THE JOURNAL OF CHEMICAL PHYSICS **134**, 204113 (2011)

A proof of Clausius' theorem for time reversible deterministic microscopic dynamics

Denis J. Evans,[1] Stephen R. Williams,[1,a)] and Debra J. Searles[2]
[1]*Research School of Chemistry, Australian National University, Canberra, ACT 0200, Australia*
[2]*Queensland Micro- and Nanotechnology Centre and School of Biomolecular and Physical Sciences, Griffith University, Brisbane, Qld 4111, Australia*

(Received 17 February 2011; accepted 28 April 2011; published online 27 May 2011)

In 1854 Clausius proved the famous theorem that bears his name by *assuming* the second "law" of thermodynamics. In the present paper we give a proof that requires no such assumption. Our proof rests on the laws of mechanics, a T-mixing property, an ergodic consistency condition, and on the axiom of causality. Our result relies on some recently derived theorems, such as the Evans-Searles and the Crooks fluctuation theorems and the recently discovered relaxation and dissipation theorems. © *2011 American Institute of Physics*. [doi:10.1063/1.3592531]

I. INTRODUCTION

Recently, we derived some generalizations[1] of the nonequilibrium work relations due to Crooks[2] that are valid for thermal rather than mechanical changes in time reversible, deterministic systems of interacting particles. Here, we extend this work to study cyclic integrals of the generalized work and heat, and in the process we give a proof of Clausius' inequality. This proof is derived without assuming the second law of thermodynamics. It also clarifies the meaning of the "temperature" that appears in Clausius' *inequality* and shows that the equality and the inequality only apply to *periodic cyclic* processes. Clausius' inequality does not apply to aperiodic cycles.

Clausius' inequality states that *if* a system responds periodically to a cyclic thermal process, then

$$\lim_{N\to\infty}\lim_{t\to\infty}\oint_P \frac{dQ^{therm}}{T} \geq 0, \tag{1}$$

where Q^{therm} is the heat transferred to the reservoir from the system, N is the number of particles in the system, and t is the time over which the process occurs. In the past, there has been some question as to what temperature, T, should be used in this relation (especially for the strict inequality), and indeed the question of how to define temperature, in general, has been an open question (see, Ref. 3). In this paper we provide a proof of Eq. (1) which also clarifies exactly what this temperature refers to.

In Secs. II and III, we present generalized forms of the Crooks fluctuation theorem and the Jarzynski equality (JE), respectively. Section IV presents the minimum average generalized work relationship, a corollary of the Jarzynski equality, and discusses its application to cyclic integrals of the generalized work. These three relationships provide the formal results required for the derivation of Clausius' inequality, later in the paper. In Section V, we introduce a framework whereby we can apply the generalized Crooks fluctuation theorem and the Jarzynski equality to thermal processes. This then allows us in Sec. VI, to derive Clausius' inequality for the heat, and to clarify the meaning of "temperature" that appears in this inequality. Finally, in the Sec. VII, we discuss other implications of these findings. In particular, we point out that unlike dissipation, entropy does not appear to be a useful concept for deterministic nonequilibrium systems.

II. GENERALIZED CROOKS FLUCTUATION THEOREM (GCFT)

We consider two closed N-particle systems: 1, 2. These systems may have the same or different Hamiltonians, temperatures, or volumes; it does not matter. Nor does the ensemble matter: microcanonical, canonical, or isothermal isobaric. A protocol, and the corresponding time-dependent dynamics, is then defined that will eventually transform equilibrium system 1 into equilibrium system 2. The systems are distinguished by introducing a parameter, λ, which takes on a value λ_1 in system 1 and λ_2 in system 2, and the transformation is also parameterised through $\lambda(t)$ with $\lambda(0) = \lambda_1$ and $\lambda(\tau) = \lambda_2$. The equations of motion are therefore nonautonomous (i.e., they depend *explicitly* on absolute time).

Following Ref. 4, we define a *generalized dimensionless "work,"* $\Delta X_\tau(\mathbf{\Gamma})$, for a trajectory of duration τ, originating from the phase point $\mathbf{\Gamma}(0)$,

$$\begin{aligned}\exp[\Delta X_\tau(\mathbf{\Gamma}(0))] &= \lim_{d\mathbf{\Gamma}(0)\to 0} \frac{p_{eq,1}(d\mathbf{\Gamma}(0))Z(\lambda_1)}{p_{eq,2}(d\mathbf{\Gamma}(\tau))Z(\lambda_2)}\\ &\equiv \frac{f_{eq,1}(\mathbf{\Gamma}(0))d\mathbf{\Gamma}(0)Z(\lambda_1)}{f_{eq,2}(\mathbf{\Gamma}(\tau))d\mathbf{\Gamma}(\tau)Z(\lambda_2)},\quad \forall\mathbf{\Gamma}(0)\in D_1,\end{aligned} \tag{2}$$

where $Z(\lambda_i)$ is the partition function for the system $i = 1, 2$, D_1 is the accessible phase space domain for system 1 (e.g., coordinates in a fixed special range $(-L, +L)$ and momenta range $(-\infty, +\infty)$), and $f_{eq,i}(\mathbf{\Gamma})$ is the equilibrium phase space distribution for system i. Without loss of generality we assume that both equilibrium distributions functions are even

---

[a)] Author to whom correspondence should be addressed. Electronic mail: swilliams@rsc.anu.edu.au.

functions of the momentum. This simply means that we are not moving relative to either system. In Eq. (2), $d\mathbf{\Gamma}(0)$ is an infinitesimal phase space volume centred on $\mathbf{\Gamma}(0)$.

If the system is not canonical the *partition function* Z_i is just the normalization factor for the equilibrium distribution function and $f_{eq,i}(\mathbf{\Gamma}) = \exp[F_i(\mathbf{\Gamma})]/Z_i$, where $F_i(\mathbf{\Gamma})$ is some real single valued phase function.

Although the physical significance of the generalized work, X, might seem obscure at this point, we will show that for particular choices of dynamics and ensemble, it is related to important thermodynamic properties and when it is evaluated along quasi-static paths it is, in fact, a path independent state function.

Before proceeding further with the analysis, it is useful to consider precisely what the generalized work is dependent upon. First, it is a function of the *equilibrium* states 1 and 2. This occurs via the equilibrium distributions appearing in Eq. (2) and also the partition functions for those states—see Eq. (2). Second, it is a function of the endpoints of the possibly *nonequilibrium* phase space trajectory that takes phase $\mathbf{\Gamma}(0)$ to $\mathbf{\Gamma}(\tau)$. As we will see later it is also a function of how much heat is gained or lost from the system over the duration of that trajectory. This heat loss determines the ratio of phase space volumes $d\mathbf{\Gamma}(0)/d\mathbf{\Gamma}(\tau)$. Last, it is a function of the duration of the trajectories τ.

The probability of observing ensemble members within the infinitesimal phase volume $d\mathbf{\Gamma}$, centred on the phase vector $\mathbf{\Gamma}$, in the initial equilibrium distribution function, $f_{eq,i}(\mathbf{\Gamma})$ is $p_{eq,i}(d\mathbf{\Gamma}) = f_{eq,i}(\mathbf{\Gamma})d\mathbf{\Gamma}$.

It is very important to note that the time τ is the time at which the parametric change in λ is complete. This means that at time τ this system is *not* necessarily at equilibrium: $f(\mathbf{\Gamma},0) = f_{eq,1}(\mathbf{\Gamma})$, but $f(\mathbf{\Gamma},\tau) \neq f_{eq,2}(\mathbf{\Gamma})$. The generalized work is defined with respect to two different *equilibrium* distributions and the end points of *finite time* phase space trajectories: $\mathbf{\Gamma}(s):\ 0 \le s \le \tau$. The recently proved relaxation theorem[5] says that if the system is T-mixing and if the initial equilibrium distribution is an even function of the momenta, then $\lim_{t\to\infty} f(\mathbf{\Gamma},t) = f_{eq,2}(\mathbf{\Gamma})$.

In order for $\Delta X_\tau(\mathbf{\Gamma})$ to be well defined, $\forall \mathbf{\Gamma}(0) \in D_1$, then $\mathbf{\Gamma}(\tau) \in D_2$ and both $f_{eq,1}(\mathbf{\Gamma}(0)) \neq 0$ and $f_{eq,2}(\mathbf{\Gamma}(\tau)) \neq 0$. This is known as the *ergodic consistency for the generalized work*.

We identify $\|\partial\mathbf{\Gamma}(\tau)/\partial\mathbf{\Gamma}(0)\|$ as the Jacobian determinant and note that

$$\left\|\frac{\partial\mathbf{\Gamma}(\tau)}{\partial\mathbf{\Gamma}(0)}\right\| = \frac{d\mathbf{\Gamma}(\tau)}{d\mathbf{\Gamma}(0)}. \tag{3}$$

The GCFT considers the probability, $p_{eq,f}(\Delta X_t = B \pm dB)$, of observing values of $\Delta X_t = B \pm dB$ for forward trajectories starting from the initial equilibrium distribution 1, $f_1(\mathbf{\Gamma},0) = f_{eq,1}(\mathbf{\Gamma})$, and the probability, $p_{eq,r}(\Delta X_t = -B \mp dB)$, of observing $\Delta X_t = -B \pm dB$ for reverse trajectories but starting from the equilibrium given by $f_{eq,2}(\mathbf{\Gamma})$, system 2.

Consider two equilibrium ensembles from which initial trajectories can be selected with known equilibrium distributions: $f_{eq,1}(\mathbf{\Gamma})$ and $f_{eq,2}(\mathbf{\Gamma})$. If initially we select phases from $f_{eq,1}(\mathbf{\Gamma})$, employ a particular protocol (*f*) and corresponding time-dependent dynamics, defined by a parameter $\lambda_f(s)$ with $\lambda_f(0) = \lambda_1$ and $\lambda_f(\tau) = \lambda_2$, then the probability that the phase variable defined in Eq. (2) takes on the value $B \pm dB$ is given by

$$p_{eq,1}(\Delta X_{\tau,f} = B \pm dB) = \int_{\Delta X_{\tau,f} = B \pm dB} d\mathbf{\Gamma}\ f_{eq,1}(\mathbf{\Gamma}). \tag{4}$$

If initially we select phases from $f_{eq,2}(\mathbf{\Gamma})$ with a particular protocol (r) which is the time-reverse of (*f*), $\lambda_r(s) = \lambda_f(\tau - s)$, and corresponding time-dependent dynamics, so $\lambda_r(0) = \lambda_2$ and $\lambda_r(\tau) = \lambda_1$, then the probability that the phase variable defined in Eq. (2) takes on the value $-B$ is given by $p_{eq,2}(\Delta X_{\tau,f} = -B \mp dB) = \int_{\Delta X_{\tau,f} = -B \mp dB} d\mathbf{\Gamma}\ f_{eq,2}(\mathbf{\Gamma})$.

We note that a trajectory starting at point $\mathbf{\Gamma}$, and evolving forward in time with the forward protocol to the point $\mathbf{\Gamma}(\tau)$ will be related by a time reversal mapping to a trajectory starting at $M^T\mathbf{\Gamma}(\tau)$ and evolving with the time-reverse protocol. If $S^\tau_{f/r}$ is the time evolution operator with forward/reverse protocol,

$$M^T S^\tau_r M^T S^\tau_f \mathbf{\Gamma}(0) = \mathbf{\Gamma}(0). \tag{5}$$

Also, if $\Delta X_{\tau,f}(\mathbf{\Gamma}(0)) = B$, then $\Delta X_{\tau,f}(\mathbf{\Gamma}(0)) = -B$. As was shown in Ref. 1, we can therefore write

$$\frac{p_{eq,1}(\Delta X_{\tau,f} = B \pm dB)}{p_{eq,2}(\Delta X_{\tau,r} = -B \mp dB)} = \exp[B]\frac{Z(\lambda_2)}{Z(\lambda_1)}. \tag{6}$$

We note that at time τ, the system that is evolving from $f_{eq,1}(\mathbf{\Gamma})$ would not have relaxed to the distribution $f_{eq,2}(\mathbf{\Gamma})$ (or vice versa). We can calculate the generalized work *after* the parameter change has ceased but during the period in which the system relaxes, as specified by the relaxation theorem,[5] to the new equilibrium. From Eq. (2), we see that

$$\begin{aligned}
&\exp[\Delta X_{\tau+s}(\mathbf{\Gamma}) - \Delta X_\tau(\mathbf{\Gamma})] \\
&= \frac{f_{eq,1}(\mathbf{\Gamma}(0))d\mathbf{\Gamma}(0)Z(\lambda_1)}{f_{eq,2}(\mathbf{\Gamma}(\tau+s))d\mathbf{\Gamma}(\tau+s)Z(\lambda_2)}\ \frac{f_{eq,2}(\mathbf{\Gamma}(\tau))d\mathbf{\Gamma}(\tau)Z(\lambda_2)}{f_{eq,1}(\mathbf{\Gamma}(0))d\mathbf{\Gamma}(0)Z(\lambda_1)} \\
&= \frac{f_{eq,2}(\mathbf{\Gamma}(\tau))d\mathbf{\Gamma}(\tau)}{f_{eq,2}(\mathbf{\Gamma}(\tau+s))d\mathbf{\Gamma}(\tau+s)}, \quad \forall s > 0.
\end{aligned} \tag{7}$$

If we look at the last line of Eq. (7), we recognize that it is simply the integrated dissipation function, $\Omega_{eq,2,s}(\mathbf{\Gamma}(\tau))$ defined in the Evans-Searles fluctuation theorem,[6] for *equilibrium* system 2, evaluated at a phase $\mathbf{\Gamma}(\tau)$, and integrated for a time s. It is important to note that both the numerator and the denominator of Eq. (7) involve *forward* time integrations from system 2 equilibrium [i.e., there is no forward and reverse as in Eq. (2)]. Therefore,

$$\begin{aligned}
&[\Delta X_{\tau+s}(\mathbf{\Gamma}) - \Delta X_\tau(\mathbf{\Gamma})] \\
&\equiv \Delta X_s(\mathbf{\Gamma}(\tau)) = \ln\frac{f_{eq,2}(\mathbf{\Gamma}(\tau))}{f_{eq,2}(\mathbf{\Gamma}(\tau+s))} - \int_0^\tau ds\ \Lambda(\mathbf{\Gamma}(\tau+s)) \\
&= \Omega_{eq,2,s}(\mathbf{\Gamma}(\tau)) = 0,\ \forall\mathbf{\Gamma}(0) \in D_2,\ \forall s \le 0,
\end{aligned} \tag{8}$$

where Λ is the phase space expansion factor:

$$\Lambda \equiv \frac{\partial}{\partial\mathbf{\Gamma}}\cdot\mathbf{\Gamma}. \tag{9}$$

The last line in Eq. (8) is identically zero because the dissipation function $\Omega_{eq}(\mathbf{\Gamma})$ for all equilibrium systems is identically zero,[5] and we know from the relaxation theorem that the system does eventually relax to the unique, ergodic, dissipationless equilibrium state of system 2.

III. GENERALIZED JARZYNSKI EQUALITY

The generalized Jarzynski equality (GJE), can be thought of as the analogue of the nonequilibrium partition identity[6] (NPI) evaluated for the generalized work. We say "analogue" because the introduction of forward and reverse paths in the definition of the generalized work is quite different from the use of forward only paths, for the NPI.

As is the case for NPI, the GJE can be derived trivially from the corresponding fluctuation theorem, Eq. (6) giving

$$\langle \exp[-\Delta X_\tau(\mathbf{\Gamma})]\rangle_{eq,1} = \frac{Z(\lambda_2)}{Z(\lambda_1)}, \tag{10}$$

where the brackets $\langle ...\rangle_{eq,1}$ denote an equilibrium ensemble average over the initial equilibrium distribution. We also note that with the change of variables the domain of integration may change.

The GJE is very widely applicable. It relates the ensemble average of the exponential of *nonequilibrium* path integrals to *equilibrium* thermodynamic free energy differences. The validity of Eq. (10) requires that $\forall\mathbf{\Gamma}(0) \in D_1,\ f_{eq,1}(\mathbf{\Gamma}(0)) \neq 0$, and $f_{eq,2}(\mathbf{\Gamma}(\tau)) \neq 0$, which implies $\forall\mathbf{\Gamma}(\tau) \in D_2,\ f_{eq,2}(\mathbf{\Gamma}(\tau)) \neq 0$. We call this the *ergodic consistency condition for the GJE*.

Equations (6) and (10) are very general, and they even apply to stochastic dynamics (see Ref. 4). Obviously, the paths do not need to be quasi-static paths as in traditional thermodynamics. These equations are independent of the particular protocol, provided the appropriate ergodic consistency holds. In fact, it is possible to average over the initial ensemble *and* a set of protocols, since the final answer is protocol or path independent.

Although the GJE and NPI can be proved from their appropriate fluctuation relation, the reverse is not true because the corresponding fluctuation relations contain more information than either the NPI or the GJE.

As we mentioned earlier, like the CFT, the application of the GJE relies on the observation of improbable fluctuations. In order to yield reliable estimates of free energy differences, one must sample the trajectories that are the conjugate antitrajectories of the most probable trajectories. This means that these formulae are of limited use for computing free energy differences in the thermodynamic, or large system, limit.

IV. MINIMUM AVERAGE GENERALIZED WORK

We now derive a further simple corollary of the GJE. From Eq. (10), we see that

$$\begin{aligned}\frac{Z(\lambda_2)}{Z(\lambda_1)} &= \langle \exp[-\Delta X_\tau]\rangle_{eq,1}\\ &= \exp[-\langle\Delta X_\tau\rangle_{eq,1}]\langle \exp[-\Delta X_\tau + \langle\Delta X_\tau\rangle_{eq,1}]\rangle\\ &\geq \exp[-\langle\Delta X_\tau\rangle_{eq,1}]\langle 1-\Delta X_\tau + \langle\Delta X_\tau\rangle_{eq,1}\rangle_{eq,1}\\ &= \exp[-\langle\Delta X_\tau\rangle_{eq,1}].\end{aligned} \tag{11}$$

In deriving this relation the fact that $e^x \geq 1 + x,\ \forall x \in \mathbb{R}$ is used.[7] Taking the logarithms of both sides and then multiplying both sides by –1, we get

$$\langle \Delta X_\tau\rangle \geq \ln\left[\frac{Z(\lambda_1)}{Z(\lambda_2)}\right]. \tag{12}$$

This is clearly the analogue of the second law inequality[6] for ensembles of systems with changing free energy. Some authors refer to work inequalities like Eq. (12) as Clausius' inequality,[8] however, we reserve that term for *cyclic* inequalities of the *heat*, since as Planck remarked,[9] "this is the form of the second law first enunciated by Clausius."

In actual systems the right hand side will turn out to be a dimensionless free energy difference. For example, if systems 1, 2 are canonical and at the same temperature and have the same number of particles and volume, as we will see later, $\ln[Z_1/Z_2] = \beta\Delta A_{21} = \beta(A_2 - A_1)$ and $\Delta X = \beta\int_0^\tau ds\ W(s)$, where W denotes the work (i.e., the change caused by the internal energy minus the change caused by the heat) and A_i is the Helmholtz free energy of system i. The minimum average work inequality implies in this case $\Delta W_{21} \geq \Delta A_{21}$. The minimum work is expended if the path is reversible or quasi-static, in the case where that work is, in fact, the difference in the Helmholtz free energies divided by $k_B T$. In the quasi-static case there are no fluctuations in the value of the generalized work.

If the parametric protocol takes us around a closed cycle that is defined in terms of the parameter $\lambda(t)$, we see that since by definition $Z_1/Z_2 = 1$,

$$\oint ds\langle \dot{X}(s)\rangle = \oint \langle dX\rangle \geq 0. \tag{13}$$

The ensemble average of the cyclic integral of the generalized work is non-negative. Because the dynamics is microscopically reversible, the cyclic integral can only be zero if the cycle is *thermodynamically* reversible—what we term quasi-static. A pathway is traversed quasi-statically if the average work for a forward path is equal and opposite to the average work for the reversed path.

The cyclic integral of the generalized work for a quasi-static cycle is zero. In fact, this is the definition of a quasi-static process. The fact that the cyclic integral of the generalized work is zero, $\oint dX = 0$, also implies that

$$_{qs}\int_i^f dX = independent\ of\ path, \tag{14}$$

where the subscript *qs* denotes the fact that the integral is for a quasi-static or thermodynamically reversible pathway. The proof of Eq. (14) is obvious. Construct a reversible cycle $i \to f$, $f \to i$. The cyclic integral must be zero, so if we vary the path for the return leg $f \to i$, we must always get the same value for the integrated reversible work, independent of the precise path.

Finally, we can see that if the integral of the generalized work for paths is independent of the pathway, that integral must be a *state function* (i.e., a function only of the initial and final states of the system). In fact, this is why the seemingly abstract generalized work defined in Eq. (2) is so important. The generalized work for a thermodynamically reversible pathway is always a history independent function of the thermodynamic states of the system at the end points of the path.

V. NONEQUILIBRIUM WORK RELATIONS FOR THERMAL PROCESSES

We wish to consider a realistic model of a system that is driven away from equilibrium by a reservoir whose temperature is changing. For this case the simple parametric change in the Hamiltonian or the external field usually employed in the derivation of the JE or the CFT, is not applicable and care is needed in developing the physical mechanisms, see Ref. 1.

Here, we could address this issue by considering a system of interest, containing some very slowly relaxing constituents, such as soft matter or pitch, in contact with a rapidly relaxing reservoir. The reservoir may be formed from a copper block or another highly thermally conductive material. Changing the temperature of the reservoir (say with a thermostatically controlled heat exchanger) then drives the system of interest out of equilibrium. The change in the temperature is slow enough that the reservoir may be treated to high accuracy, as undergoing a quasi-static temperature change. The slowly relaxing system of interest is far from equilibrium. We employ generalized versions of the CFT and the JE to describe this system. Importantly, the quantities that appear in the theory are physically measurable variables.

Another mechanism for achieving the required result would be (following Planck[9]) to have a set of large equilibrium thermostats that can be thermally coupled to the system of interest in a protocol sequence. If these thermostats are large they can be regarded as being in thermal equilibrium. If they are sufficiently remote from the system of interest there is no way the system of interest can "know" the precise mathematical details of how heat is ultimately taken from or added to the system of interest.

For convenience from a theoretical perspective, we choose the Nosé-Hoover thermostating mechanism,[10] and the equations of motion, including the thermostat multiplier, are then

$$\begin{aligned}\dot{\mathbf{q}}_i &= \frac{\mathbf{p}_i}{m},\\ \dot{\mathbf{p}}_i &= \mathbf{F}_i(\mathbf{q}) - S_i(\alpha(\mathbf{\Gamma})\mathbf{p}_i + \boldsymbol{\gamma}_{th}),\\ \dot{\alpha} &= \left(\frac{\sum_{i=1}^{N} S_i \mathbf{p}_i \cdot \mathbf{p}_i/m}{3(N_{th}-1)k_B T(t)} - 1\right)\frac{1}{\tau_\alpha^2},\end{aligned} \tag{15}$$

where τ_α is an arbitrary Nosé-Hoover time constant. The value of $T(t)$ is the target temperature of the thermostat and $S_i = 0, 1$ is a switch that controls which particles are coupled to the Nosé-Hoover thermostat: $\sum_{i=1}^{N} S_i = N_{th}$. In our model the particles that are coupled to the thermostat can be taken to be remote from the system of interest. This ensures that the particles in the system of interest are ignorant of the precise details of this unphysical thermostat. These thermostated particles are also subject to a fluctuating force $\boldsymbol{\gamma}_{th}$ that is chosen to ensure that the total momentum of the thermostated particles $\sum_{i=1}^{N} S_i \mathbf{p}_i \equiv \mathbf{p}_{th} = \mathbf{0}$ is identically zero.

The extended, time-dependent internal energy is $H_E(\mathbf{\Gamma}, \alpha, t) = H_0(\mathbf{\Gamma}) + (3(N_{th}-1)/2)k_B T(t)\alpha^2\tau_\alpha^2$ and the extended phase space of the system is $\mathbf{\Gamma}' = (\mathbf{\Gamma}, \alpha)$. The Liouville equation states[6,10] $df/dt = -\Lambda f$, and using Eq. (15) it is easy to show that $k_B T\Lambda = k_B T\left((\partial/\partial\mathbf{\Gamma})\dot{\mathbf{\Gamma}} + (\partial/\partial\alpha)\dot{\alpha}\right) = -3(N_{th}-1)k_B T\alpha = -\dot{Q}^{therm}$, where $\dot{Q}^{therm}$ is the rate of decrease in H_E due to the thermostat or equivalently the rate of increase of energy by the imaginary external thermostat. From the relaxation theorem, the unique equilibrium distribution function for this system at a fixed temperature T is then[5]

$$f_{eq}(\mathbf{\Gamma}, T, \alpha) = \frac{\tau_\alpha\sqrt{3(N_{th}-1)/(2\pi)}}{Z(T)} \times \exp(-\beta H_E(\mathbf{\Gamma}, T, \alpha))\delta(p_{th}), \tag{16}$$

where $Z(T)$ is the canonical partition function and $\lambda(t) \equiv T(t)$.

We now consider applying the GCFT, Eq. (6), when a thermal rather than a mechanical process occurs. Consider a thermostated system of N particles whose target kinetic temperature is changed from T_1 to T_2 over a period $0 < t < \tau$. We do not change the Hamiltonian during this process. For simplicity we consider a canonical ensemble for the two equilibrium states, Eq. (16), and use the equations of motion, Eq. (15). The temperature dependence of the reservoir is achieved by making the Nosé-Hoover target temperature $T(t)$ in Eq. (15) a time dependent variable.

From Eqs. (15) and (16), we see that the generalized dimensionless work is

$$\Delta X_\tau(\mathbf{\Gamma}'; 0, \tau) = \beta_2 H_E(\mathbf{\Gamma}'(\tau)) - \beta_1 H_E(\mathbf{\Gamma}'(0)) + \int_0^\tau dt\ \beta(t)\dot{Q}^{therm}(\mathbf{\Gamma}'(t)), \tag{17}$$

where $\beta(t) = 1/(k_B T(t))$ is the inverse, time-dependent target temperature. (Note: as noted above $dE = dW - dQ^{therm} = dW + dQ^{soi}$ because, following Planck, our change in the heat refers to the thermal reservoir (*therm*) rather than to the system of interest.) Now if we take the derivative of the extended Hamiltonian while the temperature is changing, but with no other external agent acting on the system, we obtain

$$\frac{d}{dt}H_E(\mathbf{\Gamma}'(t)) = -\dot{Q}^{therm}(\mathbf{\Gamma}'(t)) + \frac{3}{2}(N_{therm}-1)k_B\dot{T}(t)\alpha^2(t)\tau_\alpha^2. \tag{18}$$

We then obtain

$$\frac{d}{dt}[\beta(t)H_E(\mathbf{\Gamma}'(t))] = -\beta(t)\left[H_0(\mathbf{\Gamma}(t))\frac{\dot{T}(t)}{T(t)} + \dot{Q}(\mathbf{\Gamma}'(t))\right], \tag{19}$$

and combining Eqs. (19) and (17), the generalized "power" for a change in the target temperature with time is

$$\dot{X}(\mathbf{\Gamma}(t)) = \dot{\beta}(t)H_0(\mathbf{\Gamma}(t)). \tag{20}$$

Note that the right hand side of Eq. (20) only depends upon physical variables and not the unphysical thermostat multiplier α or the extended Hamiltonian. Equation (10) then becomes

$$\left\langle \exp\left(-\int_0^\tau dt\ \dot{\beta}(t)H_0(\mathbf{\Gamma}(t))\right)\right\rangle_1 = \frac{Z_2}{Z_1} = \exp[-\beta_2 A_2 + \beta_1 A_1]. \tag{21}$$

Since for quasi-static processes the target kinetic temperature of the Nosé-Hoover thermostat is, in fact, equal to the equilibrium thermodynamic temperature, one can see that this equation is consistent with thermodynamics because in the quasi-static limit, equilibrium thermodynamics gives us the relation

$$\begin{aligned}\int_0^\tau dt\ \dot{\beta}(t)U(t) &= \int_0^\tau dt\ \dot{T}(t)\frac{d}{dT}[\beta(t)A(t)]\\ &= \int_0^\tau dt\ \frac{d}{dt}[\beta(t)A(t)]\\ &= \beta_2 A_2 - \beta_1 A_1.\end{aligned} \tag{22}$$

For this case we see that the dimensionless work is the rate of change of dimensionless free energy.

VI. CLAUSIUS' INEQUALITY AND THE THERMODYNAMIC TEMPERATURE

We now turn our attention away from work to heat. As before we consider a periodic protocol. However for the heat (and unlike work), we can only deduce useful results if the system responds periodically to the cyclic protocol. We note that if we periodically cycle a given protocol, not all systems will respond periodically. The necessary and sufficient conditions for the system to respond periodically are not known. Clausius' inequality only applies if, in the long time limit ($t \to \infty$), the average system response is periodic.

If we now substitute Eq. (17) into Eq. (13) and apply it to a *periodic cycle* after any cyclic transients have decayed, we can deduce that

$$\begin{aligned}\lim_{t\to\infty}\oint_P ds\ \langle \dot{X}(t+s)\rangle &= \lim_{t\to\infty}\oint_P ds\ \langle \beta(t+s)\dot{Q}^{therm}(t+s)\rangle\\ &= \lim_{t\to\infty}\oint_P \left\langle \frac{dQ^{therm}}{k_B T}\right\rangle \geq 0,\end{aligned} \tag{23}$$

where we use the notation: $\oint_P ds$ to denote the cyclic integral of a *periodic* function. We note that the cycle is driven by a periodic parametric change. In this equation t is the time when you start the cyclic integral. Because the cycle is periodic in Eq. (17), the change in βH_E around the cycle is identically zero. Thus for periodic cycles, Eq. (13) reduces to an integral of the heat divided by the target temperature.

In more usual notation Eq. (23) implies that in the large system limit, $N \to \infty$, where fluctuations are negligible, we obtain the very well-known Clausius' inequality,[9] given in Eq. (1), that is, $\lim_{N\to\infty}\lim_{t\to\infty}\oint_P \frac{dQ^{therm}}{T} \geq 0$.

In these equations, Eqs. (1) and (23), the time dependent temperature is the *target temperature* of the Nosé-Hoover thermostat. At any instant the numerical value of the target temperature is, in fact, the equilibrium thermodynamic temperature that the entire system would relax to, if at that same moment, this parameter was fixed at its current value. We know that this is so from the relaxation theorem for T-mixing systems. We will often use the phrase that the temperature appearing in Eqs. (1) and (23) is at any instant of time, the equilibrium thermodynamic temperature of the *underlying* equilibrium state.

If the cycle is traversed rapidly so that the system of interest is not in thermodynamic equilibrium, the actual thermodynamic temperature of the system of interest is of course not defined.

If the thermostat is composed of a large Hamiltonian region coupled to the system of interest and a remote Nosé-Hoover thermostated region, we can argue that the precise details of the thermostat cannot possibly be "known" to the system of interest and are therefore unimportant.

If the thermostat is comparable in size to the system of interest and if the cycle is traversed quickly, both the system of interest and the thermostat will be away from equilibrium. At any point in the cycle there is a profound difference between the nonequilibrium state generated by Gaussian processes and Nosé-Hoover processes. However, for both types of thermostat, Eqs. (1) and (23) take the same form. For Gaussian thermostats the change in the kinetic temperature of the thermostat is instantaneous, whereas for Nosé-Hoover thermostats there is a variable phase lag $\sim \tau_\alpha$ in Eq. (15). (The value of this feedback time constant is arbitrary.) The *only* "temperature" any of these systems have in common is the equilibrium thermodynamic temperature that they will relax to if at any point in the cycle the parametric change is halted. At any point in the cycle the precise nature of the nonequilibrium state (e.g., the instantaneous average pressure or energy) is highly dependent on the phase lag τ_α, or whether the thermostat is Gaussian or Nosé-Hoover like.

In Planck's discussion of Clausius' inequality,[9] at any instant in the cycle, T is the equilibrium thermodynamic temperature of the particular large equilibrium reservoir with which the system of interest is currently in contact.

Clausius' thermodynamic inequality equation (1), is of course only exact in the thermodynamic limit, and in small systems it can occasionally be violated as in Eq. (23). The probability ratio that for a finite system the work integral takes on a value A compared to $-A$ can be computed from a time dependent version of the fluctuation theorem.

Equations (1) and (23) show that on average we cannot construct a perpetual motion machine of the second kind. A perpetual motion machine of the second kind would require that $\oint_P \langle dQ^{therm}/T\rangle < 0$, so that ambient heat from the reservoir is converted into useful work. Thus, the proof of Eqs. (1) and (23) constitutes a direct mechanical proof of Clausius' statement of the second "law" of thermodynamics.

If the cycle is reversible we can apply Eq. (1) to the forward cycle and to the reversed cycle that must have the same value for magnitude of the integral but opposite sign. The only possible value for both integrals is therefore zero:

$$\lim_{N\to\infty} \oint_{qs} \frac{dQ^{therm}}{T} = 0. \qquad (24)$$

The subscript *qs,* denotes a quasi-static cycle. We note that a quasi-static cycle cannot have any transients and is always periodic.

Applying the same arguments as we did for the quasi-static cyclic integral of the generalized work shows that the quasi-static integral from an initial to a final state,

$$\lim_{N\to\infty} \int_{qs\, i}^{f} \frac{dQ^{therm}}{T} \equiv S_f^{therm} - S_i^{therm} \qquad (25)$$

is, in fact, a state function, denoted by the symbol $S^{therm.}$

The state function S^{therm}, defined in Eq. (25) is known as the change in the *equilibrium entropy* of the thermostating reservoir.

If we combine Eqs. (10), (17), and (21), we see that in the thermodynamic limit where fluctuations are negligible,

$$\lim_{N\to\infty} \int_{qs\, i}^{f} \frac{dQ^{therm}}{T} = -\left(\frac{A_f - H_{E,f}}{T_f}\right) + \left(\frac{A_i - H_{E,i}}{T_i}\right)$$
$$= -S_f^{soi} + S_i^{soi}, \qquad (26)$$

where in the second term we have used our previously proven[5] relation between the partition function, the Helmholtz free energy, and the equilibrium thermodynamic temperature. Now if we compare Eqs. (25) and (26), we see that in the thermodynamic limit, the equilibrium entropy, the Helmholtz free energy, and the energy must be related by the following equation:

$$A_{f/i} = H_{E,f/i} - S_{f/i}^{soi}\, T_{f/i}, \qquad (27)$$

where we have used the fact that the heat gained by the thermostat is equal and opposite to the heat gained by the system of interest.

Equation (25) tells us another very important piece of information. The integration factor for the heat in quasi-static (i.e., reversible) processes is the time dependent thermodynamic temperature. The fact that the equilibrium temperature is the integrating factor for the heat ultimately comes from the form of the canonical equilibrium distribution function. The relaxation theorem says that this distribution is *unique*[5] for T-mixing systems. The consequence of this is that the integrating factor for the heat is also *unique*.

We make a final observation on Clausius' inequality. Suppose instead of cycling the temperature we perform a work cycle keeping the temperature of the thermal reservoir fixed. In this case we can make the thermal reservoir very large and in the limit where it becomes infinitely large compared to the system of interest, it can be viewed as being in thermodynamic equilibrium at a fixed equilibrium thermodynamic temperature.

At a great physical distance from the system of interest we could fix the reservoirs temperature with either a Nosé-Hoover or a Gaussian isokinetic thermostat—it does not matter. The system of interest cannot distinguish the fine details of how the heat is ultimately and remotely removed.

For this case, if and only if the system settles into a periodic work cycle, Eq. (23) becomes

$$\lim_{t\to\infty} \oint_P ds\ \langle \dot{X}(t+s)\rangle = \lim_{t\to\infty} \beta \oint_P ds\ \langle \dot{Q}^{therm}(t+s)\rangle$$
$$\Rightarrow \lim_{t\to\infty} \oint_P \langle dQ^{therm}\rangle \geq 0, \qquad (28)$$

where T is the fixed equilibrium thermodynamic temperature of the reservoir. This shows that for periodic work processes carried out in contact with either a large thermal reservoir at a fixed equilibrium thermodynamic temperature or a small nonequilibrium thermal reservoir with a fixed underlying equilibrium temperature, the cyclic integral of $\langle dQ\rangle$ is positive on average, and for quasi-static processes, $\int_{qs} \langle dQ\rangle$ is a state function.

VII. CONCLUSIONS

We have proved Clausius' thermodynamic inequality from time reversible microscopic equations of motion. Clausius of course proved his theorem *assuming* the second "law" of thermodynamics. Our proof requires no such assumption. Our proof rests on the laws of mechanics, on the axiom of causality, ergodic consistency, and the assumption that the system is T-mixing. This is quite a different logical position from that used by Clausius in 1854. The requirement of causality is extensively discussed in Ref. 6. Basically, causality is required for the proof of all fluctuation theorems.

The necessary and sufficient conditions for a system to be T-mixing are not presently known. However, it is extremely easy to test whether a system is T-mixing. In thermostated systems driven by a dissipative field, if the system is T-mixing it will relax to a time independent nonequilibrium steady state where the averages of suitably smooth phase functions become time independent at long times. This is easily seen from the dissipation theorem.[12] Similarly in T-mixing relaxing systems that possibly interact with a thermostat but which are not driven by external dissipative fields, averages of suitably smooth phase functions become time independent at long times.

Our derivation allows us to understand the meaning of the "temperature" appearing in Clausius' inequality for *nonequilibrium* systems. Temperature can only be defined for equilibrium systems but Clausius' (strict) *inequality* refers to nonequilibrium systems. We also now know that Clausius' inequality is only valid for periodic cycles. Planck[9] also realized this. Most textbooks do not state this requirement.

When subject to a periodic parametric cycle, not all systems will settle into a periodic response. For small systems (not covered by Clausius' proof of his inequality), we know that the inequality is only valid if we take ensemble averages. We now also know that the integrating factor for the heat is unique for T-mixing systems. Other conclusions follow.

Our proof used the vehicle of the Nosé-Hoover thermostat. It provided a simple means of constructing a time dependent thermal protocol. The formal results are independent of how near or far this Nosé-Hoover thermostated region is from the system of interest. If the thermostat is moved progressively further from the system of interest, there is no way that the system of interest can "know" the details of how the thermostating is achieved. This means that the formal results are independent of the Nosé-Hoover thermostat. Indeed, we could have derived the same formal results using a variation of a Gaussian thermostat where the kinetic energy of the thermostated particles was forced to be a periodic function of time with no fluctuations away from the periodic function.

If we consider the thermal equilibration of a T-mixing system that has a nonuniform initial temperature distribution, we see from the relaxation theorem[5] that the final equilibrium state is the one with a spatially uniform thermodynamic temperature distribution—as in the canonical or microcanonical distributions. The proof shows[5] that in T-mixing systems these equilibrium distribution functions are unique. This is equivalent to a proof of the zeroth "law" of thermodynamics. Subsystems of an equilibrium system must all be at the same thermodynamic temperature.

We now look again at Eq. (23). If our system is subject to a periodic thermal protocol and if the system settles into a periodic cycle, the ensemble averaged heat absorbed by the thermostat (dQ^{therm}) is non-negative $\lim_{t\to\infty}\oint_P\langle dQ^{therm}/T\rangle \geq 0$. If the sign was reversed we would have been able to construct a perpetual motion machine of the second kind. So we have given a proof of the second "law" of thermodynamics, since Clausius' statement of that "law" refers to the impossibility of constructing such a machine.

There is a complementary inequality for the system of interest (*soi*), namely,

$$\lim_{t\to\infty}\oint_P\langle dQ^{soi}/T\rangle \leq 0. \tag{29}$$

If we combine the system of interest and the thermostat we see that for the combined system (*uni*) we find that for all periodic cycles, nonequilibrium and quasi-static,

$$\lim_{t\to\infty}\oint_P\langle dQ^{uni}/T\rangle = 0. \tag{30}$$

Because the temperature appearing in Eqs. (29) and (30) is the temperature of the underlying equilibrium state, this temperature is the same for the system of interest and for the thermal reservoir.

At first sight Eq. (30) seems somewhat unexpected, at least for nonequilibrium systems, but we have known since Gibbs' work[10,11] that the time derivative of the fine grained Gibbs entropy $\dot{S}$ of a thermally isolated nonequilibrium system is zero:

$$\dot{S}^{uni} = -k_B\frac{d}{dt}\int_D d\mathbf{\Gamma}\ f^{uni}(\mathbf{\Gamma},t)\ln[f^{uni}(\mathbf{\Gamma},t)] = \frac{\dot{Q}^{uni}}{T} = 0. \tag{31}$$

In our proof of the relaxation theorem for systems in contact with a heat bath,[5] we proved that the Helmholtz free energy and the partition function are related by

$$A = -k_BT\ln\left[\int_D d\mathbf{\Gamma}\ \exp[-H_E(\mathbf{\Gamma})/k_BT]\right]. \tag{32}$$

We also proved for equilibrium systems, the equivalence of the target kinetic temperature employed in a Nosé-Hoover thermostat and the thermodynamic temperature for canonical systems.

Now we have arrived at a completely new logical position. We have *proved* the zeroth, first, and second "laws" of thermodynamics; the latter in the form of Clausius' equality. This means that logically we can now construct thermodynamics without any assumptions except the laws of mechanics, the assumption of T-mixing, ergodic consistency, and the axiom of causality.

ACKNOWLEDGMENTS

We would like to thank the referees for their close reading of an earlier draft of this manuscript. We would also like to thank the Australian Research Council who supported part of this research.

[1] S. R. Williams, D. J. Searles, and D. J. Evans, Phys. Rev. Lett. **100**, 250601 (2008).

[2] G. E. Crooks, J. Stat. Phys. **90**, 1481 (1998); Phys. Rev. E **60**, 2721(1999).

[3] G. P. Morriss and L. Rondoni, Phys. Rev. E **59**, R5 (1999).

[4] J. C. Reid, E. M. Sevick, and D. J. Evans, Europhys. Lett. **72**, 726 (2005).

[5] D. J. Evans, D. J. Searles, and S. R. Williams, "A Simple Mathematical Proof of Boltzmann's Equal *a priori* Probability Hypothesis," in *Diffusion Fundamentals III*, edited by C. Chmelik, N. Kanellopoulos, J. Karger, and D. Theodorou (Leipziger Universitatsverlag, Leipzig, 2009), p. 367–374; J. Stat. Mech.: Theory Exp. 2009, P07029 (2009).

[6] D. J. Evans and D. J. Searles, Adv. Phys. **51**, 1529 (2002).

[7] D. Chandler, *Introduction to Modern Statistical Mechanics* (Oxford University Press, New York, 1987), p. 137.

[8] See, for example, p. 332 of C. Jarzynski, Annu. Rev. Condens. Matter Phys. **2**, 329 (2011).

[9] M. Planck, *Treatise on Thermodynamics*, 3rd Ed., *translated from the 7th German Edition* (Dover, New York, 1945), p. 99.

[10] D. J. Evans and G. P. Morriss, *Statistical Mechanics of Nonequilibrium Liquids* (Academic, London, 1990), pp. 251 and 252.

[11] P. Ehrenfest and T. Ehrenfest, *The Conceptual Foundations of the Statistical Approach in Mechanics*, translated by M. Kac and G. E. Uhlenbeck (Dover, New York, 1990) [P. Ehrenfest and T. Ehrenfest, "Begriffliche Grundlagen der Statistischen Auffassung in der Mechanik," in *The Encyklopadie der Mathematischen Wissenschaften* (B. G. Teubner, Leipzig, 1912), vol. IV:2:II, No. 6.]

[12] D. J. Evans, D. J. Searles, and S. R. Williams, J. Chem. Phys. **128**, 014504 (2008); (Erratum) **128**, 249901 (2008).