

# Time-dependent response theory and nonequilibrium free-energy relations

Stephen R. Williams\* and Denis J. Evans†

Research School of Chemistry, The Australian National University, Canberra, ACT 0200, Australia

(Received 5 May 2008; published 15 August 2008)

The mathematics of time-dependent nonlinear response theory frequently leads to results which although formally exact, are not amenable to experimental application or even to use in computer simulations. Here we give a rigorous derivation of a tractable expression for the thermostatted nonlinear response of classical many body systems to time-dependent dissipative fields. The theory also allows for the concurrent parametric transformation of the system Hamiltonian. Our analysis shows once again the intimate relationship between nonequilibrium free-energy relations such as the Jarzynski equality and nonlinear response theory. We make a few remarks concerning the maximum entropy production approximation.

DOI: 10.1103/PhysRevE.78.021119

PACS number(s): 05.20.-y, 05.45.-a, 05.60.Cd, 61.20.Lc

## I. INTRODUCTION

There are two widely used [1–5] forms of nonlinear response theory for constant dissipative fields that have been employed in computer simulations: the so-called Kawasaki formalism [6] and the Evans and Morriss transient time correlation formalism [7]. In spite of a number of attempts [8–10], to date there has not been any tractable rigorous extension of these methods to the case of time-dependent dissipative fields. In the present paper we provide a rigorous extension of the Kawasaki formalism for the nonlinear response of thermostatted  $N$ -particle classical systems to large, time-dependent, dissipative fields. We also allow the system to possibly undergo a concurrent parametric change in its Hamiltonian. This allows us to discuss the nonlinear response of systems to dissipative fields where the underlying equilibrium state is changing concurrently. We use the term *changing state* to specifically refer to the situation where the underlying equilibrium state (that would be generated if the current Hamiltonian parameters were fixed and the system allowed time to equilibrate), is changing due to a change in the system's Hamiltonian, volume, etc. This allows us to study the similarities and differences between the dissipative flux that appears in nonlinear response theory and the work function employed in the Jarzynski equality [11].

## II. GENERALIZED KAWASAKI DISTRIBUTION FUNCTION (KDF)

We consider a system of  $i=1, N$  classical particles subject to the nonautonomous equations of motion,

$$\dot{\mathbf{q}}_i = \frac{\mathbf{p}_i}{m_i} + \mathbf{C}_i(\Gamma) \cdot \mathbf{F}_e(t),$$

$$\dot{\mathbf{p}}_i = \mathbf{F}_i(\mathbf{q}, \lambda(t)) + \mathbf{D}_i(\Gamma) \cdot \mathbf{F}_e(t) - S_i \alpha \mathbf{p}_i, \quad (1)$$

where  $\mathbf{C}_i$  and  $\mathbf{D}_i$  are second rank tensors coupling the time-dependent external field  $\mathbf{F}_e(t)$  to the equations of motion, the

parametrically dependent force is given by  $\mathbf{F}_i(\mathbf{q}, \lambda) = -\nabla_{\mathbf{q}_i} H(\Gamma, \lambda)$  and the thermostat multiplier, which holds the kinetic energy of a reservoir region ( $S_i=1$  if  $i \in$  reservoir and  $S_i=0$  otherwise) fixed, at a value of  $K(\Gamma)=K_0$  is

$$\alpha = \frac{\sum_{i=1}^N \mathbf{F}_i(\mathbf{q}, \lambda) \cdot \mathbf{p}_i S_i + S_i \mathbf{p}_i \cdot \mathbf{D}_i(\Gamma) \cdot \mathbf{F}_e(t)}{\sum_{i=0}^N S_i \mathbf{p}_i \cdot \mathbf{p}_i}. \quad (2)$$

The reservoir region is far enough removed from the system of interest that the synthetic thermostat cannot possibly introduce any artifacts (see [12,13] for details). The Hamiltonian is given by

$$H(\Gamma, \lambda) = \sum_{i=1}^N \frac{\mathbf{p}_i \cdot \mathbf{p}_i}{2m_i} + \Phi(\mathbf{q}, \lambda), \quad (3)$$

where  $\Phi(\mathbf{q}, \lambda)$  is the parametrically dependent potential. We now appeal to physics to classify what part of the external influence is represented by the potential  $\Phi(\lambda(t))$  and what part is represented by the dissipative vector field  $\mathbf{F}_e(t)$ . If the field is set to zero  $\mathbf{F}_e=0$  and the parameter  $\lambda$  is held fixed, we can obtain the following time-independent solution to the Liouville equation:

$$\begin{aligned} f_{eq}(\Gamma, \lambda) &= \frac{\exp[-\beta H(\Gamma, \lambda)] \delta[K(\Gamma) - K_0]}{\int d\Gamma \exp[-\beta H(\Gamma, \lambda)] \delta[K(\Gamma) - K_0]} \\ &= \exp\{-\beta[H(\Gamma, \lambda) - A(T, \lambda)]\} \delta[K(\Gamma) - K_0], \end{aligned} \quad (4)$$

where

$$K(\Gamma) \equiv \frac{1}{2} \sum_{i=1}^N S_i \mathbf{p}_i \cdot \mathbf{p}_i / m_i,$$

$$A = -k_B T \ln \left\{ \int d\Gamma \exp[-\beta H(\Gamma, \lambda)] \delta[K(\Gamma) - K_0] \right\}$$

is the thermodynamic potential and  $k_B T \equiv \beta^{-1} = 2K_0 / (DN_R - 1)$ , where  $D$  is the Cartesian dimension,  $N_R \equiv \sum_{i=1}^N S_i$  is the

\*swilliams@rsc.anu.edu.au

†evans@rsc.anu.edu.au

number of particles in the reservoir region,  $k_B$  is Boltzmann's constant, and  $T$  is the temperature.

Whether the system relaxes to the equilibrium given by Eq. (4) or not, depends on the physics of the situation. For example, if we consider a system consisting of equal numbers of oppositely charged atoms as found in a molten salt, a constant electric field could be used to drive the system into a dissipative nonequilibrium steady state. On the other hand if we lower the system's temperature and then let it equilibrate, the system may crystallize and then if acted on by the same electric field would relax to a new (polarized) equilibrium state in which there is no net dissipation of energy. This is in spite of the fact that both systems are governed by precisely the same equations of motion with the only difference being, say, a change in the temperature  $T$ . So the decision as whether to represent an external influence using the parametric dependence of the potential  $\lambda(t)$  or the external field  $\mathbf{F}_e$  must be made by appealing to physics rather than mathematics.

We require that this decision be such that if we turn the field off  $\mathbf{F}_e = \mathbf{0}$  and hold the parameter  $\lambda$  constant then the system relaxes to the equilibrium distribution function given by Eq. (4), which we call *the underlying equilibrium state*. Any part of the external influence that satisfies this condition is represented by the parametric dependence of the Hamiltonian if possible, otherwise it is represented by the external field.

The Lagrangian form of the Liouville equation [14] is given by

$$\frac{d}{dt}f(\mathbf{\Gamma}(t), t) = -f(\mathbf{\Gamma}(t), t)\Lambda(\mathbf{\Gamma}(t)), \quad (5)$$

where the phase space compression rate  $\Lambda$  turns out to be directly related to the rate at which energy is increased in the system by the thermostat (usually a negative quantity). This rate of heat gain is given by  $\dot{Q}(\mathbf{\Gamma}(t)) = k_B T \Lambda(\mathbf{\Gamma}(t))$ . We note that the total change in value of the Hamiltonian as the trajectory evolves satisfies the equation

$$\begin{aligned} H(\mathbf{\Gamma}(0), \lambda(0)) &= H(\mathbf{\Gamma}(t), \lambda(t)) - \int_0^t ds \dot{\lambda}(s) \frac{\partial H(\mathbf{\Gamma}(s), \lambda(s))}{\partial \lambda} \\ &\quad - \int_0^t ds \dot{Q}(\mathbf{\Gamma}(s)) + \int_0^t ds \mathbf{J}(\mathbf{\Gamma}(s)) V \cdot \mathbf{F}_e(s), \end{aligned} \quad (6)$$

where  $\dot{Q} = -\alpha \sum_{i=1}^N S_i \mathbf{p}_i \cdot \mathbf{p}_i / m_i$  is the rate of heat gain by the system from the thermostat and the dissipative flux  $\mathbf{J}(\mathbf{\Gamma})$  is thus defined by Eq. (6) given the external field  $\mathbf{F}_e(t)$ , which is specified in the equations of motion, Eqs. (1), and the expression for the Hamiltonian, Eq. (3). Close to equilibrium where linear irreversible thermodynamics applies the term  $-\beta \mathbf{J} V \cdot \mathbf{F}_e$  is the *spontaneous entropy production*. The solution to Eq. (5), for a system initially in equilibrium at time  $t=0$  is

$$f(\mathbf{\Gamma}(t), t) = f_{eq}(\mathbf{\Gamma}(0), \lambda(0)) \exp\left(-\beta \int_0^t ds \dot{Q}(\mathbf{\Gamma}(s))\right). \quad (7)$$

This equation is quite general; if we start in equilibrium  $f_{eq}(\mathbf{\Gamma}(0), \lambda(0))$  and the equations of motion, Eq. (1), are obeyed, then Eq. (7) will be applicable. Combining Eqs. (4)–(7) we obtain

$$\begin{aligned} f(\mathbf{\Gamma}(t), t) &= f_{eq}(\mathbf{\Gamma}(t), \lambda(t)) \exp\left(-\beta \left[ \int_0^t ds \mathbf{J}(\mathbf{\Gamma}(s)) V \cdot \mathbf{F}_e(s) \right. \right. \\ &\quad \left. \left. - \dot{\lambda}(s) \left( \frac{\partial H(\mathbf{\Gamma}(s), \lambda(s))}{\partial \lambda} - \frac{\partial A(\lambda(s))}{\partial \lambda} \right) \right] \right), \end{aligned} \quad (8)$$

where the thermodynamic potential  $A(\lambda)$  is given in terms of the underlying equilibrium state. We can transform Eq. (8) such that it is compatible with the standard (Heisenberg) form of the Kawasaki distribution. The choice of the initial phase  $\mathbf{\Gamma} \equiv \mathbf{\Gamma}(0)$  is quite arbitrary. Whatever that choice, Eq. (8) gives us an expression for the distribution function at the time evolved phase at the later time. Let us introduce the substitution for the phase variable  $\mathbf{\Gamma}' = \mathbf{\Gamma}(t)$ , then  $\mathbf{\Gamma}'(s) \equiv \mathbf{\Gamma}(t+s)$ , and  $\mathbf{\Gamma}(s) = \mathbf{\Gamma}'(s-t)$ , so

$$\begin{aligned} f(\mathbf{\Gamma}, t) &= f_{eq}(\mathbf{\Gamma}, \lambda(t)) \exp\left(-\beta \left[ \int_0^t ds \mathbf{J}(\mathbf{\Gamma}(s-t)) V \cdot \mathbf{F}_e(s) \right. \right. \\ &\quad \left. \left. - \dot{\lambda}(s) \left( \frac{\partial H(\mathbf{\Gamma}(s-t), \lambda(s))}{\partial \lambda} - \frac{\partial A(\lambda(s))}{\partial \lambda} \right) \right] \right), \end{aligned} \quad (9)$$

where we have dropped the primes to make the notation simpler [15]. We can make the substitution  $v = t-s$  and upon recognizing that  $v$  is a dummy variable we obtain

$$\begin{aligned} f(\mathbf{\Gamma}, t) &= f_{eq}(\mathbf{\Gamma}, \lambda(t)) \exp\left(-\beta \int_0^t ds \mathbf{J}(\mathbf{\Gamma}(-s)) V \cdot \mathbf{F}_e(t-s)\right) \\ &\quad \times \exp\left(\beta \int_0^t ds \dot{\lambda}(t-s)\right) \\ &\quad \times \left[ \frac{\partial H(\mathbf{\Gamma}(-s), \lambda(t-s))}{\partial \lambda} - \frac{\partial A(\lambda(t-s))}{\partial \lambda} \right], \end{aligned} \quad (10)$$

where  $\mathbf{\Gamma}(-s)$  is the phase space vector arrived at if we start at  $\mathbf{\Gamma}$  and run the equations of motion back in time for duration  $s$  using the protocol  $\mathbf{F}_e(t-u)$ ,  $\lambda(t-u)$ , initially with the time variable  $u=0$  and finally with  $u=s$ . In this form it is explicit which part of the distribution function depends on the time  $t$  and which part depends upon the phase space vector  $\mathbf{\Gamma}$ . This is important if we wish to take the partial derivative of the distribution function with respect to time. Equation (10) is the generalized time-dependent Kawasaki distribution function (KDF).

There are a number of special cases of Eq. (10) that were previously known. If the process is such that the parameter is always held fixed,  $\dot{\lambda}=0$ ,  $\lambda=\lambda_0$ , and the free energy does not change,  $\Delta A=0$ , the distribution function reduces to

$$f(\Gamma, t) = f_{eq}(\Gamma, \lambda_0) \exp\left(-\beta \int_0^t ds \mathbf{J}(\Gamma(-s)) V \cdot \mathbf{F}_e(t-s)\right). \quad (11)$$

This equation was previously derived [16] by assuming that the nonequilibrium fine grained Gibbs entropy,  $S(t) = -k_B \int d\Gamma f(\Gamma, t) \ln f(\Gamma, t)$ , is a minimum, subject to the constraints that the distribution is normalized, that the average energy is fixed, and that the average dissipative flux is fixed at any time and also is a continuous function of time. This result is now confirmed without assumptions about the entropy or the free energy.

Let us now additionally suppose the external field is held constant,  $\mathbf{F}_e(t) = \mathbf{F}_e \forall t > 0$ , then we recover the standard KDF,

$$f(\Gamma, t) = f_{eq}(\Gamma, \lambda_0) \exp\left(-\beta \int_0^t ds \mathbf{J}(\Gamma(-s)) V \cdot \mathbf{F}_e\right). \quad (12)$$

This equation was first derived for adiabatic systems in 1967 [6] and for thermostatted systems in 1985 [7].

We can also obtain a generalized linear response theory by taking the first order Taylor expansion for the exponential in Eq. (10) which gives

$$\begin{aligned} \langle B \rangle_{\lambda(t), t} &= \langle B \rangle_{\lambda(t), eq} - \beta V \int_0^t ds \langle \mathbf{J}(\Gamma(-s)) B(\Gamma) \rangle_{eq} \cdot \mathbf{F}_e(t-s) \\ &+ \beta \int_0^t ds \left\langle \left( \frac{\partial H(\Gamma(-s), \lambda=0)}{\partial \lambda} \right. \right. \\ &\left. \left. - \frac{\partial A(\lambda=0)}{\partial \lambda} \right) B(\Gamma, \lambda(t)) \right\rangle_{\lambda(0), eq} \dot{\lambda}(t-s). \end{aligned} \quad (13)$$

This result is new since it includes contributions from the parametric change in the Hamiltonian. Here we have ignored any terms of  $\mathcal{O}(F_e^2)$ ,  $\mathcal{O}(\lambda^2)$ , or higher where  $F_e$  is the magnitude of the field  $\mathbf{F}_e$ . For this reason we can consider the average  $\langle \dots \rangle_{\lambda(0), eq}$  taken at equilibrium with the Hamiltonian parameter held fixed at the value of  $\lambda(0)$ . To see why equilibrium averages may be used in Eq. (13) consider that their leading order change is  $\mathcal{O}(F_e)$ ,  $\mathcal{O}(\lambda)$ , or higher. The two relevant averages are multiplied by the terms  $F_e$  and  $\dot{\lambda}$ , respectively, resulting in the change to the products,  $\langle \dots \rangle \cdot \mathbf{F}_e$  and  $\langle \dots \rangle \dot{\lambda}$  being at least quadratic in the field or perturbation, respectively. So to linear order the nonequilibrium change in these averages does not contribute to the left-hand side of Eq. (13).

### III. PROOF OF THE JARZYNSKI EQUALITY FROM THE KDF

The Jarzynski equality (JE) [11,17] (for recent reviews of fluctuation theorems including the JE, see [18,19]) is a well known nonequilibrium free-energy relation. Traditionally changes in the equilibrium free energy are computed by integrating over a single quasistatic path. In contrast the JE

shows how equilibrium free-energy differences can be calculated from the work integrated along an ensemble of nonequilibrium paths. The JE is a statement about ensemble averages, not single trajectories. The equality gives the change in free energy  $\Delta A$  as,  $\langle \exp(-\beta \Delta w_t) \rangle = \exp(-\beta \Delta A)$ , where the ensemble average  $\langle \dots \rangle$  is over all trajectories which transform the system during the time interval  $0 < \tau < t$  and the work  $\Delta w_t$  is computed over this interval for each trajectory. At time  $t$  the nonequilibrium protocol stops, no further work is done, and the system relaxes to the new equilibrium whose free energy differs from the initial equilibrium's by the amount  $\Delta A$ .

There are three observations that we can make concerning Eq. (10) which are relevant to the JE. For a quasistatic process where  $f(\Gamma, t) = f_{eq}(\Gamma, \lambda(t))$ , we have

$$\left\langle \int_0^t ds \dot{\lambda}(t-s) \left( \frac{\partial H(\Gamma(-s), \lambda(t-s))}{\partial \lambda} - \frac{\partial A(\lambda(t-s))}{\partial \lambda} \right) - \mathbf{J}(\Gamma(-s)) V \cdot \mathbf{F}_e(t-s) \right\rangle_{qs} = 0. \quad (14)$$

So the term under the integral is the purely irreversible, or the dissipative, work because it “dissipates” to zero for quasistatic processes. Secondly, since  $\int_0^t ds \dot{\lambda}(t-s) \frac{\partial A(\lambda(t-s))}{\partial \lambda} = A(\lambda(0)) - A(\lambda(t)) = \Delta A$  is the reversible work, it follows that

$$\begin{aligned} \int_0^t ds -\mathbf{J}(\Gamma(-s)) V \cdot \mathbf{F}_e(t-s) + \dot{\lambda}(t-s) \frac{\partial H(\Gamma(-s), \lambda(t-s))}{\partial \lambda} \\ = \Delta w_t \end{aligned} \quad (15)$$

is the work performed over a time  $t$ .

Thirdly, since the KDF distribution function is normalized we have,

$$\left\langle \exp\left(\beta \left[ \int_0^t ds \mathbf{J}(\Gamma(-s)) V \cdot \mathbf{F}_e(t-s) - \dot{\lambda}(t-s) \left( \frac{\partial H(\Gamma(-s), \lambda(t-s))}{\partial \lambda} - \frac{\partial A(\lambda(t-s))}{\partial \lambda} \right) \right] \right) \right\rangle_{eq, \lambda(t)} = 1, \quad (16)$$

where the subscript to the angle brackets denotes that the average is taken over the equilibrium distribution function,  $f_{eq}(\Gamma, \lambda(t))$ .

We note that the equilibrium distribution function is invariant to a time reversal transformation  $f_{eq}(\Gamma, \lambda_0) = f_{eq}(\Gamma^*, \lambda_0)$  and that the Jacobian for this transformation is unity. By exploiting the time reversal symmetry in the equations of motion Eq. (1) we can see that Eq. (16) is valid for an ensemble initially at equilibrium, with  $\lambda_0 = \lambda(t)$ , which is then driven by the reverse protocol,  $\lambda(t-s)$ ,  $\mathbf{F}_e(t-s)$ . The time-dependent protocol is completely arbitrary, as is the equilibrium given by  $\lambda(t)$ , and so no new restrictions are introduced. Equation (16) can now be written much more concisely in terms of the purely irreversible or dissipative work with the time transformed,  $t' = t-s$ , as

$$\langle \exp(-\beta \Delta w_{t'}^{irr}) \rangle_{eq, \lambda_0} = 1, \quad (17)$$

where  $\Delta w_{t'}^{irr}$  is the irreversible work done for the trajectories driven by the reverse protocol, starting in equilibrium  $f_{eq}(\mathbf{\Gamma}, \lambda_0 = \lambda(t))$ ,

$$\Delta w_{t'}^{irr} \equiv \int_0^{t'} ds \left[ \dot{\lambda}(s) \left( \frac{\partial H(\mathbf{\Gamma}(s), \lambda(s))}{\partial \lambda} - \frac{\partial A(\lambda(s))}{\partial \lambda} \right) - \mathbf{J}(\mathbf{\Gamma}(s)) V \cdot \mathbf{F}_e(s) \right]. \quad (18)$$

Equation (17) is a generalized version of the nonequilibrium partition identity. It shows that the ensemble average of the negative exponential integral of the purely irreversible work (including contributions from the dissipative flux  $\mathbf{J}$  and from the parametric change to the Hamiltonian) is unity. This result holds for thermostatted dynamics where the dissipative field is time dependent. For the special case of processes where the internal Hamiltonian does not change and the field is time independent, this equation could easily be deduced for adiabatic systems from the result of Yamada and Kawasaki [6] and for thermostatted systems was derived by Morriss and Evans [7].

We now treat  $t'$  as a dummy variable and thus replace it with  $t$  and then using Eq. (18) we may express the change in work as

$$\Delta w_t = \Delta w_t^{irr} + \int_0^t ds \dot{\lambda}(s) \frac{\partial A(\lambda(s))}{\partial \lambda}, \quad (19)$$

and then, using Eq. (17), recalling that the free energy is not a phase variable and thus may be taken through the average, we arrive at

$$\langle \exp(-\beta \Delta w_t) \rangle = \exp(-\beta \Delta A), \quad (20)$$

which is a generalized JE. The first such equality was written down by Jarzynski [11,17] in 1997. This generalization includes contributions to the work from the dissipative flux. This result shows the close relationship between the JE and the KDF.

The second law can be obtained from this result [17]. We provide a simple proof of this by noting the obvious mathematical identity  $\exp(x) \geq 1+x$  and that in turn  $\langle \exp(x - \langle x \rangle) \rangle \geq 1$  or equivalently that  $\langle \exp(x) \rangle \geq \exp(\langle x \rangle)$ . So from Eq. (20) we have  $\exp(-\beta \langle \Delta w_t \rangle) \leq \exp(-\beta \Delta A)$  and upon taking the logarithm of both sides we obtain

$$\langle \Delta w_t \rangle \geq \Delta A, \quad (21)$$

which is the generalized second law inequality.

#### IV. GIBBS ENTROPY

We noted above that Eq. (11) was first derived by minimizing the fine grained Gibbs entropy of a system subject to constraints. We can also do the converse: we can calculate the time-dependent Gibbs entropy directly from that equation. Since the Gibbs entropy is defined as

$$S(t) \equiv -k_B \int d\mathbf{\Gamma} f(\mathbf{\Gamma}, t) \ln[f(\mathbf{\Gamma}, t)]. \quad (22)$$

Substituting Eq. (11) for the distribution function gives

$$S(t) = -k_B \int d\mathbf{\Gamma} f(\mathbf{\Gamma}, t) \left\{ -\beta H(\mathbf{\Gamma}, \lambda_0) + \beta A(\lambda_0) - \beta \int_0^t ds \mathbf{J}(\mathbf{\Gamma}(-s)) V \cdot \mathbf{F}_e(t-s) \right\}. \quad (23)$$

Unrolling the time dependence of the distribution function using the Schrodinger Heisenberg equivalence gives

$$S(t) = -k_B \int d\mathbf{\Gamma} f(\mathbf{\Gamma}, 0) \left\{ -\beta H(\mathbf{\Gamma}(t), \lambda_0) + \beta A(\lambda_0) - \beta \int_0^t ds \mathbf{J}(\mathbf{\Gamma}(t-s)) V \cdot \mathbf{F}_e(t-s) \right\}. \quad (24)$$

Evaluating the averages and changing variables gives

$$TS(t) = U(t) - A(\lambda_0) + \int_0^t ds \langle \mathbf{J}(\mathbf{\Gamma}(s)) V \cdot \mathbf{F}_e(s) \rangle_{\lambda_0}. \quad (25)$$

Rearranging gives

$$TS(t) = TS(0) + \Delta U + \int_0^t ds \langle \mathbf{J}(\mathbf{\Gamma}(s)) V \cdot \mathbf{F}_e(s) \rangle_{\lambda_0}. \quad (26)$$

Using Eq. (6),

$$\Delta S(t) = \int_0^t ds \langle \dot{Q}(\mathbf{\Gamma}(s)) \rangle_{\lambda_0} / T. \quad (27)$$

This remarkable equation is obviously consistent with the classical thermodynamic expression for the entropy change in a quasistatic process. However, this equation (27) is valid for fast processes. As always, it is important to remember that the temperature appearing in Eq. (27) is the equilibrium temperature of the large thermal reservoir with which the system of interest is in contact. The entropy appearing in Eq. (27) is the fine grained Gibbs entropy Eq. (22), which diverges to negative infinity as a nonequilibrium steady state is approached for deterministic classical systems. The Gibbs entropy is extremely sensitive to the distribution function which is why it is a constant of the motion for Hamiltonian systems [20]. For perfectly time reversible deterministic dynamics there is no loss of information as time increases. The trajectory can always be reversed to its particular original phase and the Gibbs entropy is constant. The Gibbs entropy and the thermodynamic entropy are only equivalent at equilibrium and of course the thermodynamic entropy is only defined at equilibrium.

If one attempts to compute the thermodynamic entropy change in the system of interest by monitoring the heat exchange and the temperature of the system of interest for systems approaching the quasistatic limit [21], because (appealing to thermodynamics) the average temperature of the

system of interest is on average higher than the reservoir temperature you will always underestimate the change in the entropy

$$\Delta S(t) \geq \int_0^t ds \langle \dot{Q}(\Gamma(s)) / T_{st}(s) \rangle_0. \quad (28)$$

Lastly, Eq. (27) implies something else that is quite remarkable. The decrease in the fine grained Gibbs entropy of the system (system of interest plus reservoir) is precisely equal and opposite to the increase in the *thermodynamic* entropy of the large quasistatic heat reservoir. For details about our treatment resulting in a quasistatic heat reservoir see Ref. [12].

We have not considered the effect of including the parametric  $\lambda$  parameter here. Such a generalization is straightforward and will not add any new physics to the above discussion.

### V. EXAMPLE: MOLECULAR DYNAMICS WITH AN OSCILLATING FIELD

We consider the case of color conductivity for two particles. We use the equations of motion, Eq. (1), with the field  $\mathbf{F}_e$  in the direction of the  $x$  Cartesian axis and the second rank tensors set to  $\mathbf{C}_i(\Gamma) = \mathbf{0}$  and  $\mathbf{D}_i(\Gamma) = (-1)^i \mathbf{I}$ , where  $\mathbf{I}$  is the identity matrix. For a fluid this results in a purely dissipative process. The underlying thermodynamic potential does not change and the Hamiltonian has no parametric  $\lambda$  dependence. The relevant distribution function is given by Eq. (11).

Molecular dynamics simulations were then computed using a fourth order Runge-Kutta algorithm. A Weeks-Chandler-Andersen (WCA) potential for the pair interaction between the particles is used,  $U(\Delta r_{ij}) = 4\epsilon [(\sigma/\Delta r_{ij})^{12} - (\sigma/\Delta r_{ij})^6] + \epsilon \forall \Delta r_{ij} < \sigma\sqrt{2}$ ,  $U(\Delta r_{ij}) = 0 \forall \Delta r_{ij} \geq \sigma\sqrt{2}$ . The length unit is  $\sigma$ , the energy unit is  $\epsilon$ , and the time unit is  $\sigma\sqrt{m}/\epsilon$ . The number density is  $\rho = 0.39685$ , the temperature is such that  $k_B T = \epsilon$ , and the time step is  $dt = 0.002$ . The magnitude of the field, starting at time  $s=0$ , evolves by the protocol  $F_e(s) = F_0 \sin(\omega s)$  with  $F_0 = 1$ ,  $\omega = 2\pi$ , and duration  $t = 5$ .

We consider how the pressure  $P$  changes due to the non-equilibrium process as a function of time. This is an intrinsically nonlinear response. To linear order there is no field induced change to the pressure:  $\Delta P = \mathcal{O}(F_0^2)$ . The pressure is obtained by taking the average of

$$P(\Gamma) = \frac{1}{DV(t)} \left( \frac{\sum_{i=1}^N \mathbf{p}_i \cdot \mathbf{p}_i}{m_i} + \sum_{i=1}^N \mathbf{F}_i \cdot \mathbf{q}_i \right), \quad (29)$$

where  $D=2$  is the Cartesian dimension. The ensemble averaged pressure was obtained by computing ensembles of  $5 \times 10^7$  trajectories. To do this a long equilibrium trajectory was computed and for each interval of duration  $\Delta t = 5$  a non-equilibrium transient trajectory was spawned.

In Fig. 1(a) the results from the simulations for the average pressure as a function of time are shown. The directly

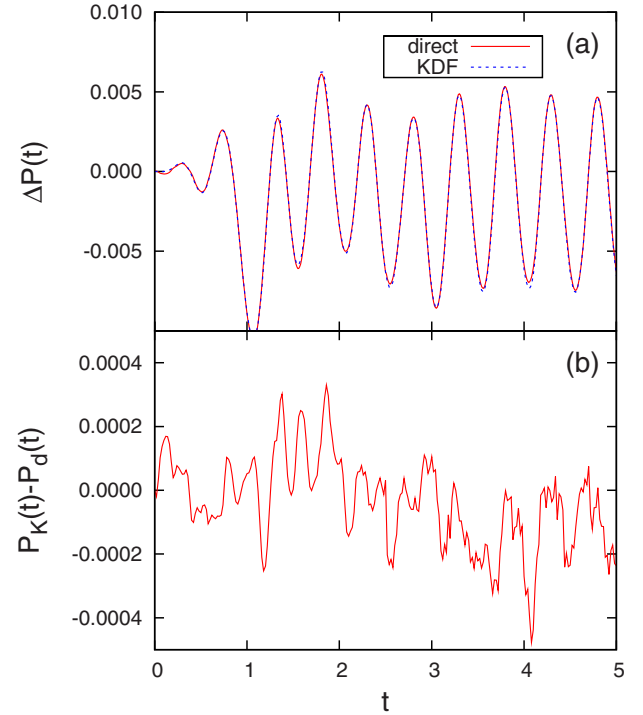


FIG. 1. (Color online) (a) The difference in the average pressure,  $P(t)$ , relative to the equilibrium value  $P_0$ ,  $\Delta P(t) = P(t) - P_0$ , as a function of time for both direct and KDF averaging. The field  $F_e = F_0 \sin(\omega t)$ ,  $\omega = 2\pi$ , is switched on at time  $t=0$ . Due to the very good agreement between the two curves, the differences are very difficult to distinguish on this graph. (b) The difference between the average pressure obtained from the KDF,  $P_K(t)$ , Eq. (11) and the direct ensemble average,  $P_d(t)$ , as a function of time.

averaged results were obtained by ensemble averaging the pressure, Eq. (29), at various discrete times, over the ensemble of transient trajectories. This is compared to the values obtained for the pressure by use of Eq. (11). To use this equation the pressure of the initial configuration, which the transient trajectories were spawned from, was computed. For each of these configurations the integral  $\int_0^t ds \mathbf{J}(\Gamma(-s)) \cdot \mathbf{F}_e(t-s)$ , which appears in Eq. (11), was then computed with  $s$  denoting the time along the transient trajectory. This required different trajectories to be computed for different times,  $t$ , as the initial phase angle of the field,  $F_e(t-s) = F_0 \sin[\omega(t-s)]$  with  $s=0$  appearing in the above integral, explicitly depends upon this. This need for different trajectories may seem a little surprising but recall that we sample the equilibrium distribution at time  $t$  and then use the spawned trajectories to account for the difference between this and the desired nonequilibrium distribution at time  $t$ . For this reason the spawned trajectories  $\Gamma(-s)$ , starting from the same initial point  $\Gamma(0)$ , which evolve by the field  $\mathbf{F}_e(t-s)$  will depend explicitly upon the value of  $t$ . Using this integral with the equilibrium distribution, which is directly sampled by the equilibrium trajectory, allows us to obtain the KDF average.

If one looks closely at the peaks in the oscillations [Fig. 1(a)] small differences can be discerned due to finite statistical sampling. These differences can be seen more clearly in Fig. 1(b) where the difference between the two methods is

plotted directly. The difference between the two methods was noted to decrease with the number of trajectories in a manner that appeared to roughly scale as  $1/\sqrt{N_t}$ , where  $N_t$  is the number of trajectories. It can be concluded that the agreement between the directly averaged results and those obtained using the KDF, Eq. (11), is very good. Given the irregular form of the pressure response curves and the fact that the pressure response is entirely nonlinear it is exceedingly unlikely that this level of agreement is accidental. This confirms the correctness of the theoretical analysis given above.

## VI. CONCLUSIONS

We have provided a generalization of the nonlinear response theory that applies to both time-dependent dissipative and state changing processes. This generalization is valid for systems being driven arbitrarily far from equilibrium by some external agent. It is assumed that the thermal reservoir which regulates the system is held to a fixed temperature. Our model for this reservoir contains unphysical terms in the equations of motion. However, these particles can be removed arbitrarily far from the system of interest so the mathematical details of how the heat is ultimately (and only on average) removed from the system of interest has no physical significance.

The close relationship between this generalized KDF and the Jarzynski equality has been shown by using the KDF to prove the Jarzynski equality. An example of the time-dependent nonlinear response to an external field, using molecular dynamics simulations, has been presented. This showed excellent agreement for the change in the pressure between results obtained directly from the simulations and those computed using the KDF. The generalized KDF has also been used to obtain a form of linear response theory which can be applied to state changing processes.

We need to make a few further comments about our definition of the work. Using Eq. (6) we see that our definition of the work, Eq. (15), gives the total change in the energy minus that part of the change due to the heat.

$$\begin{aligned} \Delta w_t &\equiv H(\mathbf{\Gamma}(t), \lambda(t)) - H(\mathbf{\Gamma}(0), \lambda(0)) - \int_0^t ds \dot{Q}(\mathbf{\Gamma}(s)) \\ &= \int_0^t ds \dot{\lambda}(s) \frac{\partial H(\mathbf{\Gamma}(s), \lambda(s))}{\partial \lambda} - \int_0^t ds \mathbf{J}(\mathbf{\Gamma}(s)) \mathbf{V} \cdot \mathbf{F}_e(s). \end{aligned} \quad (30)$$

Thus our definition is consistent with the first law definition of work. A much more general definition was given by Reid *et al.* [22]. They defined a generalized dimensionless “work”  $\Delta X_t$ , in terms of probabilities of observing infinitesimal initial and final volume elements  $d\mathbf{\Gamma}$ ,  $d\mathbf{\Gamma}(t)$ , along a phase space trajectory together with ratios of initial and final partition functions  $Z^{(1)}$ ,  $Z^{(2)}$ ,

$$\exp[\Delta X_t(\mathbf{\Gamma})] \equiv \frac{P_F(d\mathbf{\Gamma})Z^{(1)}}{P_R(d\mathbf{\Gamma}(\tau))Z^{(2)}} = \frac{f_{1,eq}(\mathbf{\Gamma}) \|\partial\mathbf{\Gamma}/\partial\mathbf{\Gamma}(\tau)\| Z^{(1)}}{f_{2,eq}(\mathbf{\Gamma}(\tau)) Z^{(2)}}. \quad (31)$$

In this equation  $f_{2,eq}(\mathbf{\Gamma}(\tau))$  is the probability of observing the

phase space trajectory end point in the *equilibrium* ensemble of the end-point Hamiltonian. This definition of the generalized variable applies to quite arbitrary ensembles and dynamics (including stochastic dynamics) and gives a trivial derivation of the generalized nonequilibrium free-energy relation as follows:

$$\begin{aligned} \langle \exp(-\Delta X_t) \rangle_{1,eq} &= \int d\mathbf{\Gamma} f_{1,eq}(\mathbf{\Gamma}) \frac{f_{2,eq}(\mathbf{\Gamma}(\tau)) \|\partial\mathbf{\Gamma}(\tau)/\partial\mathbf{\Gamma}\| z^{(2)}}{f_{1,eq}(\mathbf{\Gamma}) z^{(1)}} = \frac{Z^{(2)}}{Z^{(1)}}. \end{aligned} \quad (32)$$

For microcanonical ensembles with constrained total energy this generalized work is in fact related to the change in the heat divided by the reservoir temperature. In the isothermal isobaric ensemble the generalized work is the change in the enthalpy minus the change due to the thermostat, divided by the thermostat temperature. So we have to be very careful about using the term generalized work. In Eq. (32) the equation only refers to initial and final phase space vectors and to equilibrium distribution functions and the corresponding partition functions for the end-point equilibrated states. Arbitrary protocols and time dependencies can be used to arrive at these phase points and (eventually) equilibrated states.

In the special case where both ensembles are canonical and the systems 1,2 are in contact with a heat bath maintained at a fixed temperature it is trivial to see using Eq. (30) that the generalized variable coincides with the work, Eq. (30) as follows:

$$\Delta X_t = \beta \Delta w_t, \quad (33)$$

and Eq. (32) reduces to the Jarzynski equality, Eq. (20).

Recently there has been some interest in what has become known as the maximum entropy production (MEP) approach [23,24] to nonlinear dynamical systems. This approach asserts that nonequilibrium systems arrange themselves in a way that maximizes the rate at which entropy is produced—subject to a set of constraints. As Hoover pointed out in 1986 [25], the problem with these theories is that there is no objective way in which to comprehensively list the set of such constraints. Combining Eq. (9) with the approach reported in [16], shows that the nonequilibrium distribution function for a dissipative system does indeed minimize the fine grained entropy of the nonequilibrium system. Subject to the constraints placed on the system (the distribution is normalized, the initial average energy is fixed, and the average dissipative flux is constrained at all times and is a *continuous* function of time) the exact distribution function, Eq. (11), minimizes the Gibbs entropy for all times including the transient behavior of the system. This is in spite of the fact that for a thermostatted deterministic steady state the entropy diverges towards negative infinity as time increases [26]. This is a consequence of the fact that the steady state distribution (which is only approached and never actually reached) is a strange attractor of lower dimension than the ostensible dimension of phase space [27,28]. It can be seen in [16] that at any finite time, assuming the entropy to be a minimum (subject to the correct constraints of course) results in the same expression, Eq. (11), which is rigorously derived here. In the steady state

the average rate of change of the entropy is constant. Thus the rate of decrease of the Gibbs entropy is maximal. However, the approach in [16] is far more complex than the MEP approximations. MEP only employs sets of Lagrange multipliers for the corresponding sets of constraints. In order to achieve temporal continuity in the average dissipative flux, [16] employs a Lagrange multiplier *functional* (i.e., the set of constraints becomes infinite). This functional becomes a memory function that conditions what happens in the future by what has happened in the past, ensuring that the dissipative flux is a continuous function of time.

If we take the logarithm of Eq. (11) and assume the initial ensemble is canonical we see that

$$\ln[f(\Gamma, t)] = \beta A(\lambda_0) - \beta H(\Gamma, \lambda_0) - \beta \int_0^t ds \mathbf{J}(\Gamma(-s))V \cdot \mathbf{F}_e(t-s). \quad (34)$$

Since the logarithm is a monotonic increasing function of its argument we see that the probability of observing a phase  $\Gamma$  at time  $t$  is increased if the dissipation integrated along the phase space trajectory that terminates at  $\Gamma$  at time  $t$  is large and positive. What is important here is that the probability is

influenced by the path integral and not just the instantaneous value. In certain time-dependent systems (e.g., viscoelastic systems) the instantaneous entropy production is not always positive. However, the second law inequality [29] guarantees that the time integral is positive and subject to the constraints is actually maximal.

We also see that the probability is increased if the value of the Hamiltonian at the current phase point, namely  $H(\Gamma, \lambda_0)$ , is also low. In the long time limit we expect that the integrated dissipation will dominate over the Hamiltonian term. This is probably why MEP provides a reasonable approximation at long times in some circumstances [23]. We note as always that the reciprocal temperature  $\beta$  is not directly related to the temperature of the system of interest (this temperature may not be well defined far from equilibrium) but rather to that of the large, effectively equilibrium, heat bath to which the heat eventually dissipates.

#### ACKNOWLEDGMENTS

We thank the Australian Research Council for supporting this work and the Australian Partnership for Advanced Computing (APAC) for computational facilities.

- 
- [1] I. Borzsa, P. T. Cummings, and D. J. Evans, *Mol. Phys.* **100**, 2735 (2002).
- [2] B. D. Todd, *Phys. Rev. E* **56**, 6723 (1997).
- [3] J. Delhommelle, P. T. Cummings, and J. Petravac, *J. Chem. Phys.* **123**, 114505 (2005).
- [4] J. Delhommelle and P. T. Cummings, *Phys. Rev. B* **72**, 172201 (2005).
- [5] G. Pan and C. McCabe, *J. Chem. Phys.* **125**, 194527 (2006).
- [6] T. Yamada and K. Kawasaki, *Prog. Theor. Phys.* **38**, 1031 (1967).
- [7] G. P. Morriss and D. J. Evans, *Mol. Phys.* **54**, 629 (1985).
- [8] D. J. Evans and G. P. Morriss, *Mol. Phys.* **64**, 521 (1988).
- [9] J. Petravac and D. J. Evans, *Phys. Rev. E* **58**, 2624 (1998).
- [10] J. Olsen and P. Jørgensen, *J. Chem. Phys.* **82**, 3235 (1985).
- [11] C. Jarzynski, *Phys. Rev. Lett.* **78**, 2690 (1997).
- [12] S. R. Williams, D. J. Searles, and D. J. Evans, *Phys. Rev. E* **70**, 066113 (2004).
- [13] S. R. Williams, D. J. Searles, and D. J. Evans, *Mol. Phys.* **105**, 1059 (2007).
- [14] D. J. Evans and G. P. Morriss, *Statistical Mechanics of Non-equilibrium Liquids*, 2nd ed. (Cambridge University Press, Cambridge, 2008).
- [15] For this phase shift to work requires that  $\Gamma' = \Gamma(t)$  spans the entire phase space.
- [16] D. J. Evans, *Phys. Rev. A* **32**, 2923 (1985).
- [17] C. Jarzynski, *Phys. Rev. E* **56**, 5018 (1997).
- [18] L. Rondoni and C. Mejia-Monasterio, *Nonlinearity* **20**, R1 (2007).
- [19] E. M. Sevick, R. Prabhakar, S. R. Williams, and D. J. Searles, *Annu. Rev. Phys. Chem.* **59**, 603 (2008).
- [20] O. Penrose, *Ann. N.Y. Acad. Sci.* **373**, 211 (1981).
- [21] Here we assume that we are close enough to equilibrium for the postulate of local thermodynamic equilibrium to hold.
- [22] J. C. Reid, E. M. Sevick, and D. J. Evans, *Europhys. Lett.* **72**, 726 (2005).
- [23] R. Lorenz, *Science* **299**, 837 (2003).
- [24] *Non-Equilibrium Thermodynamics and the Production of Entropy: Life, Earth, and Beyond*, edited by A. Kleidon and L. R. D (Springer, Berlin/Heidelberg, 2005).
- [25] W. G. Hoover, *J. Stat. Phys.* **42**, 587 (1986).
- [26] D. J. Evans and L. Rondoni, *J. Stat. Phys.* **109**, 895 (2002).
- [27] R. Klages, *Microscopic Chaos, Fractals and Transport in Non-equilibrium Statistical Mechanics* (World Scientific, Singapore, 2007).
- [28] D. J. Evans, W. G. D. Cohen, D. J. Searles, and F. Bonetto, *J. Stat. Phys.* **101**, 17 (2000).
- [29] S. R. Williams, D. J. Evans, and E. Mittag, *C. R. Phys.* **8**, 620 (2007).