Fluctuation theorem for stochastic systems

Debra J. Searles  
Department of Chemistry, University of Queensland, Brisbane, Queensland 4072, Australia

Denis J. Evans  
Research School of Chemistry, Australian National University, Canberra, Australian Capital Territory 0200 Australia  
(Received 1 December 1998)

The fluctuation theorem describes the probability ratio of observing trajectories that satisfy or violate the second law of thermodynamics. It has been proved in a number of different ways for thermostatted deterministic nonequilibrium systems. In the present paper we show that the fluctuation theorem is also valid for a class of stochastic nonequilibrium systems. The theorem is therefore not reliant on the reversibility or the determinism of the underlying dynamics. Numerical tests verify the theoretical result. [S1063-651X(99)04007-6]

PACS number(s): 05.20.–y, 05.70.Ln, 47.10.+g, 47.70.–n

I. INTRODUCTION

The fluctuation theorem (FT) states that the ratio of the probability of observing nonequilibrium trajectory segments of duration $\tau$ with a time-averaged rate of entropy production per unit volume, $\bar{\sigma}_\tau$, to the probability of observing segments with an average entropy production rate per unit volume, $-\bar{\sigma}_\tau$, is

$$\ln \left( \frac{p(\bar{\sigma}_\tau)}{p(-\bar{\sigma}_\tau)} \right) = \frac{1}{k_B} \bar{\sigma}_\tau V \tau. \quad (1)$$

The FT is interesting in that it gives an analytic expression for the probability that, for a finite system and for a finite time, the second law of thermodynamics will be violated. This expression has been tested numerically and predicts the expected long-time and large-system behavior, that is second-law-violating trajectories will not be observed in the thermodynamic and/or long-time limits.

There are three approaches that have been used to derive this expression for deterministic systems. This relationship was derived in 1993 for nonequilibrium steady states by Evans, Cohen, and Morriss (ECM2) [1] from a natural invariant measure [2] which was proposed heuristically for steady-state trajectories. In 1995 Gallavotti and Cohen [3,4] gave a proof of the steady-state FT, demonstrating that this FT can be derived from the Sinai-Ruelle-Bowen (SRB) measure [5], if one employs the so-called chaotic hypothesis [3,4]. The steady-state FT (1) is only true asymptotically, $\tau \to \infty$. In ECM2 it was shown [1] that the steady-state fluctuation theorem (1) was consistent with computer simulation results for an atomic fluid undergoing reversible thermostatted shear flow far from equilibrium.

A transient FT had already been developed in 1994 [6–8]. It considers transient trajectories which are generated from initial phases sampled from an equilibrium microcanonical ensemble and which evolve in time towards the steady state which is assumed to be unique. Unlike steady-state FT’s, this transient fluctuation theorem is exact for all trajectory durations $\tau$. If long-time steady-state averages are independent of the initial phase vector used to generate a steady-state trajectory segment (for a given volume, number of particles, and energy), then one might expect that, in the limit $\tau \to \infty$, averages over transient trajectory segments would approach those taken over nonequilibrium steady-state segments. Further, one would expect the asymptotic convergence of the transient and steady-state fluctuation theorems. However, there has been some recent discussion of this point, and not all parties agree about this asymptotic convergence [9].

A third approach that has been used to derive the FT is valid only in the linear response regime close to equilibrium and is also only valid asymptotically ($\tau \to \infty$). In a footnote to ECM2 [1], the authors noted that the FT can be proved using the Green-Kubo relations for the linear response together with an application of the central limit theorem to the distribution of $\{\bar{\sigma}_\tau\}$, in the $\tau \to \infty$ limit [7,10]. This third approach, although limited to the linear response regime, is quite general with respect to the nature of the thermostating. In an obvious limit this approach applies to unthermostatted systems. This is because Green-Kubo relations are robust with respect to thermostating (see p. 116 of [11]).

Most theoretical and numerical studies of the FT have concentrated on reversible, deterministic dynamics although recently theoretical studies on stochastic systems have been carried out [12–14]. Kurchan [12] has shown that the FT is valid for Langevin dynamics, and Lebowitz and Spohn [13] showed that it could be extended to apply to steady-state Markov processes. Maes recently demonstrated [14] that a FT can be obtained if the steady state is regarded as a Gibbs state.

In the present paper, the transient FT is generalized so that it applies to stochastic systems. Furthermore, it is demonstrated that by considering the transient response of a system that is initially in a state with a known distribution function, rather than directly treating a steady-state system, a formula that is valid at all times is obtained. This approach is different from the steady-state approach of Lebowitz and Spohn [13] in that an exact, finite-time transient FT is derived. As in the deterministic case, if the steady state is unique, we expect that the transient FT will asymptotically converge to the steady-state stochastic FT. Also, we provide the first numerical tests of a stochastic FT. Given that the FT is valid for stochastic systems, reversibility and determinism are clearly not prerequisites for the FT.
II. DERIVATION OF THE FT USING THE LIOUVILLE MEASURE FOR A CLASS OF STOCHASTIC SYSTEMS

Consider the equations of motion for a stochastic system given by

\[
\dot{\mathbf{q}}_i = \frac{\mathbf{p}_i}{m} + C_i(\mathbf{q}) F_e, \\
\dot{\mathbf{p}}_i = \mathbf{F}_i(\mathbf{q}) + D_i(\mathbf{q}) F_e + \xi_i(t) - \alpha \mathbf{p}_i, 
\]

where \(\xi_i(t)\) is a random variable. The first term on the right-hand side, \(G(\mathbf{q})\), is deterministic and assumed to be reversible. As an example, consider the transient response of a system, initially at equilibrium, to an applied field with a random term \(\xi\), contributing to the equations of motion for the momenta. The system is thermostatted to ensure that a steady state can be reached, and the equations of motion are

\[
\dot{\mathbf{q}}_i = \frac{\mathbf{p}_i}{m} + C_i(\mathbf{q}) F_e, \\
\dot{\mathbf{p}}_i = \mathbf{F}_i(\mathbf{q}) + D_i(\mathbf{q}) F_e + \xi_i(t) - \alpha \mathbf{p}_i, 
\]

where \(\mathbf{q}_i\) and \(\mathbf{p}_i\) are the coordinates and momenta of the \(i\)th particle, respectively, \(\mathbf{F}_i\) is the interparticle force on that particle, \(F_e\) is an external field applied to the system, \(C_i\) and \(D_i\) describe the coupling of the system to the field, and \(\alpha\) is a Gaussian thermostat multiplier [11] that fixes the internal energy:

\[
\alpha = \left. \frac{N}{\sum_{i=1}^{N} p_i^2/m} \left[ \sum_{i=1}^{N} F_e D_i(\mathbf{q}) \cdot \frac{\mathbf{p}_i}{m} + \xi_i \cdot \frac{\mathbf{p}_i}{m} - F_e C_i(\mathbf{q}) \cdot \mathbf{F}_i(\mathbf{q}) \right] \right|_{t=0}^{t=\infty}.
\]

To ensure that the system remains on a constant-energy, zero-total momentum, hypersurface, the thermostat multiplier contains the random term and the restriction \(\sum_{i=1}^{N} \xi_i = 0\) is imposed. The phase space of the nonequilibrium system is therefore a subset of that of the initial equilibrium ensemble. In Eq. (3) the stochastic term can be regarded either as a random force that is added to the equation for the rate of change of momentum, or it can be regarded as contributing a random term to the thermostat. The difference between these two interpretations is purely semantic.

If the adiabatic incompressibility of phase space \((A/\Gamma)\) condition is satisfied [11], then the Liouville equation for this system reads:

\[
\frac{df(\Gamma,t)}{dt} = -f(\Gamma,t) \frac{\partial}{\partial \Gamma} \cdot \dot{\mathbf{r}} = -\Lambda(\Gamma)f(\Gamma,t),
\]

where \(\Lambda(\Gamma)\) is the phase-space compression factor. For the system described by Eqs. (3) and (4), \(\Lambda(\Gamma) = -dN\alpha(\mathbf{q}) + O(1)\), where \(d\) is the number of Cartesian coordinates considered. The solution of Eq. (5) can be written [7]

\[
f(\mathbf{q},t) = \exp \left[ -\int_0^t \Lambda(s) ds \right] f(\mathbf{q},0).
\]

The FT considers the probabilities of observing trajectories with entropy production rates which are equal in magnitude but opposite in sign. In the proof of this theorem using the Liouville measure [6,7,8,10], it was necessary, for every possible trajectory, to identify a conjugate trajectory which had this property: that is, all trajectories were sorted into conjugate pairs. In a reversible, deterministic system this identification was straightforward and was accomplished by carrying out time-reversal mappings [6,7,8,10]. It is now shown how this procedure can be modified for stochastic systems.

Consider a trajectory segment \(n(s), 0<s<t\), and its time-reversed trajectory \(n(s), 0<s<t\), which we call an antisegment. The sign in the subscript reflects the sign of the integral of the thermostat multiplier (or entropy production) along the trajectory segment. For a reversible system, these trajectories are simply related by a time reversal mapping: each conjugate trajectory \(n^+(s)\) is generated from the original trajectory \(n^-(s)\) by carrying out a time-reversal mapping of the phase at the midpoint of the trajectory and integrating the equations of motion backward and forward in time [6,7,8,10]. Without loss of generality if the field is assumed to be even with respect to the time-reversal mapping, then the flux, entropy production rate and the thermostat multiplier will odd. We use the notation that the averages of a phase variable, \(A\), along a forward trajectory and its conjugate, time-reversed, trajectory are given by

\[
\bar{A}_+(t) = \int_0^t ds A(n^+(s))
\]

and

\[
\bar{A}_-(t) = \int_0^t ds A(n^-(s)),
\]

respectively. Depending on the parity of the phase function \(A(\mathbf{q})\) under time-reversal symmetry, there may be a simple relation between \(\bar{A}_+(t)\) and \(\bar{A}_-(t)\).

In a stochastic system the conjugate trajectory can no longer be generated by simply carrying out a time-reversal mapping and solving the equations of motion. After the time-reversal mapping at the midpoint of the original trajectory, integration of the equations of motion forward and backward in time will, with overwhelming probability, result in the observation of a different set of random numbers than was observed for the original trajectory and the trajectories will not be conjugate. Clearly, a mapping of the sequence of random numbers observed for the forward trajectory must be carried out for the conjugate trajectory. The necessary mapping of the random numbers will depend on the function \(\xi(R)\) where \(R\) is a random number. Figure 1 gives a diagrammatic representation of the way in which conjugate trajectories are generated for stochastic systems.

If the sequence of random numbers \(R_1,R_2,R_3,R_4\) is observed for the original trajectory, then this sequence must be...
appropriately mapped in the conjugate trajectory. For example, if the random term contributes only to the equation of motion for $p_{x,t}$, which is even under a time-reversal mapping, then the sequence $R_4, R_3, R_2, R_1$ must be observed for the conjugate trajectory to be generated. Similarly, if the random term contributes only to $q_{x,t}$, which is odd under time reversal, then the sequence $-R_4, -R_3, -R_2, -R_1$ must be observed. Provided the mapped sequence is allowed by the random number generator, the antisegment will be a solution of the equations of motion. It should be noted that in the first case no restrictions on the random number generator are required; however in the second case, a symmetry restriction on the range of numbers is necessary for this process to be valid.

It can be observed from Fig. 1 that $M^{(T)} \Gamma_3 = \Gamma_4$, $M^{(T)} \Gamma_1 = \Gamma_6$, and $M^{(T)} \Gamma_2 = \Gamma_5$ where $M^{(T)}$ used to represent a time-reversal mapping. At all points along the trajectory, the fluxes of conjugate trajectories are related by $J(t; \Gamma_i, 0 < t < \tau) = -J(t; \Gamma_{i-1}, 0 < t < \tau)$, and therefore fluxes averaged over the duration of the segment are related by, $J_i = -J_{i-1}$. It is straightforward then to see that in order to divide the trajectories into conjugate pairs, the equations of motion do not have to be reversible [that is, $M^{(T)} e^{iL(t)t}, M^{(T)} e^{iL(t)t}$, $\Gamma(0) = \Gamma(0)$, where $M^{(T)} M^{(T)} = 1$], but it is necessary that the antisegment be a solution of the equations of motion. This condition is equivalent to that required in the derivation by Lebowitz and Spohn [13], in which case it is assumed that if the rate constant for a forward step is nonzero, then the rate of the reverse process must also be nonzero. That is, in both derivations it is required that the reverse process be able to be observed.

Now it is shown how the probability of observing the conjugate trajectories can be determined. For the system considered, the initial phases are distributed microcanonically; so the probability of observing an initial phase inside a small phase volume, $\delta V(\Gamma(0))$ about $\Gamma(0)$ is proportional to

$$V(\Gamma(t), t) = \exp \left[ \int_0^t \Lambda(\Gamma(s)) ds \right] V(\Gamma, 0). \tag{9}$$

This is illustrated in Fig. 2.

The ratio of volumes the $\delta V_1$ and $\delta V_4$ gives the ratio of the probability of observing initial phase points. The probability of observing a trajectory is equal to the product of the probability of observing the initial phase point and the probability of observing the sequence of random numbers:

$$\text{prob}(\Gamma(s); 0 < s < t) = \text{prob}(\delta V(0)) \text{prob}(R_1 \cdots R_n). \tag{10}$$

The probability of observing a trajectory segment with a particular time-averaged value of $\Lambda$ is then given by the sum over all trajectories with that value, and the probability ratio is given by

\[
\begin{align*}
p(\bar{\lambda}_+(\tau)) &= \sum_{i:fA(\Gamma_i) = \bar{\lambda}_+(\tau)} \delta V(\Gamma_i + (0)) p(R_1, \ldots, R_n)_i; \\
p(\bar{\lambda}_-(\tau)) &= \sum_{i:fA(\Gamma_i) = \bar{\lambda}_-(\tau)} \delta V(\Gamma_i + (0)) p(R_1, \ldots, R_n)_i; \\
&= \sum_{i:fA(\Gamma_i) = \bar{\lambda}_-(\tau)} \delta V(\Gamma_i - (0)) p(M^{(T)}(R_1 \cdots R_n)_i); \\
&= \sum_{i:fA(\Gamma_i) = \bar{\lambda}_+(\tau)} \exp [\bar{\lambda}_+(\tau) \tau] \delta V(\Gamma_i + (0)) p(M^{(T)}(R_1 \cdots R_n)_i), \\
&= \exp [-\bar{\lambda}_+(\tau) \tau], \tag{11}
\end{align*}
\]
where the notation $i \int \Lambda(\Gamma_i) = \bar{\Lambda}_+(\tau)$ is used to represent all trajectories $i$ for which the time-averaged value of the phase-space compression factor is equal to $\bar{\Lambda}_+$, and it is assumed that $p(R_1, \ldots, R_n) = p(M(T)(R_1, \ldots, R_n))$. The resulting fluctuation formula given by Eq. (11) for this stochastic system is identical to that for the deterministic, reversible systems.

As in the deterministic case, there may be many different pairs of conjugate trajectories which each have the same value for $\bar{\Lambda}_+(\tau)$.

The FT derived above is valid for transient trajectory segments of arbitrary length. Averages of phase variables over the transient trajectory segments approach the averages over steady-state trajectory segments in the long-time limit; therefore, the stochastic FT will also apply to steady-state systems.

III. NUMERICAL TESTS OF THE FLUCTUATION THEOREM APPLIED TO TRANSIENT STOCHASTIC SYSTEMS

Transient NEMD simulations of Couette flow using the SLLOD algorithm and the usual Lees-Edwards periodic boundary conditions, were carried out employing a stochastic force in the $x$ direction and the corresponding Gaussian isoenergetic thermostat. The equations of motion for this system are

\[
\dot{q}_i = p_i + i \gamma \gamma_i, \\
\dot{p}_i = F_i - i \gamma p_i + i \gamma_i - \alpha p_i, \tag{12}
\]

with the thermostat multiplier given by

\[
\alpha = \frac{\sum_{i=1}^{N} \dot{\gamma}_i p_{x_i} - \gamma(p_{x_1} p_{x_1} + F_{x_1} \gamma_1)}{\sum_{i=1}^{N} p_i \cdot p_i}. \tag{13}
\]

The system consisted of $N=32$ particles in two Cartesian dimensions and the particles interacted with the Weeks-Chandler-Anderson short-ranged, repulsive pair potential [15]. Lennard-Jones units are use throughout. The internal energy per particle was set at $E/N=1.560$ 32 (i.e., $T \sim 1.0$) and the particle density at $n=0.8$. A strain rate of $\gamma = 0.5$ was applied.

FIG. 3. Ensemble-averaged response of the flux for a system of 32 particles in two Cartesian dimensions to which a strain rate is applied at time zero and for which the SLLOD algorithm is used to model the shear flow. The internal energy per particle was set at $E/N=1.560$ 32 (i.e., $T \sim 1.0$) and the particle density at $n=0.8$. A strain rate of $\gamma = 0.5$ was applied.

FIG. 4. Histograms of $\bar{\Lambda}_+(\tau) = (1/\tau) \int_0^\tau \alpha(\Gamma_i(s)) ds$ for a system undergoing transient response to an applied strain rate of $\gamma = 0.5$. The internal energy per particle was set at $E/N=1.560$ 32 (i.e., $T \sim 1.0$) and the particle density at $n=0.8$. Trajectory segments of (a) $\tau = 0.1$, (b) $\tau = 0.4$, and (c) $\tau = 0.6$ were used.

Since for this system $\Lambda(\Gamma) = -2N \alpha(\Gamma) + O(1) = \alpha(\Gamma) V/k_B$, the fluctuation theorem becomes

\[
\frac{p(\bar{\Lambda}_+(\tau))}{p(\bar{\Lambda}_-(\tau))} = \exp[2N \bar{\Lambda}_+(\tau) \tau], \tag{14}
\]

where $O(1)$ terms are omitted since they are negligible in the thermodynamic limit and to reduce the complexity of the expression. The system studied here is sufficiently small that these effects cannot be neglected, and they are included in the data presented.
In the simulations, the stochastic term \( \xi_i \) was the product of a random number \( \xi \) and a \( \delta \) function at each time step. The random numbers were selected from a Gaussian distribution with \( \sigma = 0.5 \), a standard deviation of 1.0, and were restricted within the range \([-10.0, 10.0]\).

Figure 3 shows the ensemble averaged response of the flux which indicates that a steady state is approached and the initial transient response has a Maxwell time of approximately \( \tau_M \approx 0.07 \).

In the transient response simulations, many initial equilibrium phases were generated and the response to an applied strain rate was monitored for various trajectory segment lengths. Histograms of

\[
\bar{\alpha}_+(\tau) = 1/\tau \int_0^\tau ds \alpha(G_+(s))
\]

were obtained with \( \tau = 0.1, 0.4, \) and \( 0.6 \). These histograms are shown in Fig. 4. The FT predicts that a plot of \( \ln[p(\bar{\alpha}_+(\tau))/p(-\bar{\alpha}_+(\tau))] / (2N\tau) \) versus \( \bar{\alpha}_+(\tau) \) should give a straight line of unit slope. For each of the trajectory segment lengths considered in Fig. 3, the FT was tested with \( O(1/N) \) corrections included. The normalized probability ratios are shown in Fig. 5. In each case, a slope of unity is obtained and the FT is verified.

These results show that the FT is valid for finite averaging times of this transient, stochastic system.
0.2, 0.3, and 0.4 were calculated. The results for the results, and Fig. 7 plots the slope as a function of in the limit straight line of unit slope.

The FT was also examined for steady-state systems evolving with the stochastic equations of motion considered in Sec. III. Histograms of

$$\bar{\alpha}_+(\tau) = \frac{1}{\tau} \int_0^\tau ds \alpha(\Gamma_+(s))$$

for steady-state trajectory segments of length $\tau = 0.05, 0.1, 0.2, 0.3,$ and 0.4 were calculated. The results for $\tau = 0.05, 0.2,$ and 0.4 are shown in Fig. 6.

For steady-state trajectories, the FT predicts that a plot of $\{\ln[p(\bar{\alpha}_+(\tau))/p(-\bar{\alpha}_+(\tau))]/(2N\tau)\}$ versus $\bar{\alpha}_+(\tau)$ gives a straight line of unit slope in the limit $\tau \to \infty$. Figure 6 shows the results, and Fig. 7 plots the slope as a function of $\tau$, indicating that it approaches unity in the long-time limit.

V. CONCLUSIONS

The present work shows that the fluctuation theorem is quite general and applies to both deterministic and stochastic nonequilibrium systems. As was found to be the case for deterministic systems, the fluctuation theorem applies (i) at all times to finite (transient) trajectory segments which are initially sampled from the equilibrium microcanonical ensemble and then move isoenergetically towards a steady state and (ii) asymptotically to long-time steady-state trajectory segments. In all cases—transient or steady state, stochastic, or deterministic—the fluctuation theorem applies in both the linear and nonlinear response regimes.

As a final comment, we note that although the theory and simulations presented here apply to systems in which every particle is ergostatted (so-called homogeneous thermostatting), the theory presented here applies equally well to systems where a fluid obeying Newtonian mechanics (i.e., no thermostat) flows inside thermostatted solid walls, using the theory presented here. To treat such a system, consider a mixture of two types of particles where at the temperature and density studied one set of particles, the wall particles, is in the solid phase and is thermostatted and the other set of particles is liquid and is not thermostatted. In such cases the only difference to the theory above is that in equations such as Eq. (6), above the $N$ refers to the number of thermostatted particles and not to the total number of particles.

ACKNOWLEDGMENTS

We would like to thank the Australian Research Council for the support of this project. Helpful discussions and comments from Professor E. G. D. Cohen are also gratefully acknowledged.