On the probability of violations of Fourier’s law for heat flow in small systems observed for short times

Denis J. Evans, Debra J. Searles, and Stephen R. Williams

I. INTRODUCTION

In a nonequilibrium system a thermodynamic force, \( X_i \), or a mechanical field, \( F_i \), acts on the system, preventing it from relaxing to equilibrium. Close to equilibrium this results in the spontaneous production of entropy, that per unit time is proportional to the product of the thermodynamic or mechanical force, the system volume, \( V \), and the dissipative flux, \( J \), divided by the absolute temperature, \( T \). The Second Law of Thermodynamics implies that for large systems, the instantaneous production of entropy per unit time is non-negative. This is in spite of the fact that the microscopic equations of motion are time reversible. Over the past 15 years there has been considerable progress toward understanding the microscopic origin of this irreversibility—especially for mechanical nonequilibrium systems. The fluctuation theorem\(^2\)–\(^6\) (FT) gives a formula for the probability ratio that in a thermostated nonequilibrium system subject to a dissipative mechanical field, the time averaged dissipative flux, \( \langle J \rangle \), takes on a value, \( A \), to minus the value, namely, \(-A\). This formula is an exact analytic expression for the probability—that for a finite system and for a finite time, the dissipative flux flows in the reverse direction to that required by the Second Law of Thermodynamics. This theorem is quite general. It applies arbitrarily far from equilibrium even where the spontaneous production of entropy or the thermodynamic temperature cannot be defined. It has been shown to apply to both deterministic\(^2\)–\(^5\) and stochastic systems.\(^6\)

Most of the work on the FT has been devoted to mechanical processes. In the present paper we derive a FT for a system undergoing the thermal process of heat conduction. The flow of heat is maintained by two thermal reservoirs that are maintained at different temperatures. The boundary regions are, in the steady state, maintained at the different temperatures using Nosé–Hoover thermostats.\(^7\) These time reversible deterministic thermostats are well known in computer simulations, but they are unnatural (i.e., no atoms obey these equations of motion in natural systems). The equations of motion for the system of interest, the atoms sandwiched between the thermostats, are entirely natural (i.e., they are completely Newtonian). In contrast to nonequilibrium mechanical systems, there is no external mechanical field appearing in the equations of motion for the atoms in the system of interest. In this thermal system it is the boundary conditions that prevent the system relaxing to equilibrium.

The thought experiment we have in mind is the following. At \( t=0 \) we have three contiguous equilibrium systems, which from left to right are labeled H, 0, C. Initially all three systems are at a common temperature \( T_0 \). For simplicity we assume that each of the systems is composed of atoms with the same interatomic interactions, that there is interfacial thermal contact between the H and 0 systems and the 0 and C systems. The number of atoms in the H and C systems is \( N_H = N_C = N_T \), \( \neq N_0 \), and the number of atoms in region “0” is \( N_0 \). At \( t=0 \) the two outer systems, H and C are brought into contact with thermal reservoirs that rapidly drive their temperatures to \( T_H \) and \( T_C \), respectively, where again for simplicity the thermal reservoir temperatures are related to the initial temperature by the relation \( T_0 = (T_H + T_C) / 2 \). After this thermal contact we expect that the 0 system will be driven away from equilibrium as heat flows on average, from the hot reservoir H, through the 0 system toward the cold reservoir, C. We also expect that system 0 will be subject to...
essentially random heat fluctuations due to the fact that the two heat reservoirs are uncorrelated and the precise heat flowing in from the hot reservoir does not instantaneously equal that flowing out to the cold reservoir. We call these heat fluctuations “boundary” fluctuations. The system is selected to be sufficiently small so that fluctuations in the system properties are observable and the relaxation time is conveniently short.

Throughout this paper we assume that our systems are transient mixing (T-mixing). Our definition of T-mixing is slightly different from the usual mathematical definition of mixing. We assume that transient correlation functions \( \langle A(t)B(t) \rangle \), of smooth zero mean phase functions \( A(\Gamma(t)), B(\Gamma(t)) \), decay to zero at long times. In these transient time correlation functions (TTCFs), \( t = 0 \) corresponds to the time when the distribution function is in its initial form and \( t \rightarrow \infty \) corresponds to the steady state. The mathematical definition of mixing involves time invariant distribution functions, and therefore only equilibrium or steady state time correlation functions are relevant. We take the domain of phases \( \Gamma \) over which the system is T-mixing to include all the allowed phase space within the volume of the system under study.

After relaxation of initial transients, we expect the 0 system to relax, not to equilibrium but because the system is T-mixing, to a unique nonequilibrium steady state defined by \( N_0', T_{01}, T_C \), and the conduction cell’s geometrical dimensions. A preliminary account of some of the results presented here was given in a conference proceeding.\(^3\)

The FT states\(^9\) that if a system satisfies the following conditions:

- it has deterministic time reversible dynamics,
- satisfies ergodic consistency, and
- its initial averaged dissipation function is even in the momenta,

then, the averaged dissipation function\(^9\) defined in terms of the initial phase \( \Gamma = (q_1, \ldots, q_N, p_1, \ldots, p_N) \) and the initial distribution function, \( f(\Gamma, 0) \) as

\[
\int_0^t ds \Omega(s) = \ln \left( \frac{f(\Gamma(0), 0)}{f(\Gamma(t), 0)} \right) - \int_0^t ds \frac{\partial}{\partial \Gamma} \Gamma(\Gamma(s)) = \bar{\Omega}_i(\Gamma(0)) t, \tag{1}
\]

satisfies the following time reversal symmetry:

\[
\frac{\text{Pr}(\bar{\Omega}_i = A)}{\text{Pr}(\bar{\Omega}_i = -A)} = e^{\Lambda t}. \tag{2}
\]

Here \( \text{Pr}(\bar{\Omega}_i = A) \) is the probability that \( \bar{\Omega}_i \) takes on a value \( A \pm dA \). We note that the second term in the middle equality of Eq. (1) is the time integral of the phase space expansion factor:\(^7\) \( \Lambda = \left( \frac{\partial}{\partial \Gamma} \right) \cdot \bar{\Gamma} \). In closed Hamiltonian systems it is zero. In open Hamiltonian systems that gain or lose heat to a surrounding reservoir region that is so large that it may be regarded as being in thermodynamic equilibrium at a temperature \( T_{\text{res}} \), it is known that

\[
\Lambda = \frac{\dot{Q}}{k_B T_{\text{res}}}, \tag{3}
\]

where \( \dot{Q} \) is the heat added to the system from the surrounding reservoir region. In accord with the Second Law of Thermodynamics for nonequilibrium steady states, this is a negative quantity on average. Boltzmann’s constant is denoted as \( k_B \).

For mechanical systems that are sufficiently close to equilibrium that local thermodynamic equilibrium can be assumed to hold, the time averaged dissipation function multiplied by Boltzmann’s constant equals the so-called spontaneous rate of entropy production \( \lim_{t \rightarrow \infty} \lim_{\tau \rightarrow 0} \frac{\dot{S}}{k_B} = k_B \bar{\Omega}_i \) as defined in linear irreversible thermodynamics. In the weak gradient limit the total entropy production per unit time is given by the usual expression from linear irreversible thermodynamics: \( \dot{S} = \sigma V = \Sigma I_i V_i / T \), where \( V \) is the system volume, \( \sigma \) is the so-called entropy source strength, and the sum is over the product of all conjugate thermodynamic fluxes, \( J_i \), and thermodynamic forces, which is \( X_i \) divided by the temperature of the system, \( T \).

There are two different statements of the FT. If the time averages appearing in Eq. (2), \( \bar{\Omega}_i(t) \), are calculated from \( t = 0 \) and the ensemble averages [used to compute the probability distribution \( \text{Pr}[\bar{\Omega}_i(t)] \)] are taken over the initial ensemble, Eq. (2) is exact for all averaging durations \( t \), and the theorem is called the transient FT (TFT).\(^3,4\)

On the other hand, if the time and ensemble averaging are carried out in a nonequilibrium steady state, after the relaxation of the initial transients, Eq. (2) is termed the steady state FT (SSFT),\(^9\) and furthermore it is only true asymptotically, \( (t \rightarrow \infty) \). If the steady state is unique, the probability distribution in Eq. (2) can either be obtained from many segments taken from a single continuous phase space trajectory, or it can be obtained by taking a single segment from an ensemble of steady state trajectories, each of which originates from different initial phases a long time in the past.

We will show that if the system is T-mixing, the system will relax, except for a set of microscopic initial conditions of measure zero, to a nonequilibrium steady state whose properties and statistics are independent of the initial phase from which nonequilibrium steady state trajectories are constructed. Thus the system relaxes to a unique steady state. We say that such steady states are ergodic with respect to the initial distribution \( f(\Gamma, 0) \):

\[
\lim_{t \rightarrow \infty} \langle B(t) \rangle (\Gamma(0)) = \lim_{t \rightarrow \infty} \lim_{\tau \rightarrow 0} \{ f(\tau ds B(\Gamma(t + s))) / \tau \} \text{ for almost all } \Gamma(0). \]

It is “ergodic” because we are able to equate an ensemble average with a time average. This definition of ergodicity is thus slightly different from the usual mathematical definition, which concerns only invariant measures.

In this paper we derive the TFT for the thermal process of heat conduction. We test both the TFT and SSFT using nonequilibrium molecular dynamics computer simulation. We also prove that in the weak gradient limit, the dissipation function is given exactly by a microscopic expression for the spontaneous entropy production identified in linear irreversible thermodynamics for the heat conduction process.
The Second Law of Thermodynamics is violated, i.e., for some trajectories \( \bar{\Omega}_t < 0 \). If \( \langle \cdots \rangle_{\bar{\Omega}_t > 0} \) denotes a transient ensemble average over those trajectories where the time integrated dissipation function is positive, then

\[
\frac{\langle e^{-\bar{\Omega}_t} \rangle_{\bar{\Omega}_t > 0}}{\langle e^{-\bar{\Omega}_t} \rangle_{\bar{\Omega}_t < 0}} = \langle e^{-\bar{\Omega}_t} \rangle_{\bar{\Omega}_t < 0} \geq 1, \quad \forall t, \tag{4}
\]

and the probability of second law violations become exponentially small with an increased duration of violation, \( t \), and with the number of particles (since \( \bar{\Omega}_t \) is extensive in \( N \) and \( t \)).

The second law inequality states that transient ensemble averages of time integrals of the dissipation cannot be negative,

\[
\langle \bar{\Omega}_t \rangle \geq 0, \quad \forall t. \tag{5}
\]

This is true regardless of how close or far the system is to equilibrium and regardless of system size. Equation (5) does not imply that the ensemble averaged instantaneous dissipation function is positive at all times.

### II. MICROSCOPIC DESCRIPTION OF THERMAL CONDUCTION

Experimentally there are a number of ways in which walls can be thermostated. If the walls are made of highly thermally conductive material, a coolant may be circulated through channels in the reservoirs. Alternatively if the heat capacity of the reservoirs is huge compared to that of the thermal conduction cell, then the temperature variation in the reservoirs over relevant observation times may be regarded as insignificant. For theoretical analysis both of these mechanisms are too complex. Instead we employ the so-called Nosé–Hoover thermostat in the reservoir regions in order to maintain these regions at a fixed temperature. Its impact on the system of interest, namely, the thermal conduction cell, is only indirect. In an experiment, the material properties of the thermal conduction cell are independent of whether the reservoirs are maintained at a fixed temperature by virtue of the circulation of a coolant or the use of large heat capacity reservoirs. The thermal conductivity is a material property. It is independent of the precise chemical composition of the walls of a conduction cell. The theory that follows is also independent of the thermostating mechanism. The reason for this independence is that the formal fluctuation formulae are independent of precisely how far removed the thermostating region is from the system of interest. Thus we can move the thermostating region arbitrarily far from the system of interest and still generate the same fluctuation relation. There is no way that the system of interest can "know" precisely how the heat is ultimately removed by the remote thermostat. We note that in low-dimensional anharmonic chains, it is well known that there can be long ranged spatial correlations for heat flow (see Ref. 11, p. 343). In typical physical systems such correlations are much shorter ranged.

The aim is to derive fluctuation formulae for the transient response of the thermal conduction cell. We consider the system initially at equilibrium (because then the phase space distribution function is known). At this stage the whole system is isothermal. The temperature gradient is then applied and a heat flux develops.

The equations of motion for all the particles in the combined systems, \( H, 0, \) and \( C \), are

\[
\mathbf{q}_i = \mathbf{p}_i/m, \quad \mathbf{p}_i = \mathbf{F}_i - \alpha_H \mathbf{p}_A - \alpha_C \mathbf{p}_B, \tag{6}
\]

where \( \alpha_{HC} \) are the thermostat multipliers, \( T_{HC} \) are the required temperatures of the hot and cold regions, \( A_i \) and \( B_i \) are switches equal to 1 or 0. \( A_i \) is only nonzero for particles in region \( H \), and \( B_i \) is only nonzero for particles in region \( C \). The multipliers themselves satisfy the following equations of motion:

\[
\frac{d\alpha_{HC}}{dt} = \frac{1}{\xi} \left( \sum_{i \in HC} \frac{\mathbf{p}_i^2}{m} - \left( dN_T + 1 \right) k_B T_{HC} \right), \tag{7}
\]

where \( d \) is the number of Cartesian dimensions and \( N_T \) is the number of particles in each reservoir. The constant \( \xi \) controls the timescale, for fluctuations in the kinetic temperatures, of regions \( H \) and \( C \). The time constant is given by \( \tau_{HC} = O(\sqrt{\xi/dN_T k_B T_{HC}}) \). We always choose \( \xi = O(N_T) \) so that \( \tau_{HC} \) is intensive.

For simplicity, we assume that the walls are sufficiently dense that the particles from region 0 do not penetrate either of the reservoir regions—the walls are effectively solid. The details of the interatomic forces implicit in \( \{ \mathbf{F}_i \} \) will be described later. It is important to note that in the 0 region and the \( H \) and 0, and \( C \) and 0 interfaces, the equations of motion can be made arbitrarily realistic by improved modeling of the interatomic forces. In the 0 region there are no unnatural forces. Our system is very similar to that studied by Searles and Evans in 2001 and by Petravic and Harrowell in 2005, and although the dimensionality and the particle dynamics are very different, it has the same form for the dissipation function as in the system studied by Mejía-Monasterio and Rondoni in 2008.

The additional Nosé–Hoover thermostat ensures that in a steady state the reservoir regions are maintained at constant kinetic temperatures, \( T_H \) and \( T_C \). In a nonequilibrium steady state,

\[
\lim_{t \to \infty} \left( \frac{d\alpha_{HC}}{dt} \right) = 0 \Rightarrow \left( 1/(dN_T + 1) k_B \right) \sum_{i \in HC} \frac{\mathbf{p}_i^2}{m} = T_{HC}. \tag{8}
\]

### III. TFT FOR HEAT CONDUCTION

Since the system is T-mixing, it must be ergodic since otherwise we could trivially identify TTFs of smooth variables (namely, phase space domain occupation numbers) that never decay, thereby contradicting the T-mixing assumption. With \( T_H = T_C = T_0 \), we know that any arbitrary distribution will eventually relax to the unique equilibrium distribution.
We assume that at $t=0$ the initial phase space distribution, $f(\Gamma,0)$, is that unique equilibrium canonical distribution.a,7,14

$$f(\Gamma,0) = \frac{\exp[-\beta_0 H_0(\Gamma) + \xi(\alpha_{\Gamma}^2 + \alpha_{\Gamma}^2)/2]}{\int d\Gamma^{\prime} \exp[-\beta_0 H_0(\Gamma) + \xi(\alpha_{\Gamma}^2 + \alpha_{\Gamma}^2)/2]}.$$  \tag{9}

where $\beta_0 = 1/(k_B T_0)$, $H_0 = \Sigma_{i=1}^{2} 2m + \Phi(q)$ is the internal energy, $H_0' = H_0 + \xi(\alpha_{\Gamma}^2 + \alpha_{\Gamma}^2)/2$ is the extended internal energy, and $\Gamma' = (\Gamma', \alpha_{\Gamma} , \alpha_{\Gamma} )$ is the extended phase space vector.

The phase space expansion factor, $\Lambda(\Gamma)$, appearing in the Liouville equation,

$$\frac{df(\Gamma,t)}{dt} = -f(\Gamma,t)\Lambda(\Gamma),$$  \tag{10}

is

$$\Lambda = -dN_T \alpha_H - dN_T \alpha_C.$$  \tag{11}

Thus the formal Lagrangian solution of the Liouville equation is

$$f(\Gamma^{\prime}(t),t) = f(\Gamma^{\prime}(0),0) \exp \left[ \int_0^t ds \left( dN_T \alpha_H(s) + dN_T \alpha_C(s) \right) \right].$$  \tag{12}

From the equations of motion we see that the rate of change of the extended internal energy is

$$\dot{H}_0' = -dN_T k_B(T_H \alpha_H + T_C \alpha_C).$$  \tag{13}

Using Eq. (9) and substituting Eqs. (11) and (13) into Eq. (1) give the time averaged dissipation function as

$$\bar{\Omega}(\Gamma) = -\int_0^t ds \left( dN_T \beta_0 k_B(T_H \alpha_H(s) + T_C \alpha_C(s)) \right) \frac{-dN_T(\alpha_H(s) + \alpha_C(s))}{-dN_T(\alpha_H(s) + \alpha_C(s))}$$

$$= \int_0^t ds dN_T(\alpha_H(s) + \alpha_C(s))(T_H - T_C)$$

$$= dN_T \frac{(T_H - T_C)}{(T_H + T_C)} \int_0^t ds [\alpha_C(s) - \alpha_H(s)],$$  \tag{14}

where the second line follows from the imposed relationship between the initial temperature and target temperatures of the thermostats, $2T_0 = T_H + T_C$. Essentially the same expression for this dissipation function appears in Ref. 8 and more recently in Ref. 13. From the TFT (2) we see that the probability ratio of observing conjugate values for the time averaged difference in the thermostat multipliers is

$$\frac{\Pr(\alpha_{\Gamma} - \alpha_{\Gamma} = A)}{\Pr(\alpha_{\Gamma} - \alpha_{\Gamma} = -A)} = \exp \left[ dN_T \frac{T_H - T_C}{T_C + T_H} \right].$$  \tag{15}

The TFT for heat flow given by Eq. (15) is exact for arbitrary system size, observation time, $t$, and also arbitrarily far from equilibrium.

Later we will show that for weak applied temperature gradients, Eq. (14) for the dissipation function is the entropy production computed using linear irreversible thermodynamics. Equation (15) is a statement of the TFT for heat flow between Nosé–Hoover thermostatted walls. Since the system is T-mixing, as we shall see, the system will relax to a unique steady state. Therefore we can consider the SSFT\textsuperscript{13,15}

$$\lim_{t \to \infty} \ln \left[ \frac{\Pr(\alpha_{\Gamma} - \alpha_{\Gamma} = A)}{\Pr(\alpha_{\Gamma} - \alpha_{\Gamma} = -A)} \right] = \frac{dN_T (T_H - T_C)}{T_H + T_C}.$$  \tag{16}

Equations (15) and (16) are valid outside the linear regime. For our thermal conduction setup and with our initial conditions and thermostats, the only caveat is that the steady state formula requires the system to be T-mixing. Equations (15) and (16) are clearly consistent with the Second Law of Thermodynamics in that it is exponentially more probable for heat to flow from the hot reservoir to the cold reservoir, in which case $\alpha_{\Gamma} > 0$ and $\alpha_{\Gamma} < 0$ and from Eq. (13) we see that in the steady state limit $\lim_{t \to \infty} \alpha_{\Gamma} > \lim_{t \to \infty} \alpha_{\Gamma}$. In either the large system or the long time limit, the time averaged heat will only flow from hot to cold.

The dissipation theorem\textsuperscript{16} gives an exact TTCF expression for the ensemble average of the nonlinear response of an arbitrary phase variable, $B(\Gamma)$, as

$$\langle B(t) \rangle = \langle B(0) \rangle - \frac{dN_T (T_H - T_C)}{T_H + T_C}$$

$$\times \int_0^t ds \langle B(s) [\alpha_H(0) - \alpha_C(0)] \rangle.$$  \tag{17}

In this equation the angle brackets denote an average over the initial (i.e., $t=0$) ensemble and $\langle B(t) \rangle = \langle B(\Gamma(t)) \rangle$. Unlike the FTs, the dissipation theorem does not require ergodicity. The linearized weak field version of this equation is essentially identical to Eq. (11) in the paper by Petrvac and Harrowell.\textsuperscript{12} By comparing with the usual Kawasaki distribution function for a system driven by an external mechanical force, we see that although the system is a thermal non-equilibrium system where boundary conditions rather than external mechanical forces drive the system away from equilibrium, there is a formal resemblance of the nonlinear response to that obtained if we applied a fictitious mechanical field,

$$F = \frac{k_B (T_H - T_C)}{2}$$  \tag{18}

to the system. In this case the intensive dissipative flux $J$ can be identified as the fictitious function

$$J(\Gamma) = dN_T (\alpha_H(\Gamma) - \alpha_C(\Gamma)), $$  \tag{19}

where $n_T = N_T/V_T$ is the number density of the thermostat volumes. Equation (17) contains a great deal of information. Since the system is T-mixing, the TTCF appearing on the right hand side decays to zero at long times. More precisely we require that time integrals of TTCFs of phase functions and the time zero dissipation function (17) should converge in the infinite time limit. This implies that the ensemble average of any smooth phase function becomes time independent at sufficiently long time. The system must relax to a steady state and again by T-mixing that steady state must be unique.
This proves that T-mixing systems are ergodic with respect to the initial distribution. This simple result is quite remarkable.

We note that in the weak field [see Eq. (18)] limit, the linear response of the system to thermal conduction can be computed exactly from the time integral of an equilibrium time correlation function. In this limit T-mixing reduces to the mixing condition met in ergodic theory.

Choosing the phase function $B(\Gamma)$ to be the dissipation function itself, and using the second law inequality shows that at late times, the ensemble averaged dissipation function equals the time averaged dissipation function, and that average value is nonnegative. In fact it must be positive because the equilibrium state is the unique dissipationless state for T-mixing systems.\textsuperscript{14}

If the transient autocorrelation function for the dissipation function is positive for all times, then the relaxation to the equilibrium state is the unique dissipationless state for T-mixing systems.\textsuperscript{14}

The subtraction of the second term is ignored, ever longer times are required for convergence on the basis of the weak field entropy production appearing in linear irreversible thermodynamics. Difficulties are associated with this approach because there are many different microscopic quantities that yield the same average behavior. These difficulties are highlighted in this section.

Our system considers the transient response of the three regions H, 0, and C that are initially at the same temperature, $T_0$. At $t=0$ systems H and C are instantly brought into contact with Nosé–Hoover thermostats, which rapidly ($\tau_{W1} = O(\sqrt{d} k_B T H_H C)$ bring systems H and C to temperatures $T_H$ and $T_C$, respectively. Unlike in the preceding sections, the following analysis assumes that the temperature differences are small so that local thermodynamic equilibrium exists.

Without loss of generality we assume that the three regions H, 0, and C have a rectangular cross section of area $A$ and wall normals parallel to the $x$-axis, and the distance separating the thermostatted reservoirs is $L$. In the limit of small temperature gradients, linear irreversible thermodynamics gives the total spontaneous entropy production inside the thermal conduction cell, 0, as

\begin{equation}
\Sigma_{\text{therm}}(t) = \int_V d\mathbf{r} \sigma(\mathbf{r}, t) = \int_V d\mathbf{r} J_{Q}(t) \cdot \nabla T^{-1} = A \int_{-L/2}^{+L/2} dx d\mathbf{r} J_{Q}(t) \frac{d(1/T(x))}{dx} = A \int_{-L/2}^{+L/2} dx \frac{dJ_{Q}(x,t)}{dx} = A \int_{-L/2}^{+L/2} dx \frac{dJ_{Q}(x,t)}{dx} + O\left(\frac{d^3}{dx^3}\right).
\end{equation}

In the weak gradient limit, we expect that the temperature gradient will be a linear function of position and terms $O(d^3/dx^3)$ may be ignored (see Ref. 19 for details).

First we note that if the temperature difference is zero $\Sigma_{\text{therm}}(t; T_H - T_C = 0) = 0$, $\forall t$. Due to the fact that the two thermostats are uncorrelated, the first term contains contributions from random energy fluctuations in the thermostats. The first term is not identically zero even when the entire system is at equilibrium. However the spontaneous entropy production (like the dissipation function) must be identically zero at equilibrium.\textsuperscript{14} The subtraction of the second boundary term ensures that when $T_H = T_C$, the spontaneous entropy production is indeed identically zero at equilibrium.

The relative importance of the boundary terms only increases as the temperature gradient becomes smaller. As equilibrium is approached the energy fluctuations due to the uncorrelated thermostats is of constant magnitude, while the pure dissipation vanishes as $(T_H - T_C)^2$. Therefore if this second term is ignored, ever longer times are required for convergence of the SSFT and the transient fluctuation relation (TFR) itself [Eq. (2)] will also only be satisfied asymptotically in contradiction to the TFT. [Note: we refer to a

It is clear that the dissipation function is not instantaneously related to the time derivative of the fine grained Gibbs entropy.

**IV. CONNECTION WITH LINEAR IRREVERSIBLE THERMODYNAMICS**

Most previous work on the FT for heat flow has concentrated on one-dimensional systems.\textsuperscript{17,18} Furthermore, the approach taken by these groups has been to define the argument of the FT on the basis of the weak field entropy production appearing in linear irreversible thermodynamics.

Without loss of generality we assume that the three regions H, 0, and C have a rectangular cross section of area $A$ and wall normals parallel to the $x$-axis, and the distance separating the thermostatted reservoirs is $L$. In the limit of small temperature gradients, linear irreversible thermodynamics gives the total spontaneous entropy production inside the thermal conduction cell, 0, as

\begin{equation}
\Sigma_{\text{therm}}(t) = \int_V d\mathbf{r} \sigma(\mathbf{r}, t) = \int_V d\mathbf{r} J_{Q}(t) \cdot \nabla T^{-1} = A \int_{-L/2}^{+L/2} dx d\mathbf{r} J_{Q}(t) \frac{d(1/T(x))}{dx} = A \int_{-L/2}^{+L/2} dx \frac{dJ_{Q}(x,t)}{dx} = A \int_{-L/2}^{+L/2} dx \frac{dJ_{Q}(x,t)}{dx} + O\left(\frac{d^3}{dx^3}\right).
\end{equation}

In the weak gradient limit, we expect that the temperature gradient will be a linear function of position and terms $O(d^3/dx^3)$ may be ignored (see Ref. 19 for details).

First we note that if the temperature difference is zero $\Sigma_{\text{therm}}(t; T_H - T_C = 0) = 0$, $\forall t$. Due to the fact that the two thermostats are uncorrelated, the first term contains contributions from random energy fluctuations in the thermostats. The first term is not identically zero even when the entire system is at equilibrium. However the spontaneous entropy production (like the dissipation function) must be identically zero at equilibrium.\textsuperscript{14} The subtraction of the second boundary term ensures that when $T_H = T_C$, the spontaneous entropy production is indeed identically zero at equilibrium.

The relative importance of the boundary terms only increases as the temperature gradient becomes smaller. As equilibrium is approached the energy fluctuations due to the uncorrelated thermostats is of constant magnitude, while the pure dissipation vanishes as $(T_H - T_C)^2$. Therefore if this second term is ignored, ever longer times are required for convergence of the SSFT and the transient fluctuation relation (TFR) itself [Eq. (2)] will also only be satisfied asymptotically in contradiction to the TFT. [Note: we refer to a
fluctuation relation when the mathematical form of a FT is proposed in conjunction with the substitution of a variable (usually with the same average value) for which the corresponding theorem has not been proved.] This is a standard problem with fluctuating boundary terms and is related to the well known convergence problems for the Gallavotti–Cohen FT for thermostatted steady states as equilibrium is approached.13,15,20,21

The ultimate fluxes into and out of our system are given by the energy gain or loss by the thermostats themselves. These are the only nonconservative elements of our system. The thermostating terms are analogous to the coolant in a physical thermostat. Once the energy is taken up by the circulating coolant, the physical circulation of that coolant removes that energy from the system of interest. With this in mind it seems natural to evaluate the H/C “heat fluxes” as

$$J_{\phi,CH}(t)A = ± dN_f k_b a_C(t) k_B T_{CH}.$$  \hspace{1cm} (22)

These are the heat fluxes across the area, A, immediately before or after the heat is removed or injected by the thermostats themselves. [\(\text{lim}_{\tau \to \infty} \bar{a}_H(t) = \text{lim}_{\tau \to \infty} \bar{a}_C(t) = 0\); note the difference in signs!]

Substituting these heat fluxes into Eq. (21) gives

$$\Sigma_{therm}(t) = \left( dN_f k_b a_H(t) - dN_f k_b a_C(t) \right) - \frac{1}{T_0} \left( dN_f k_b T_H a_H(t) - dN_f k_b T_C a_C(t) \right)$$

$$= \left( dN_f k_b a_H(t) (T_H + T_C)/2 + dN_f k_b a_C(t) (T_H + T_C)/2 - dN_f k_b a_H(t) - dN_f k_b a_C(t) \right) \frac{2}{(T_H + T_C)}$$

$$= - dN_f k_b (a_H(t) - a_C(t)) \left( \frac{T_H - T_C}{T_C + T_H} \right) = \Omega(t) k_B.$$  \hspace{1cm} (23)

Our microscopic expression for the weak field thermodynamic entropy production \(\Sigma_{therm}(t)\) exactly equals the instantaneous dissipation function multiplied by Boltzmann’s constant, \(k_B\Omega(t)\). This confirms for thermal conduction processes that the dissipation function is a microscopic expression for the generalized (i.e., arbitrarily far from equilibrium) spontaneous entropy production. This appears to be true for both mechanically and thermally driven nonequilibrium systems.

In previous studies,17,18 this second term in the entropy production (21) was not included; it was assumed that there is a perfect balance between the ingoing and outgoing fluxes \(J_{OH}(t) = J_{OC}(t)\), and many earlier workers used an entropy production

$$\Sigma_A(t) = -A J_{q}(t) \left( \frac{1}{T_H} - \frac{1}{T_C} \right).$$  \hspace{1cm} (24)

Although the equality of these two heat fluxes \(J_{OH}(t)\) and \(J_{OC}(t)\) is true on average in a steady state, it is not true instantaneously and should not be used here since the FT requires us to analyze fluctuations in the entropy production. For transient experiments, the averaging time \(\tau_{TD}\) required for the equality of \(\Sigma_A(t)\) with \(k_B\bar{\Omega}(t)\) is of the order of the time it takes heat to diffuse across the whole system. Thus if

$$\tau_{TD} = O(L^2 \rho c_p/\lambda),$$  \hspace{1cm} (25)

where \(\rho\) is the mass density, \(c_p\) the intensive constant pressure heat capacity, and \(\lambda\) the thermal conductivity, then

$$\lim_{\tau_{TD} \to \infty} \Sigma_A(t) = k_B\bar{\Omega}(t).$$  \hspace{1cm} (26)

It is important to note that the thermal diffusion time \(\tau_{TD}\) is macroscopic, scaling as the system dimension squared. The neglect of the second term in Eq. (21) may partially explain the poor results obtained by Sano.18,22,23 For steady states the time required for the equality of the two heat fluxes is expected to be microscopic and is related to the heat capacity of the system of interest.

In summary, the substitution of \(\Sigma_A(t)\) for the actual dissipation function \(\Omega(t) = \Sigma_{therm}(t)\) can be expected to make the TFR only an asymptotic relation and increase the averaging times required for the already asymptotic steady state fluctuation relation to converge. Both of these convergence times diverge to infinity as equilibrium is approached. In contrast the convergence time for the SSFT does not diverge upon approaching equilibrium.

The limitations of obtaining the FT from irreversible thermodynamics are now quite clear. First, unlike the derivation of the FT leading to Eqs. (15) and (21), which is valid at all fields and for all averaging times, the thermodynamic approach is only possibly valid close to equilibrium. Also, extreme care is required so as not to neglect terms that vanish on average in the steady state but have nonzero instantaneous values. Although we have proven that the dissipation function is a quantity that is recognizable as the spontaneous entropy production, if we did not know the dissipation function in advance and could only use macroscopic forces and fluxes, our chances of guessing the correct microscopic dissipation function would be very small. A key clue that many previous workers failed to note is that the instantaneous dissipation function must be identically zero at equilibrium.14 This is a necessary but not sufficient condition.

One can also show that the change in the fine grained Gibbs entropy (20) \(-\dot{S}_G(t) = dN_f k_b (a_H(t) + a_C(t)) = -k_b (\Lambda(t))\) is equal to the ensemble average of the phase space expan-
sion factor. The time averaged rate of change of the fine grained Gibbs entropy is in the steady state also equal, for sufficiently long averaging times, to minus one times the steady state average of the thermodynamic entropy production,

$$\bar{S}_{G,t} = -k_B \langle \bar{A} \rangle' = -A \langle J_{G,t} \rangle \left( \frac{1}{T_H} - \frac{1}{T_C} \right) = \left( \bar{S}_{\text{therm}}, \right)_{t \to \infty} = k_B \langle \bar{A} \rangle'.$$

The second equality is simply a restatement of Eq. (20); the third equality is true for sufficiently long averaging times. We note that $\Lambda(t)$ is the argument of the steady state Gallavotti–Cohen FT. The discussion above shows that long averaging times are required for this asymptotic FT to converge, and worse, as equilibrium is approached, the convergence times are required for this asymptotic FT to converge, and worse, as equilibrium is approached, the convergence factor. The time averaged rate of change of the fine grained Gibbs entropy is in the steady state also equal, for sufficiently long averaging times, to minus one times the steady state average of the thermodynamic entropy production,

The second equality is simply a restatement of Eq. (20); the third equality is true for sufficiently long averaging times. We note that $\Lambda(t)$ is the argument of the steady state Gallavotti–Cohen FT. The discussion above shows that long averaging times are required for this asymptotic FT to converge, and worse, as equilibrium is approached, the relative importance of the boundary terms grows like $(T_H - T_C)^{-2}$.13,20,21

Table I gives a summary of the various expressions (some approximate) that have been used in fluctuation relations. Only two of those expressions will satisfy the exact TFT, but not instantaneously.

<table>
<thead>
<tr>
<th>Definition</th>
<th>Equation</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_B \Omega(t) = dN_j k_B (T_H - T_C) (\alpha_j(t) - \alpha_j(t)) / 2T_0$</td>
<td>(14)</td>
<td>Satisfies an exact TFT (15); argument in nonlinear response theory; subject of TFT and ES SSFT</td>
</tr>
<tr>
<td>$\dot{S}_{G,t} = -dN_j k_B (\alpha_j(t) + \alpha_j(t))$</td>
<td>(20)</td>
<td>Time derivative of Gibbs entropy; averages to $-\dot{\Sigma}<em>{A,t}$ for $t \gg \tau</em>{\text{HFT}}$; subject of GCFT</td>
</tr>
<tr>
<td>$\Sigma_{\text{therm}}(t) = -A \left( J_{G,t} T_H - J_{G,t} T_C \right) + A \left( J_{G,t} T_C - J_{G,t} T_H \right) / T_0$</td>
<td>(21)</td>
<td>Is the natural expression for instantaneous dissipation $\Sigma_{\text{therm}}(t) = k_B \Omega(t)$</td>
</tr>
<tr>
<td>$\dot{\Sigma}<em>{A,t} = -A J</em>{G,t} \left( \frac{1}{T_H} - \frac{1}{T_C} \right)$</td>
<td>(24)</td>
<td>Equivalent to $-\dot{S}<em>{G,t}$ for $t \gg \tau</em>{\text{HFT}}$; averages to $k_B \Omega(t)$ for $t \gg \tau_{\text{TD}}$</td>
</tr>
</tbody>
</table>

Applying the second law inequality (5) to our system shows that in the long time limit where the time averaged dissipation function equals the time averaged $\Sigma_{A,t}$ [Substitute Eq. (27) into Eq. (5)], we obtain

$$\lim_{t \to \infty} \frac{\langle J_{G,t} \rangle}{\Delta T} = -\lambda(\partial T / \partial x) = \frac{\lambda}{\partial T / \partial x} > 0 \Rightarrow \lambda(0) > 0,$$

where $\lambda(0)$ is the limiting zero gradient thermal conductivity and the temperature gradient is negative. Since the system is $T$-mixing, a unique steady state is generated for any given temperature gradient. If we assume that in the weak temperature gradient limit the thermal conductivity is finite (there are divergent systems especially in one dimension11 where it is thought that the limiting zero gradient thermal conductivity scales with system size), then the limiting thermal conductivity must be positive. (The thermal conductivity cannot be zero since in $T$-mixing systems, the only state that has zero dissipation is the equilibrium state.14)

V. COMPUTER SIMULATION RESULTS

In order to test the fluctuation formula given by Eqs. (15) and (16), simulations of a two-dimensional fluid between walls were carried out. The dissipation function is described in terms of the dissipative flux and external field as per Eqs. (18) and (19) (i.e., $\bar{\Omega}_G = -\beta \vec{V}_F \tau_E$, where $V_F$ is the volume of each thermostating region). The system consisted of three sections: a fluid region of 64 particles between two walls each containing 32 particles. The complete system was initialized in a cubic box with periodic boundary conditions in the direction parallel to the walls. The equations of motion for all the particles are given by Eqs. (6) and (7). The relation was tested for both transient and steady state trajectory segments.

For the particles in the fluid region (labeled as region 0), the switches $A_1$ and $B_1$ were set to zero at all times, and therefore these particles obeyed Newtonian mechanics. The forces on these particles were solely due to their interactions with other particles via the dimensionless Weeks–Chandler–Andersen (WCA) pair potential $(m, e, \sigma)$.7 The particle density of the fluid region was initially set to $n=0.4$.\]
The wall particles were thermostatted using the Nosé–Hoover thermostat, and forces were applied so that their density was maintained at a higher value of $n = 1.2$. One wall was designated the hot wall, $H$, and the other the cold wall, $C$. In the hot wall, the switches were set to $A_1 = 1$ and $B_1 = 0$; whereas in the cold wall they were set to $A_1 = 0$ and $B_1 = 1$. These particles interacted with other particles via a WCA pair potential. In addition, a spring potential was applied to prevent wall particles from diffusing outside of their respective regions $[U(r_{ij}) = \frac{1}{2} k (r_{ij} - r_{eq})^2$, where $k = 57$ and each layer of particles in the wall was subject to a layer force]. These forces are nonphysical and designed to ensure the wall remained intact throughout the simulation. During an equilibration period, the temperature in the Nosé–Hoover thermostat was set to $T = 1.0$ for both the hot and the cold walls. After this period, the Nosé–Hoover thermostat was set to $T_H = 1.1$ and $T_C = 0.9$ to create a temperature gradient across cell 0.

In order to test the TFT, approximately $8 \times 10^7$ trajectories were simulated. For each trajectory, this involved sampling a starting point from the equilibrium distribution, applying the temperature gradient and measuring the value of $\overline{J_t}$ for a trajectory of length $t = 1.6$. The $8 \times 10^7$ values of $\overline{J_t}$ obtained were then used to construct a frequency histogram (see Fig. 1) from which the probabilities required for testing Eq. (15) could be obtained.

Equation (15) was tested by plotting $-(1/\beta_0 V_F) \ln [Pr(\overline{J_t} = A)/Pr(\overline{J_t} = -A)]$ versus $A$, as is shown in Fig. 2. According to the FT for this system (see Eq. (15)) and using results (18) and (19), this plot should yield a straight line of unit slope. Clearly the numerical data are consistent with the theoretical prediction.

In order to test the FT for steady states, a single steady state trajectory was simulated with the temperature gradient applied. This trajectory was divided into trajectory segments of duration from $t = 2.5$ to $t = 25$, and the value of $\overline{J_t}$ was determined for each segment. Approximately $8 \times 10^4$ segments were obtained. Histograms for these steady state averages are shown in Fig. 3. The probabilities required for testing Eq. (16) were then obtained from the histogram. Equation (16) was tested by plotting $-(1/\beta_0 V_F) \ln [Pr(\overline{J_t} = A)/Pr(\overline{J_t} = -A)]$ versus $A$, as is shown in Fig. 4. According to the SSFT and using results (18) and (19), this plot should be a straight line of unit slope in the
as was done in Refs. 17 and 18, the TFR in terms of this invalid convergence to a slope of unity is expected in the long time limit.

Figure 5 shows the convergence of the slope toward 1, and clearly the numerical result is consistent with the theoretical prediction.

We note that if the second term in Eq. (21) is neglected, as was done in Refs. 17 and 18, the TFR in terms of this expression, $\Sigma_A$ [see Eq. (24)], for the entropy production is invalid [see Fig. 6(a)]—this relation is not exact for all times and only becomes correct asymptotically. For steady state simulations the steady state fluctuation relation in terms of $\Sigma_A$ will converge at sufficiently long times. However for the times accessible to our simulations, it is a long way from convergence [see Fig. 6(b)]. (Remember too that as the averaging times increase, the amplitude of the fluctuations decrease, making it ever harder to verify the fluctuation relation in question.)

We have derived and numerically verified the TFT and SSFT for heat flow in a Newtonian system that is bounded by thermostatted walls driven to different but fixed temperatures. This work shows that the FT is applicable to thermal transport processes that are driven by boundary conditions rather than by external mechanical fields.

Our work points out that extreme care is required to measure the actual dissipation function. The actual dissipation function contains no nondissipative components. These nondissipative boundary components are expected to be ubiquitous in boundary driven thermal systems and become ever more dominant the closer the system is to thermodynamic equilibrium.13

The theorems are consistent with the Second Law of Thermodynamics. The FTs so derived give exact expressions for the probability that in a finite system observed for a finite time, heat will flow in the reverse direction to that required by Fourier’s law for heat flow and equivalently the Second Law of Thermodynamics (namely, from the cold toward the hot region). The expressions were tested and verified using numerical simulation of a two-dimensional fluid containing particles evolving in time according to purely Newtonian dynamics and interacting via a WCA pair potential.

For thermal conduction we have verified that a microscopic expression for the so-called entropy production deduced from linear irreversible thermodynamics is instantaneously equal to the weak field dissipation function multiplied by Boltzmann’s constant. Further these quantities are on average equal to the negative of the average rate of change of the fine grained Gibbs entropy of the system. As the system collapses toward a steady state attractor whose dimension is less than the ostensible dimension of phase space,7,11 the Gibbs entropy diverges at a constant rate toward negative infinity.

We remark that when Eq. (21) is used as the expression for the dissipation function, the resulting FT (2) does not refer to any thermostating variables [e.g., $\alpha_{\text{WAC}}(t)$]. This form of the FT refers only to natural thermodynamic functions (heat fluxes and bath temperatures) that are characteristic of a natural system undergoing heat flow. In this form, the FT will be independent of the precise nature of the thermostats used to remove heat at the walls. Therefore in this form, the FTs (2), (21), and (23) are valid in naturally occurring thermal conduction processes.

Previous approximations to the dissipation function for heat flow lead to the breakdown of the conditions for the TFT, generating instead an asymptotic TFR. The convergence times for these asymptotic relations diverge to infinity as equilibrium is approached.

It has been known previously that for systems that are driven away from equilibrium by mechanical fields, close to equilibrium the average of the dissipation function is equal to the average spontaneous entropy production. Now we know that the same is true for boundary driven heat flow. We now conjecture that this equality is always true. Thus it would seem that the dissipation function that is the argument of the
fluctuation, dissipation, and relaxation theorems is in fact a
generalized spontaneous entropy production valid arbitrarily
far from equilibrium.

Finally, we have given a “physicist’s proof” that for our
T-mixing boundary driven heat flow system, the system will
eventually relax to a unique steady state with time independent
ensemble averages for sufficiently smooth phase functions.
Further, this steady state will be ergodic over the initial
distribution (except for a set of initial phases of measure
zero). In this steady state the average dissipation must be
positive; it cannot be negative or zero. If the autocorrelation
function for the dissipation function is always positive, then
the steady state will be the state of maximal dissipation, but
in general where these autocorrelation functions are not posi-
tive for all times, there is no extremal property for the non-
equilibrium steady state.

ACKNOWLEDGMENTS

We would like to thank the Australian Research Council
for their support of this project. Very helpful discussions and
comments from Lamberto Rondoni and Rainer Klages are
also gratefully acknowledged.

1If the force acts to change the equilibrium state of the system, the dissi-
pative flux is calculated as the difference is the total flux from the equi-
librium value at the instantaneous value of the force. The dissipative flux
could thus be termed the “purely dissipative flux.” For example, the
dissipative flux for a shearing fluid is given by the viscous pressure
tensor, not the total pressure tensor that contains the diagonal contribu-
tions of the equilibrium hydrostatic pressure.

(1993).

3D. J. Evans and D. J. Searles, Phys. Rev. E 50, 1645 (1994); 52, 5839
112, 9727 (2000); E. M. Sevick, R. Prabhakar, S. R. Williams, and D. J.


6J. L. Lebowitz and H. Spohn, J. Stat. Phys. 95, 333 (1999); C. Maes,
ibid. 95, 367 (1999); J. Kurchan, J. Phys. A 31, 3719 (1998); F. Bonetto,
G. Gallavotti, and P. L. Garrido, Physica D 105, 226 (1997); F. Bonetto,

7J. D. Evans and G. P. Morriss, Statistical Mechanics of Nonequilibrium


11R. Klages, Microscopic Chaos, Fractals and Transport in Nonequilib-
rium Statistical Mechanics (World Scientific, Singapore, 2007) A. J. Li-
chtenberg and M. A. Lieberman, Regular and Chaotic Dynamics, 2nd ed.


Exp. 2009, P07029.

(2007).

16D. J. Evans, D. J. Searles, and S. R. Williams, J. Chem. Phys. 128,
014504 (2008).


18D. J. Evans and G. P. Morriss, Statistical Mechanics of Nonequilibrium

19G. Gallavotti and E. G. D. Cohen, J. Stat. Phys. 95, 333 (1999); C. Maes,
ibid. 95, 367 (1999); J. Kurchan, J. Phys. A 31, 3719 (1998); F. Bonetto,
G. Gallavotti, and P. L. Garrido, Physica D 105, 226 (1997); F. Bonetto,

20D. J. Evans and G. P. Morriss, Statistical Mechanics of Nonequilibrium


24R. Klages, Microscopic Chaos, Fractals and Transport in Nonequilib-
rium Statistical Mechanics (World Scientific, Singapore, 2007) A. J. Li-
chtenberg and M. A. Lieberman, Regular and Chaotic Dynamics, 2nd ed.


Exp. 2009, P07029.

(2007).

29D. J. Evans, D. J. Searles, and S. R. Williams, J. Chem. Phys. 128,
014504 (2008).


32D. J. Evans, D. J. Searles, and S. R. Williams, J. Chem. Phys. 128,
014504 (2008).

and D. J. Evans, J. Chem. Phys. 112, 9727 (2000); E. M. Sevick, R.
59, 603 (2008).


36J. L. Lebowitz and H. Spohn, J. Stat. Phys. 95, 333 (1999); C. Maes,
ibid. 95, 367 (1999); J. Kurchan, J. Phys. A 31, 3719 (1998); F. Bonetto,
G. Gallavotti, and P. L. Garrido, Physica D 105, 226 (1997); F. Bonetto,