Time-dependent fluctuation theorem

Emil Mittag and Denis J. Evans

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

(Received 23 September 2002; published 14 February 2003)

The fluctuation theorem (FT) is a generalization of the second law of thermodynamics that applies to small systems observed for short times. For thermostated systems it gives the probability ratio that entropy will be consumed rather than produced. In the present paper, we propose a version of the FT that applies to thermostated dissipative systems which respond to time-dependent dissipative fields. In testing the time-dependent fluctuation theorem we provide convincing evidence that sets of trajectories with conjugate values for the time-integrated entropy production, \( \pm \Delta \), are indeed (for time-reversible dynamical systems such as those studied here), time-reversal images of one another. This observation verifies the deep connection between time-reversal symmetry, the fluctuation theorem, and the second law of thermodynamics.

DOI: 10.1103/PhysRevE.67.026113

PACS number(s): 05.20.-y, 47.10.+g

INTRODUCTION

The fluctuation theorem (FT) gives a mathematical expression for the ratio of probabilities that, in a finite thermostated system observed for a finite time, the time-averaged irreversible entropy production \( \langle \Sigma \rangle \) will take on an arbitrary value \( A \), compared to \(-A\). The FT was first proposed by Evans, Cohen, and Morriss in 1993 [1]. The FT was then expressed as

\[
\frac{\text{Prob}(\Sigma_t/k_B = A)}{\text{Prob}(\Sigma_t/k_B = -A)} = \exp(At) \tag{1}
\]

Thus the probability that entropy will be produced rather than consumed increases exponentially with time and with system size. The theorem applies exactly to transient systems evolving from equilibrium at \( t = 0 \) toward a nonequilibrium steady state [2], and asymptotically \( (t \to \infty) \) to nonequilibrium steady states [1, 3].

The FT is important for several reasons. It expresses the probability that the second law of thermodynamics will be violated for a finite system observed for a finite time. It is one of the few exact mathematical expressions that is valid even far from equilibrium. Close to equilibrium, Green-Kubo relations can be derived from the FT [4]. It can also be used to derive expressions for free energy differences between two equilibrium systems, where the differences are computed using nonequilibrium path integration [5, 6].

Evans, Cohen, and Morriss originally proposed the FT for ergodic systems with constant-energy dynamics [1]. They showed that the FT was applicable to systems composed of a set of steady-state subtrajectories obtained from a single very long steady-state phase-space trajectory. Their heuristic derivation used Lyapunov weights for sampling phase-space trajectory segments. This version of the FT has since been denoted as the steady-state FT (SSFT) [7]. Evans and Searles [2] subsequently gave a derivation of the FT that used the Liouville measure for a microcanonical ensemble of systems where the entropy production was averaged over an ensemble of transient nonequilibrium trajectories spawned from a single equilibrium trajectory. This transient fluctuation theorem (TFT) was subsequently shown to be valid in many other ensembles and with different dynamics [8]. Later, Gallavotti and Cohen clarified the proof of the SSFT using the Sinai-Ruelle-Bowen measure [3]. Recently, a derivation of the TFT using local Lyapunov weights applied to arbitrary ensembles and dynamics has been given [9].

Many numerical simulations have been performed verifying the FT in various ensembles and with various dynamics [1–2, 4, 7–11]. The validity of the FT has been confirmed for systems in the absence of a thermostat [11] and, most recently, the FT was verified in the isobaric-isothermal ensemble [7]. Recently the TFT has been confirmed in a laboratory experiment using optical tweezers applied to a single colloidal particle in solution [12].

The most general (i.e., ensemble-independent) version of the TFT employs the so-called dissipation function [8]

\[
\tilde{\Omega}_s = \int_0^t ds \ln \left[ \frac{f(\Gamma(t),0)}{f(\Gamma(0),0)} \right] - \int_0^t ds \Lambda(\Gamma(s)) \tag{2}
\]

where \( f(\Gamma(0),0) \) is the phase-space distribution of the initial ensemble and \( f(\Gamma(t),0) \) is the initial probability density (i.e., at time \( t = 0 \)) at the evolved phase \( \Gamma(t) \). \( \Lambda(\Gamma) = \partial \Gamma / \partial \Gamma \) is the phase-space compression factor. This general dissipation function can be used to give a general expression for the fluctuation theorem:

\[
\frac{\text{Prob}(\tilde{\Omega}_s = A)}{\text{Prob}(\tilde{\Omega}_s = -A)} = \exp(At) \tag{3}
\]

For thermostated or ergostated systems, the dissipation function \( \Omega \) is recognizable as the rate of entropy absorption or production, \( \Sigma \), by the thermostat. Equation (3) has been tested via computer simulations for a range of ensembles with a large range of dynamics [1, 2, 4, 7–11]. With only a single exception [12], all these previous simulations have tested nonequilibrium systems subjected to time-independent...
external fields. In this paper we demonstrate the validity of the FT in nonequilibrium thermostated systems with time-dependent external fields.

Consider a system of \( N \) interacting particles subject to a time-dependent color field \( F_c(t) \). The total system Hamiltonian is \( H(\Gamma) = H_0(\Gamma) + F_c(t) \sum_{i,j} c_i \mathcal{H}_{ij} \), where \( c_i = (-1)^i \) is the color field coupling constant. \( H_0(\Gamma) = p_i^2/2m + \Phi(\mathbf{q}) \) is the internal energy of the system, with \( \Phi(\mathbf{q}) \) being the inter-particle potential energy. Interparticle interactions are modeled with the Weeks-Chandler-Andersen (WCA) potential [13] \( \Phi(\mathbf{q}) = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \varphi(|q_i - q_j|) \). \( \varphi(q) = 4[q^{-12} - q^{-6}] \), \( q < 2^{1/6} \) and zero otherwise.

The equilibrium \( N \)-particle phase-space distribution function is canonical and is given by \( f(\Gamma,0) \sim e^{-\beta[H_0 + (1/2)\mathbf{q}^2]} \). Here \( Q \) is the effective mass of a heat bath, \( \zeta \) is the Nose-Hoover thermostat multiplier [14], and \( \beta \) is the Boltzmann factor \( \beta = 1/k_BT = 2K/dN + O(1/N) \), where \( d \) is the Cartesian dimension. The equations of motion can be written as

\[
\begin{align*}
\dot{\mathbf{q}}_i &= \frac{\mathbf{p}_i}{m}, \\
\dot{\mathbf{p}}_i &= \mathbf{F}_i - ic_i F_c \mathbf{p}_i - \zeta \mathbf{p}_i, \\
\dot{\zeta} &= \frac{1}{Q} \left[ \sum \frac{p_i^2}{m} (g + 1)k_BT \right],
\end{align*}
\]

where \( \mathbf{F}_i = -\partial \Phi(\mathbf{q})/\partial q_i \), \( \omega \) is the frequency of the periodic external field, \( \phi(t) \) is a periodic function, \( \phi(t+P) = \phi(t) \), \( P = 2\pi/\omega \), and \( g = 6N + O(1) \) is the number of degrees of freedom in the system. The dissipative flux [15] for this system is \( H_0^{ad} = -JVF_c \), where \( V \) is the system volume, the superscript “\( ad \)” indicates that the time derivative of the Hamiltonian is taken in the absence of a thermostat, and \( J = V^{-1} \sum_{i=1}^{N} c_i \mathbf{p}_i \). We now substitute the initial phase-space distribution function of the system into the expression for the general dissipation function [Eq. (2)]. The general dissipation function for this system is then \( \Omega_j = -\beta(1/t)^{0} ds J(s) F_c(s) V = -\beta J(t) F_c(t) V \). Substituting this expression into Eq. (2) yields

\[
\ln \left[ \frac{\text{Prob}(\beta J(t) F_c(t) V = A)}{\text{Prob}(\beta J(t) F_c(t) V = -A)} \right] = \Delta t.
\]

A time-dependent TFT can exist only if three conditions are met. First, for every trajectory starting at a phase \( \Gamma(0) \), its conjugate antitrajectory must be observable among the initial ensemble of phases (i.e., the system must be ergodically consistent). This is a standard requirement for the applicability of the FT [7]. Secondly the conjugate trajectory \( \Gamma^*(t), t = P \) where \( M^T[\Gamma(t)] = \Gamma^*(P-t) \), must be a solution of the equations of motion. [A sufficient condition for this to occur is that the equations of motion are time reversible and therefore the time-dependent external field must have a definite parity under time-reversal symmetry (i.e., \( M^T[F_c(t)] = \pm F_c(P-t) \).]

**NUMERICAL RESULTS**

We test Eq. (5) via molecular dynamics simulations. In order to test the time-dependent FT, we use calculations that are identical to previous TFT simulations [2,7,8,11] except for the time dependence of the external field. Nonequilibrium side trajectories are periodically spawned from a main equilibrium trajectory. All trajectories are thermostated using a Nosé-Hoover thermostat, which, at equilibrium, generates a canonical distribution of phases. The time-dependent external field is activated at time \( t = 0 \) for each side trajectory and the response of the system is then monitored over the length of the side trajectory, \( P \). The time average of the dissipative flux is calculated for each transient trajectory and the ensemble average of the dissipative flux is then calculated from these time averages. The conditions for our test simulations are \( T = 1.0, N = 8, \) number density \( n = 0.4 \), time step = 0.001, \( P = 2.0 \), and \( F_c = 0.15 \). A step potential with odd parity was used, as shown in Fig. 1. Initially the external field is zero, then at time \( P/4 \) the field increases to \( F_c = 0.15 \), at \( P/2 \) the field changes to \( F_c = -0.15 \), and the field changes to zero at time \( 3P/4 \).

Figure 2 shows the full ensemble average of the transient responses, with the magnitude of the external field scaled by a factor of ten for convenience. The data are qualitatively as one would expect intuitively, or on the basis of the Maxwell model. The ensemble-averaged current is zero until the field...
is turned on, at which time the current rises abruptly. At $t = 1$, the current has not yet reached its steady-state value. However, at this time the field drops abruptly to $F_e = -0.15$ and the current immediately begins to fall in an approximately exponential fashion. The ensemble-averaged response is causal in character, with changes in the ensemble-averaged current taking place after the external field is changed. The ensemble-averaged data show no anticipation of future changes in the applied field.

In Fig. 3 we confirm that the fluctuation theorem is valid for this system. As expected, the FT is verified and confirmed for time-reversible systems with time-dependent external fields. The points near the ends of the curve may appear to diverge from the FT prediction. However, this is due to insufficient averaging for those points; they gradually converge as the number of transient trajectories in the simulation increases.

By histograming the responses on the basis of the time-averaged entropy production, we are able to directly compare the character of the response as a function of the time-averaged entropy production. Figure 4 shows a histogram of the time-averaged entropy production. As expected it is approximately Gaussian. The field is comparatively weak and the averaging time is short so the mean of the distribution, although positive, differs from zero by less than one standard deviation. We divide the area under the probability distribution function for the dissipation function to the right of the $y$ axis into bins. The area to the left of the $y$ axis is divided into correspondingly symmetric bins to those on the right.

By calculating the subensemble average of the dissipative flux of an individual bin, we can compare the second law satisfying subensemble-averaged response of a bin to the right of the $y$ axis with its conjugate second-law-violating response to the left of the $y$ axis. Figure 5 shows the subensemble-averaged response to the time-dependent external field for bins 1 and 1* of Fig. 4. The plot of the external field is scaled by a factor of 5 for convenience. As expected the second-law-satisfying response of bin 1 (shown as circles) as shown in Fig. 4 is related to the second-law-violating response of bin 1* (shown as crosses) by the transformation

$$\bar{J}_{tp} = -M^T(\bar{J}_{tp}).$$

The subensemble-averaged currents in conjugate bins are time-reversal maps of each other.
We note that the subensemble-averaged dissipative fluxes in conjugate bins both appear to respond to the change in the external field before that change takes place. This anticipatory response is due to a mixing of second-law-satisfying and second-law-violating characteristics within the subensemble averages for the bin. The trajectories are binned in terms of their time-integrated entropy production. The fact that the time-integrated entropy production is positive does not imply that for all times along a trajectory the entropy production is positive.

The anticausal character of the subensemble-averaged response for bins 1 and 1* seems to be significantly greater for the bin with a negative time-averaged entropy production, namely, bin 1*. Figure 6 shows a plot of the same data as in Fig. 5. Here, however, the data for bin 1* have been time-reversal mapped so as to be more readily comparable to the data for the conjugate bin 1. As expected, there is excellent agreement between the two curves.

Figure 7 shows the subensemble-averaged dissipative flux for bins 1 through 7 of Fig. 4. The magnitude of the response increases as the bin number increases. Bin 7 is therefore the one depicted with a dashed line with periodic solid circles. The data for all of the bins show considerable anticausal character. In fact all curves except the first one (obtained from bin 1) exhibit so much anticausal character that it is hard to say which curve is most anticausal in character.

The total ensemble-averaged response (i.e., the weighted response from all bins) must be causal in character and must be second-law satisfying. As we have seen, Fig. 2 confirms this. The full ensemble-averaged dissipative flux shown in Fig. 2 is the sum of the product of the subensemble-averaged dissipative flux in each bin multiplied by the weight of that bin. We can express this as

$$\langle J(t) \rangle = \sum_{i} w_i \langle J(t) \rangle_i,$$

where "bins" indicates that the summation is performed over all bins of the probability histogram and $w_i$ is the weight of bin $i$. We know that the ensemble-averaged response for a single bin is the time-reversal mapping of the response in the conjugate bin, i.e., $\langle J(t) \rangle_i = M^{T}_i \langle J(t) \rangle_{i*}$, where $i*$ denotes the bin that is conjugate to bin $i$. The total antiresponse is
reversal mapping to the data for the total forward response
mapped normal response obtained by applying the time-
weights for a bin and the subensemble-averaged current for
In other words, the time-reversal mapping of the full
response obviously satisfies the second law and is
pels this concern.

FIG. 8. A plot of the antiresponse of the dissipative flux to the
external field (solid line). The antiresponse (circles) is the time-
reversal map of the ensemble-averaged response, i.e., $M^T(J(t))$.
Also shown (crosses) are the results obtained from pairing the
weights of the response in each bin with the subensemble-averaged
response of the conjugate bin, Eq. (8).

$$M^T(J(t)) = \sum_i \sum_j w_i M_j^T(J(t)) = \sum_i w_i J(t)_{ij}$$

$$= \sum_i w_i J(t).$$

(8)

In other words, the time-reversal mapping of the full
ensemble-averaged response is the sum of the product of the
weights for a bin and the subensemble-averaged current for the
conjugate bin. Figure 8 shows the results of the application
of Eq. (8) (shown as crosses) and the time-reversal
mapped normal response obtained by applying the time-
reversal mapping to the data for the total forward response
(shown as circles). As expected from Eq. (8), the agreement
of the two curves is very good. Numerical error is respon-
sible for any difference between the curves.

CONCLUSION

We have shown that the fluctuation theorem is satisfied
for time-reversible, time-dependent systems. The fluctuation
theorem is therefore not restricted to systems with constant
dissipative fields. This further enhances the breadth of appli-
cability of the theorem.

The standard proof of the transient FT assumes that tra-
cjectories with conjugate values of the entropy production
($\pm A \pm \delta A$) are composed of pairs of trajectories and their
corresponding time-reversed antitrajectories. It has been ar-
geted that, although the existence of trajectory-antitrajectory
pairs is sufficient for the existence of a fluctuation theorem, it
may not be a necessary condition. It is possible that Eq. (3)
may be derived by means other than through the exploitation
of time-reversal symmetry. The present paper dispels this
conjecture. Figure 5, 6, and 7 give convincing evidence that
sets of trajectories with conjugate values for the time-
integrated entropy production ($\pm A \pm \delta A$) are indeed (for
time-reversible systems such as those studied here) time-
reversal images of one another.

It is also possible that, although trajectory conjugacy may
be necessary and sufficient for the existence of a fluctuation
theorem, as a practical matter the shear complexity of a
many-particle phase-space may be so great (with many non-
contiguous islands in the initial phase space having the same
value for the time-averaged entropy production) that it may
not be possible to actually observe time-reversed responses
for subsets of trajectories with conjugate values for the time-
integrated entropy production. Again the present work dis-
spels this concern.

Finally, this work shows that, although the total ensemble-
averaged response obviously satisfies the second law and is
completely causal in character, in general, the subensemble-
averaged currents, averaged over sets of trajectories with a
specified, time-averaged value of the entropy production, ex-
hbit mixed causal and anticausal character.