma, 0)$ (i.e., the dissipation function weighted by the initial distribution is the weighted divergence of the phase space flow field).

It is very important to remember that the existence of the dissipation function $\Omega(\Gamma)$ at a phase point Γ requires that $f(\Gamma, 0) \neq 0$. The existence of the integrated form of the dissipation function requires that the dynamics is ergodically consistent [i.e., $\forall \Gamma, t$ for which $f(\Gamma, 0) \neq 0$, $f(\Gamma(t), 0) \neq 0$]. There are systems that fail to satisfy this condition. For example, if we let the initial distribution be microcanonical and further assume that the dynamics does not preserve the energy (there may be no thermostat or the thermostat may fix the kinetic temperature or so), then the ergodic consistency obviously breaks down. The existence of the dissipation function (7) only requires that the initial distribution is normalizable and that the ergodic consistency holds. To prove the Evans-Searles fluctuation theorem, requires an additional condition: namely, that the dynamics must be time reversal symmetric.

The TFT has generated much interest, as it shows how irreversibility emerges from the deterministic, reversible equations of motion,¹⁷ and is valid arbitrarily far from equilibrium. It provides a generalized form of the second law of thermodynamics that can be applied to small systems observed for short periods of time. It also resolves the long-standing Loschmidt paradox. The TFT has been verified experimentally.¹

The form of the above equation applies to any valid ensemble/dynamics combination, provided the distribution function is invariant with respect to time reversal. However, the precise expression for $\bar{\Omega}_t$ given in Eq. (7) is dependent on both the initial distribution and the dynamics. This result is extremely general. It is valid arbitrarily far from equilibrium. It leads to a number of other simple but important corollaries such as the second law inequality,¹⁸

$$\langle \bar{\Omega}_t \rangle_{\mathbf{F}_e, f(\Gamma, 0)} \geq 0, \quad \forall t, \mathbf{F}_e, f(\Gamma, 0), \quad (10)$$

and the nonequilibrium partition identity,

$$\exp(-\bar{\Omega}_t)_{\mathbf{F}_e, f(\Gamma, 0)} = 1, \quad \forall t, \mathbf{F}_e, f(\Gamma, 0). \quad (11)$$

The notation $\langle \cdots \rangle_{\mathbf{F}_e, f(\Gamma, 0)}$ implies that the ensemble average is taken over the ensemble defined by the initial distribution

$f(\Gamma, 0)$ [Eq. (6)], with any value (including 0) for the external field $\mathbf{F}_e$ in the equations of motion.

IV. THE DISSIPATION THEOREM

We now derive the *dissipation theorem*, which shows that, as well as being the subject of the TFT, the dissipation function is also the central argument of both linear (i.e., Green-Kubo theory) and nonlinear response theories. Firstly, we note the Dyson identity¹⁹ for two operators A and B ,

$$\begin{aligned} \exp[(A+B)t] &= \exp[At] + \int_0^t ds \exp[As]B \\ &\quad \times \exp[(A+B)(t-s)]. \end{aligned} \quad (12)$$

To prove this identity, we note that at $t=0$, it is clearly true and for $t>0$, the left hand side and the right hand side satisfy the same first order differential equation. Substituting $A = -iL(\Gamma)$ and $B = -\Lambda$ we find that

$$\begin{aligned} \exp[-(iL)t] &= \exp[-(iL + \Lambda)t] \\ &= \exp[-iLt] - \int_0^t ds \exp[-iLs]\Lambda \\ &\quad \times \exp[-(iL)(t-s)]. \end{aligned} \quad (13)$$

Recursive substitution and summing the resultant infinite series gives

$$\begin{aligned} f(\Gamma, t) &= \exp\left[-\int_0^t ds \Lambda(-s)\right] \exp[-iL(\Gamma)t] f(\Gamma, 0) \\ &= \exp\left[-\int_0^t ds \Lambda(-s)\right] \exp[-F(\Gamma(-t)) \\ &\quad + F(\Gamma(0))] f(\Gamma, 0). \end{aligned} \quad (14)$$

Comparing this with Eq. (7), we see that

Thus, the propagator for the N -particle distribution function $\exp[-iL(\Gamma)t]$ has a very simple relation to exponential time integral of the dissipation function. As shown below, in the case of isokinetic nonequilibrium dynamics, this equation reduces to Eq. (7.2.17) of Ref. 5. In the case of adiabatic (i.e., unthermostatted) dynamics for an ensemble that is initially a canonical ensemble, this result is equivalent to Eq. (7.2.8) of Ref. 5, which is the distribution function derived by Yamada and Kawasaki.²⁰ However, Eq. (15) is much more general and, similar to the TFT, can be applied to any initial ensemble and any time reversible, and possibly, thermostatted dynamics that satisfies AIF.

From Eq. (15), we can calculate nonequilibrium ensemble averages in the Schrödinger representation

$$\langle B(t) \rangle_{\mathbf{F}_e, f(\Gamma, 0)} = \left\langle B(0) \exp\left[-\int_0^t ds \Omega(\Gamma(s))\right] \right\rangle_{\mathbf{F}_e, f(\Gamma, 0)}, \quad (16)$$

and by differentiating and integrating Eq. (14) with respect to time, we can write the averages in the Heisenberg representation as

$$\langle B(t) \rangle_{\mathbf{F}_e f(\Gamma, 0)} = \langle B(0) \rangle_{f(\Gamma, 0)} + \int_0^t ds \langle \Omega(0) B(s) \rangle_{\mathbf{F}_e f(\Gamma, 0)}. \quad (17)$$

Equations (15)–(17) are new results. On both sides of Eq. (15)–(17), the time evolution is governed by the field dependent thermostatted equations of motion (1). The derivation of Eqs. (16) and (17) from the definition of the dissipation function (7) is called the dissipation theorem. This theorem is extremely general and allows the determination of the ensemble average of an arbitrary phase variable under very general conditions. Similar to the FT, it is valid arbitrarily far from equilibrium. As in the derivation of the FT the only unphysical terms in the derivation are the thermostating terms within the wall region. However, because these thermostating particles can be moved arbitrarily far from the system of interest, the precise mathematical details of the thermostat are unimportant. Since the number of degrees of freedom in the reservoir is assumed to be much larger than that of the system of interest, the reservoir can always be assumed to be in thermodynamic equilibrium. There is, therefore, no difficulty in defining the thermodynamic temperature of the walls. This is in marked contrast with the system of interest, which may be very far from equilibrium where the thermodynamic temperature cannot be defined.

For the special case of isokinetic dynamics where the kinetic energy $K_{\text{therm}}(\Gamma)$ of the thermostatted particles is fixed and if the initial distribution is isokinetic,

$$f(\Gamma, 0) \equiv f_K(\Gamma, 0) = \frac{\delta(2K_{\text{therm}} - (d_C N_{\text{therm}} - 1)k_B T_W) \exp[-\beta H_0(\Gamma)]}{\int d\Gamma \delta(2K_{\text{therm}} - (d_C N_{\text{therm}} - 1)k_B T_W) \exp[-\beta H_0(\Gamma)]}, \quad (18)$$

where $H_0(\Gamma)$ is the internal energy of the entire system and $\beta = 1/(k_B T_W)$. It is straightforward^{15,21} to show that the dissipation function is related to the generalized entropy production $\Sigma(\Gamma)$,

$$\Omega(\Gamma) \equiv \Sigma(\Gamma) = -\beta \mathbf{J}(\Gamma) \cdot \mathbf{V} \cdot \mathbf{F}_e. \quad (19)$$

Here, $\mathbf{V}$ is the volume of the system of interest and $\mathbf{J}(\Gamma)$ is the dissipative flux in the system of interest,

$$\mathbf{J}(\Gamma) \cdot \mathbf{V} \cdot \mathbf{F}_e = - \sum_{i=1}^{N-N_w} \left[\frac{\mathbf{p}_i}{m} \cdot \mathbf{D}_i - \mathbf{F}_i \cdot \mathbf{C}_i \right] \cdot \mathbf{F}_e. \quad (20)$$

Although here, we assumed a special dynamics where the kinetic energy of the thermostatted particles is fixed, the form of Eq. (19) must be true for other “thermostatted” dynamics (e.g., Nose-Hoover, constant energy, etc.).²¹ Furthermore, if the reservoir region does not directly interact with the field and N_{therm} is large, and much larger than the number of degrees of freedom in the system of interest, the form of Eq. (19) is true in general (e.g., for thermostats where higher order moments of the momenta are constrained, stochastic thermostats, etc.).⁹ The dissipative flux, volume, and field are properties of the system of interest and the only relevant property taken from the thermostatted region is its temperature.

Equation (17) can be written as the transient time correlation function expression,⁵ for the thermostatted nonlinear response of the phase variable B to the dissipative field $\mathbf{F}_e$,

$$\langle B(t) \rangle_{\mathbf{F}_e f(\Gamma, 0)} = \langle B(0) \rangle_{f(\Gamma, 0)} - \beta V \int_0^t ds \langle \mathbf{J}(0) B(s) \rangle_{\mathbf{F}_e f_K(\Gamma, 0)} \cdot \mathbf{F}_e. \quad (21)$$

In the weak field limit, this reduces to the well-known Green-Kubo expression⁵ for the linear response,

$$\lim_{\mathbf{F}_e \rightarrow 0} \langle B(t) \rangle_{\mathbf{F}_e f_K(\Gamma, 0)} = \langle B(0) \rangle_{f(\Gamma, 0)} - \beta V \int_0^t ds \langle \mathbf{J}(0) B(s) \rangle_{\mathbf{F}_e=0, f_K(\Gamma, 0)} \cdot \mathbf{F}_e, \quad (22)$$

where the right hand side is given by the integral of an *equilibrium* (i.e., $\mathbf{F}_e = 0$) time correlation function. Equation (19) is not new,¹⁵ and Eqs. (2) and (22) have been known for many years. However, this interpretation referring to the separation of the properties of the system of interest from those of the thermostating region (i.e., the Gedanken experiment) is new.

It is interesting to compare a number of different relationships between the distribution function, the dissipation function, and the phase space expansion factor. The first such relation is Eq. (15) above. We note that although the time argument in Eq. (15) is negative, the dynamics must still be governed by the field dependent, thermostatted equations of motion (1). Rewriting Eq. (7), we have

$$f(\Gamma(t), 0) = \exp \left[- \int_0^t ds \Omega(\Gamma(s)) + \Lambda(\Gamma(s)) \right] f(\Gamma(0), 0). \quad (23)$$

In a nonequilibrium steady state (SS), $\langle \Omega(t) \rangle_{\text{ss}} = -\langle \Lambda(t) \rangle_{\text{ss}}$. We also note that if the initial ensemble is microcanonical (has a uniform density) and the dynamics is such that the total energy (system of interest *plus* walls and thermostat) is constant, then $\Omega(t) = -\Lambda(t)$, $\forall t$. Lastly, we have the formal solution of the Liouville equation in its Lagrangian form (4),¹³

$$f(\Gamma(t), t) = \exp \left[- \int_0^t ds \Lambda(\Gamma(s)) \right] f(\Gamma(0), 0). \quad (24)$$

Rather obviously the results of the dissipation theorem [Eq. (17)] can also be used to obtain a fluctuation dissipation theorem, as described in Ref. 22, by considering the case where the phase function $B(\Gamma) = \mathbf{J}(\Gamma)$. Furthermore, following Ref. 22, we find that when the equilibrium dissipative flux autocorrelation function is δ correlated, $\langle \mathbf{J}(t_1) \mathbf{J}(t_2) \rangle_{\mathbf{F}_e=0} = \langle \mathbf{J}(t_1) \mathbf{J}(t_2) \delta(t_2 - t_1) \rangle_{\mathbf{F}_e=0}$, we obtain the fluctuation dissipation relation, $\lim_{\mathbf{F}_e \rightarrow 0} \langle \mathbf{J}(t) \rangle_{\mathbf{F}_e} = -(1/3) \beta V \langle \mathbf{J}(0) \cdot \mathbf{J}(0) \rangle_{\mathbf{F}_e=0} \cdot \mathbf{F}_e$.

V. CONCLUSIONS

In this paper, we have shown the central importance of the dissipation function to nonequilibrium statistical mechanics. It is the argument of both the FT and the dissipation theorem. These theorems are both exact arbitrarily far from

equilibrium. The FT has been confirmed in laboratory experiments¹ and we see no reason why the dissipation theorem cannot be likewise tested in the laboratory.

Originally, the dissipation function was defined in order to characterize the ratio of probabilities p_r of observing infinitesimal sets of phase space trajectories originating ($t=0$) in a volume $dV_{\Gamma(0)}$ to the probability of observing at $t=0$, their time reversed antitrajectories $dV_{\Gamma^*(0)}$, where $\Gamma^*(0) \equiv M^T \Gamma(t)$,¹⁵

$$\frac{p_r(dV_{\Gamma(0)},0)}{p_r(dV_{\Gamma^*(0)},0)} = \exp \left[\int_0^t ds \Omega(\Gamma(s)) \right]. \quad (25)$$

Combining Eq. (25) with Eq. (15) shows that the nonequilibrium N -particle distribution function at time t can be written in terms of the ratio of probabilities of observing sets of phase trajectories,

$$f(\Gamma(0),t) = \frac{p_r(dV_{M^T \Gamma(-1)},0)}{p_r(dV_{\Gamma(0)},0)} f(\Gamma(0),0). \quad (26)$$

We find it remarkable that the measure of *irreversibility* given in Eq. (25) by the dissipation function also features so centrally in the dissipation theorem. Our work shows that this measure of irreversibility is the prime function in determining how a nonequilibrium system will *respond* to a nonequilibrium perturbation or dissipative field.

We have given a derivation of the dissipation theorem [Eqs. (15)–(17)] for an exceedingly general set of time reversible equations of motion (1) and for an arbitrary initial distribution $f(\Gamma,0)$ [Eq. (6)]. If one substitutes Ω for B in Eq. (17) and then combines the resulting equation with the second law of inequality (10), one can prove the following inequality for integrals of transient correlations of the dissipation function:

$$\int_0^t ds \langle \Omega(0) \Omega(s) \rangle_{F_e, f(\Gamma,0)} \geq 0, \quad \forall t, F_e, f(\Gamma,0). \quad (27)$$

So, not only does the dissipation autocorrelation function start with a positive value, but also for all normalizable initial distributions and for any well-defined dynamics with an arbitrarily strong external field (if any). Any negative tails in the transient correlation function *cannot* be strong enough for the integral (27) to ever become negative.

The dissipation theorem can be used to calculate the ensemble average of an arbitrary phase variable and for arbitrarily strong dissipative fields, F_e . In deriving Eq. (19), we considered a system that preserves the initial (equilibrium) distribution in the absence of an external dissipative field. Our formalism is sufficiently general to also describe the situation where there is no external dissipative field $F_e=0$ but where dissipation still occurs because the initial distribution is not preserved by the dynamics. The new results given in Eqs. (15)–(17) apply to this more general circumstance (this includes systems subject to rapid temperature or pressure quenches, etc., but in which there is no applied *mechanical* dissipative field).

We have argued that although the derivation employs unphysical thermostatting terms, these are a convenient but

ultimately irrelevant device. For both the dissipation theorem and FT, dissipation takes place only in the system of interest (as measured by the dissipation function) and on average, heat is lost to a surrounding reservoir region that is arbitrarily close to thermodynamic equilibrium at a known temperature.

Recently, there have been a number of papers that have claimed that in some glassy systems, the fluctuation theorem and/or the fluctuation dissipation theorem “seems to fail.”²³ We have a number of comments to make on this matter. Firstly, if a mathematical theorem has been proved and the proof is correct (there are proofs of the theorems in question that are undoubtedly correct), then no experiment can prove that the theorem is incorrect. However, any mathematical theorem involves a set of conditions. It may well be that in certain situations, the conditions required for the theorem to apply may not hold. In such a case, the mathematical *relation* expected from the application of the theorem to the system being studied may not be satisfied.

When people say that the fluctuation theorem or the fluctuation dissipation theorem fails for glasses, they usually mean the relations obtained by applying these systems to a canonical distribution of states [i.e., the fluctuation relation of Eq. (8) with Eq. (19) as the dissipation function, etc.] With respect to linear response theory and the fluctuation dissipation theorem, there are a number of necessary conditions for these theorems to apply. For Eq. (19) to apply, these conditions include²⁴

- (1) the initial state must be preserved by field-free dynamics (there cannot be any slow relaxation);
- (2) the initial distribution should be Boltzmann distributed at least over some ergodic subdomain of phase space;
- (3) if the phase space can be divided into distinct subdomains (as is natural in systems where multiple phases exist, glassy systems, and where allotropes exist), the subdomains should be robust with respect to small changes in macroscopic conditions; and, for linear response theory to apply,
- (4) the applied fields need to be sufficiently weak that the system responds *linearly* with respect to field strength.

Recently, we have shown that if one fixes the absolute magnitude of a constant driving force and one then lowers the temperature and approaches the glass transition, the reduced magnitude of the driving force (reduced in accordance with the rapidly diverging response time of the system) increases without bound.²⁵ In this case, the expected linear response relation will fail because as the temperature is lowered at a fixed force value, the system moves into the nonlinear response regime. For a detailed discussion of the application of the transient fluctuation theorem to glassy systems, see Ref. 24.

It has been proposed that in nonequilibrium systems, the path integral of the entropy production should be an extremum (see, for example, Ref. 26). To the leading order close to equilibrium, the dissipation function is the entropy production of linear irreversible thermodynamics, $\sigma(t)V$, $\lim_{F_e \rightarrow 0} (\Omega(t) - \sigma(t)V) = O(F_e^4)$. Our work shows that in natural systems that exchange heat with their surroundings, the

maximum entropy production hypothesis, at best, can only be an approximation. In none of our expressions for the non-equilibrium N -particle distribution function [see Eqs. (15), (23), and (24)], the probability density is controlled *solely* by the path integral of the entropy production.

ACKNOWLEDGMENTS

We would like to thank Lamberto Rondoni for his valuable comments on this work. We also thank the Australian Research Council for its support of this work.

- ¹G. M. Wang, E. M. Sevick, E. Mittag, D. J. Searles, and D. J. Evans, Phys. Rev. Lett. **89**, 050601 (2002); D. M. Carberry, J. C. Reid, G. M. Wang, E. M. Sevick, D. J. Searles, and D. J. Evans, *ibid.* **92**, 140601 (2004); J. C. Reid, D. M. Carberry, G. M. Wang, E. M. Sevick, D. J. Evans, and D. J. Searles, Phys. Rev. E **70**, 016111 (2004); D. Collin, F. Ritort, C. Jarzynski, S. B. Smith, I. Tinoco, Jr., and C. Bustamante, Nature (London) **437**, 231 (2005); J. Liphardt, S. Dumont, S. B. Smith, I. Tinoco, Jr., and C. Bustamante, Science **296**, 1832 (2002); E. H. Trepagnier, C. Jarzynski, F. Ritort, G. E. Crooks, C. J. Bustamante, and J. Liphardt, Proc. Natl. Acad. Sci. U.S.A. **101**, 15038 (2004); S. Schuler, T. Speck, C. Tietz, J. Wrachtrup, and U. Seifert, Phys. Rev. Lett. **94**, 180602 (2005); D. M. Carberry, M. A. B. Baker, G. M. Wang, E. M. Sevick, and D. J. Evans, J. Opt. A, Pure Appl. Opt. **9**, S204 (2007).
- ²D. J. Evans, D. J. Searles, and L. Rondoni, Phys. Rev. E **71**, 056120 (2005).
- ³G. E. Crooks, Phys. Rev. E **61**, 2361 (2000).
- ⁴G. P. Morriss and D. J. Evans, Phys. Rev. A **35**, 792 (1987); D. J. Evans and G. P. Morriss, *ibid.* **38**, 4142 (1988); I. Borzsak, P. T. Cummings, and D. J. Evans, Mol. Phys. **100**, 2735 (2002); B. D. Todd, Phys. Rev. E **56**, 6723 (1997); J. Delhommelle, P. T. Cummings, and J. Petrávic, J. Chem. Phys. **123**, 114505 (2005); J. Delhommelle and P. T. Cummings, Phys. Rev. B **72**, 172201 (2005); G. Pan and C. McCabe, J. Chem. Phys. **125**, 194527 (2006); J. Delhommelle, P. T. Cummings, and J. Petrávic, *ibid.* **123**, 114505 (2005).
- ⁵D. J. Evans and G. P. Morriss, *Statistical Mechanics of Nonequilibrium Liquids* (Academic, London, 1990), also available at <http://rsc.anu.edu.au/~evans/evansmorrissbook.htm> (http://epress.anu.edu.au/sm_citation.html).
- ⁶W. G. Hoover, A. J. C. Ladd, and B. Moran, Phys. Rev. Lett. **48**, 1818 (1982).
- ⁷D. J. Evans, J. Chem. Phys. **78**, 3297 (1983).
- ⁸G. Ayton, D. J. Evans, and D. J. Searles, J. Chem. Phys. **115**, 2035 (2001).
- ⁹S. R. Williams, D. J. Searles, and D. J. Evans, Phys. Rev. E **70**, 066113 (2004).
- ¹⁰C. Jarzynski, J. Stat. Phys. **98**, 77 (2000), G. Gallavotti, arXiv:cond-mat/0601049v2.
- ¹¹S. R. Williams, D. J. Searles, and D. J. Evans, Mol. Phys. **105**, 1059 (2007).
- ¹²R. C. Tolman, *The Principles of Statistical Mechanics* (Dover, New York, 1979).
- ¹³D. J. Evans and D. J. Searles, Phys. Rev. E **52**, 5839 (1995).
- ¹⁴D. J. Evans and D. J. Searles, Phys. Rev. E **50**, 1645 (1994).
- ¹⁵D. J. Evans and D. J. Searles, Adv. Phys. **51**, 1529 (2002).
- ¹⁶D. J. Searles and D. J. Evans, J. Chem. Phys. **113**, 3503 (2000).
- ¹⁷By time-reversible equations of motion, we mean that there exists a time reversal mapping M^T , such that $\Gamma = M^T \exp(iLt) M^T \exp(iLt) \Gamma$.
- ¹⁸D. J. Searles and D. J. Evans, Aust. J. Chem. **57**, 1119 (2004).
- ¹⁹F. J. Dyson, Phys. Rev. **75**, 486 (1949).
- ²⁰T. Yamada and K. Kawasaki, Prog. Theor. Phys. **38**, 1031 (1967).
- ²¹D. J. Searles, L. Rondoni, and D. J. Evans, J. Stat. Phys. **127**, 184101 (2007).
- ²²R. Kubo, Rep. Prog. Phys. **29**, 255 (1966).
- ²³A. Crisanti and F. Ritort, J. Phys. A **36**, R181 (2003); F. Zamponi, F. Bonetto, L. F. Cugliandolo, and J. Kurchan, J. Stat. Mech.: Theory Exp. P09013 (2005); G. Diezemann and R. Böhmer, J. Chem. Phys. **124**, 214507 (2006); F. Romá, S. Bustingorry, P. M. Gleiser, and D. Domínguez, Phys. Rev. Lett. **98**, 097203 (2007).
- ²⁴S. R. Williams and D. J. Evans, J. Chem. Phys. **128**, 1337 (2007).
- ²⁵S. R. Williams and D. J. Evans, Phys. Rev. Lett. **96**, 015701 (2006).
- ²⁶R. Dewar, J. Phys. A **36**, 631 (2003).