Extension of the Fluctuation Theorem

R. van Zon and E. G. D. Cohen
The Rockefeller University, 1230 York Avenue, New York, New York 10021, USA
(Received 21 May 2003; published 8 September 2003)

Heat fluctuations are studied in a dissipative system with both deterministic and stochastic components for a simple model: a Brownian particle dragged through water by a moving potential. An extension of the stationary state fluctuation theorem is derived. For infinite time, this reduces to the conventional fluctuation theorem only for small fluctuations; for large fluctuations, it gives a much larger ratio of the probabilities of the particle to absorb rather than supply heat. This persists for finite times and should be observable in experiments similar to a recent one carried out by Wang *et al.*

DOI: 10.1103/PhysRevLett.91.110601 PACS numbers: 05.40.–a, 02.50.–r, 05.70.–a, 44.05.+e

There is a lack of unifying principles in nonequilibrium statistical mechanics, compared to the equilibrium case. So it is not surprising that the *fluctuation theorem* has received a lot of attention, as it gives a property of fluctuations of entropy production for a large class of systems, possibly arbitrarily far from equilibrium [1–10].

Here, we prefer to use the term heat rather than entropy production, so that the conventional stationary state fluctuation theorem (SSFT) [11] states that the probability $P_\tau(Q_\tau)$ to find a value of Q_τ for the amount of heat dissipated in a time interval τ satisfies, in a nonequilibrium stationary state [1,3],

$$\frac{P_\tau(Q_\tau)}{P_\tau(-Q_\tau)} \sim e^{\beta Q_\tau}, \tag{1}$$

where $\sim$ indicates the behavior for large τ. Here $\beta = (k_B T)^{-1}$, with k_B Boltzmann's constant and T the (effective) temperature of the system. In contrast, the transient fluctuation theorem (TFT) considers fluctuations Q_τ in time, when the system is initially in equilibrium [2]. These theorems were first demonstrated in (isoenergetic) deterministic many particle systems in an external field [1–3], but later in stochastic systems as well [4,5].

In contrast to our formulation in terms of heat, in the literature, Eq. (1) is interpreted as a theorem for fluctuations of entropy production far from equilibrium, by identifying Q_τ/T as a (generalized) entropy production. Furthermore, the fluctuation theorem holds for arbitrary values of Q_τ, i.e., also far from its average. Hence, it is often referred to as a large deviation theorem.

Recently, a laboratory experiment was carried out by Wang *et al.* [8]. They measured fluctuations in the work done on a system in a transient state of a Brownian particle in water, subject to a moving, confining potential. The TFT for work fluctuations was confirmed.

In the deterministic models, dissipation is often modeled by including a damping term in the equations of motion, chosen such as to keep, e.g., the total energy of the system fixed. This damping term is a mechanical expression for what in reality is an external thermostat. Any work done by the external field on the system is absorbed by this "internal" thermostat. Therefore, the external work done on the system equals the heat dissipated in the system.

While Wang *et al.* intended to study the entropy production (or heat) fluctuations, in fact, the work fluctuations were studied [10]. In contrast to the above sketched purely deterministic models, work fluctuations differ from heat fluctuations in their system due to the presence of a confining potential. Thus, some of the external work done is converted into potential energy and only the rest is converted into heat. In fact, in this system the work fluctuations in the stationary state satisfy the conventional SSFT [10,12], but the heat fluctuations do not, as we shall show. As it turns out, in the presence of a deterministic together with a stochastic component, the resulting behavior of the heat fluctuations coincides with the conventional SSFT only for a restricted set of small fluctuations while the behavior is very different for larger ones.

We first discuss the work-related fluctuation theorem in the experiment of Wang *et al.*, which was treated theoretically in Refs. [10,12,13]. The Brownian particle in a fluid, subject to a harmonic potential moving with constant velocity $\mathbf{v}^*$, was described by an overdamped Langevin equation:

$$\frac{d\mathbf{x}_t}{dt} = -(\mathbf{x}_t - \mathbf{x}_t^*) + \zeta_t. \tag{2}$$

Here, $\mathbf{x}_t$ is the position of the particle at time t, $\mathbf{x}_t^* = \mathbf{v}^* t$ is the position of the minimum of the harmonic potential at time t, and ζ_t is a fluctuating force with zero mean and a delta function correlation in time. We remark that the relaxation time of the position of the particle has been set equal to one. Also, we set $k_B T = 1$, so $\langle \zeta_t \zeta_s \rangle = 2\delta(t - s)$ [14]. In Ref. [10], it was shown that Eq. (2) is solvable in a comoving frame, in which it reduces to a standard Ornstein-Uhlenbeck process. Thus, the stationary probability distribution and Green's function are known, and are both Gaussian in $\mathbf{x}_t$. The work is the total amount of energy put into the system in a time τ. This is a fluctuating quantity, given by [15]

$$W_\tau \equiv \mathbf{v}^* \cdot \int_0^\tau [-(\mathbf{x}_t - \mathbf{x}_t^*)]\, dt. \tag{3}$$

Here, the time $t = 0$ denotes the initial time of an interval of length τ in the stationary state. W_τ is a linear function of the positions $\mathbf{x}_t$ and, since those have a Gaussian probability distribution function, so does W_τ. When the mean and variance of the probability distribution function P_τ^W are computed [using the stationary solution and Green's function of Eq. (2)], one finds [10]

$$\lim_{\tau\to\infty} \frac{1}{w\tau} \ln\left[\frac{P_\tau^W(pw\tau)}{P_\tau^W(-pw\tau)}\right] = p. \tag{4}$$

Here, p is a scaled value of W_τ, defined as $p = W_\tau/\langle W_\tau\rangle$, such that $\langle p\rangle = 1$. We also wrote

$$\langle W_\tau\rangle = w\tau, \tag{5}$$

with w the average work rate, which is independent of τ in the stationary state. In the current units, $w = |\mathbf{v}^*|^2$. Equation (4) is, for the work fluctuations, a more careful formulation of the SSFT in Eq. (1). A work-related TFT also holds [10,12,13].

We now turn to the heat SSFT. The heat Q_τ is that part of the work W_τ that goes into the fluid. Some work is also stored in the potential, so

$$Q_\tau \equiv W_\tau - \Delta U_\tau, \tag{6}$$

where ΔU_τ is the change in potential energy of the particle in a time τ,

$$\Delta U_\tau \equiv U_\tau - U_0, \tag{7}$$

with $U_t \equiv \frac{1}{2}|\mathbf{x}_t - \mathbf{x}_t^*|^2$. This form of U_t makes Q_τ nonlinear in $\mathbf{x}_t$. As a result, the probability distribution function $P_\tau(Q_\tau)$ of Q_τ need not be Gaussian. Nonetheless, it is possible to compute its Fourier transform.

The Fourier transform of $P_\tau(Q_\tau)$, defined as

$$\hat{P}_\tau(q) \equiv \int_{-\infty}^{\infty} dQ_\tau\, e^{iqQ_\tau} P_\tau(Q_\tau), \tag{8}$$

is computed by writing P_τ as [using Eqs. (6) and (7)]

$$P_\tau(Q_\tau) = \iint d\mathbf{x}_0\, d\mathbf{x}_\tau\, P_\tau^{W_\tau,\mathbf{x}_0,\mathbf{x}_\tau}(Q_\tau + \Delta U_\tau, \mathbf{x}_0, \mathbf{x}_\tau), \tag{9}$$

where $P_\tau^{W_\tau,\mathbf{x}_0,\mathbf{x}_\tau}$ is the joint distribution of the work W_τ, the positions $\mathbf{x}_0$ and $\mathbf{x}_\tau$ at the beginning and at the end of the time interval τ, respectively. This distribution is Gaussian because W_τ, $\mathbf{x}_0$, and $\mathbf{x}_\tau$ are all linear in $\mathbf{x}_t$. When Eq. (9) is inserted into Eq. (8), a seven-dimensional Gaussian integral is left, which after some algebra yields

$$\hat{P}_\tau(q) = \frac{\exp\{wq(i-q)[\tau - \frac{2q^2(1-e^{-\tau})^2}{1+(1-e^{-2\tau})q^2}]\}}{[1+(1-e^{-2\tau})q^2]^{3/2}}. \tag{10}$$

Once $\hat{P}_\tau(q)$ has been transformed back, one considers

$$f_\tau(p) \equiv \frac{1}{w\tau} \ln\left[\frac{P_\tau(pw\tau)}{P_\tau(-pw\tau)}\right]. \tag{11}$$

Here, p is a scaled value of Q_τ, defined as $p = Q_\tau/\langle Q_\tau\rangle$, i.e., $\langle p\rangle = 1$. We also used $\langle Q_\tau\rangle = \langle W_\tau\rangle - \langle \Delta U_\tau\rangle = w\tau$ by Eq. (5), since $\langle \Delta U_\tau\rangle = 0$ in the stationary state. The scaled logarithmic ratio $f_\tau(p)$ should be equal to p for $\tau \to \infty$ when the conventional SSFT holds.

As far as we know, there is no exact result for the inverse Fourier transform of $\hat{P}_\tau(q)$ in Eq. (10) in terms of known functions. Therefore, a completely analytic treatment did not seem feasible. Instead, we used first a numerical method, the fast Fourier transform algorithm [16], to invert Eq. (10). The resulting probability distribution function P_τ as well as the corresponding f_τ have been plotted in Fig. 1. These results do not agree very well with the straight line with slope 1, which should be approached for large τ if the conventional SSFT were to hold. One might think that this is due to τ not being large enough. However, we found that deviations of $f_\tau(p)$ from p for large p are generic, while the straight line is approached only for p values of smaller magnitudes ($|p| \lesssim 1$). Nonetheless, we cannot say anything conclusive about the large τ, large p behavior because the distribution gets very peaked and, hence, becomes smaller for large deviations, which makes the numerical method unreliable.

Therefore, we used next an analytical asymptotic approach based on large deviation theory [17] similar to the treatment by Lebowitz and Spohn [5]. One considers then

$$e(\lambda) \equiv \lim_{\tau\to\infty} -\frac{1}{w\tau} \ln\langle e^{-\lambda Q_\tau}\rangle. \tag{12}$$

This infinite-τ quantity is used to reconstruct the distribution function of Q_τ for large τ by setting

$$P_\tau(Q_\tau) \sim \exp[-w\tau\hat{e}(Q_\tau/w\tau)], \tag{13}$$

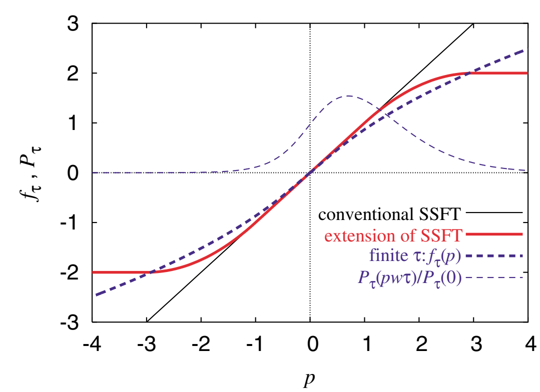

FIG. 1 (color online). Numerically obtained $f_\tau(p)$ (bold dashed line) for $v^* = 1.5$ and $\tau = 1.3$, and the (v^* independent) extension of the SSFT for $\tau \to \infty$ (bold solid line). Also plotted are the conventional SSFT (thin solid line), and the numerically obtained distribution function $P_\tau(pw\tau)$, scaled by its value at zero (thin dashed line).

where $\hat{e}(p)$ is the Legendre transform of $e(\lambda)$:

$$\hat{e}(p) = \max_{\lambda}[e(\lambda) - \lambda p]. \tag{14}$$

For a class of models, Lebowitz and Spohn proved the symmetry relation

$$e(\lambda) = e(1 - \lambda). \tag{15}$$

From this, using Eqs. (11), (13), and (14), one sees that $\lim_{\tau\to\infty} f_\tau(p) = p$, i.e., the conventional SSFT holds [5].

Our numerical results suggest, however, that for our model the conventional SSFT for the heat does not hold. We therefore expect Eq. (15) to be violated. Indeed, the following calculation of $e(\lambda)$ shows this to be the case.

The Fourier transform $\hat{P}_\tau(q)$ in Eq. (10) determines $e(\lambda)$. First, from Eq. (8), we have

$$\langle e^{-\lambda Q_\tau}\rangle \equiv \int_{-\infty}^{\infty} dQ\, e^{-\lambda Q_\tau} P_\tau(Q_\tau) = \hat{P}_\tau(i\lambda), \tag{16}$$

Thus, we need the analytic continuation of $\hat{P}_\tau$ to imaginary arguments. This poses no difficulty as long as $\hat{P}_\tau$ remains analytic. One finds from Eqs. (10) and (16)

$$\langle e^{-\lambda Q_\tau}\rangle = \frac{\exp[-w\lambda(1-\lambda)\{\tau + \frac{2\lambda^2(1-e^{-\tau})^2}{1-(1-e^{-2\tau})\lambda^2}\}]}{[1-(1-e^{-2\tau})\lambda^2]^{3/2}}. \tag{17}$$

Clearly, there are divergences at the singular points $\lambda = \pm(1-e^{-2\tau})^{-1/2}$, where the right-hand side (r.h.s.) of Eq. (16) is no longer analytic, so that Eq. (17) only holds for values of λ in between those. Using Eqs. (12) and (17), we have

$$e(\lambda) = \lambda(1-\lambda) \qquad \text{for } |\lambda| < 1, \tag{18}$$

where, taking $\tau \to \infty$ as in Eq. (12), moves the singularities to ± 1. This $e(\lambda)$ satisfies Eq. (15) for $0 < \lambda < 1$.

However, as λ approaches the singularities, the function in Eq. (17) diverges. Beyond the singularities at $\pm(1-e^{-2\tau})^{-1/2}$, the r.h.s. of Eq. (17) becomes purely imaginary, and multivalued due to the denominator. But the left-hand side of Eq. (17) remains real. Clearly, we cannot use Eq. (17) for $|\lambda| > (1-e^{-2\tau})^{-1/2}$. To determine $\langle e^{-\lambda Q}\rangle$ in that case, we first need to know why the integral in Eq. (16) diverges as $\lambda \to \pm(1-e^{-2\tau})^{-1/2}$. As we will argue next, this happens because P_τ has exponential tails. Since $P_\tau(Q_\tau)$ is a normalized distribution and $e^{-\lambda Q_\tau}$ a regular function, the divergence in Eq. (16) can only be due to the behavior of the integrand at $\pm\infty$. In fact, for $\lambda > 0$ any divergence must be due to the behavior at negative Q_τ and, for $\lambda < 0$ it must be due to the behavior at positive Q_τ. Now, for $\lambda < 0$, if the distribution function $P_\tau(Q_\tau)$ fell off faster than exponential for large positive Q_τ, the factor $e^{-\lambda Q_\tau}$ could not make the integral diverge. As it does diverge, we conclude that the distribution function falls off exponentially or slower. On the other hand, if it did fall off slower than exponential, then the exponential factor $e^{-\lambda Q_\tau}$ would always dominate the distribution function for large positive Q_τ and the integral would diverge for all $\lambda < 0$. Since there are negative values of λ for which the integral converges, the function P_τ cannot fall off slower than exponential. Hence, it must fall off exponential for large Q_τ. Considering $\lambda > 0$, one deduces along similar lines that it must also fall off exponentially for large negative values of Q_τ.

In fact, the integral in Eq. (16) diverges for all $|\lambda| \geq (1-e^{-2\tau})^{-1/2}$. If the function $P_\tau(Q_\tau)$ falls off exponentially for large positive Q_τ, say as e^{-aQ_τ}, the integral in (16) diverges for all $\lambda \leq -a$. Likewise, given that the $P_\tau(Q_\tau)$ falls off similar to e^{aQ_τ} for large negative Q_τ, the integral diverges for all $\lambda \geq a$. Hence, for $|\lambda| \geq (1-e^{-2\tau})^{-1/2}$, the quantity on the r.h.s of Eq. (12), of which the limit is taken, is minus infinity for all τ, so that $e(\lambda) = -\infty$. Thus, Eq. (18) becomes

$$e(\lambda) = \begin{cases} \lambda(1-\lambda) & \text{for } |\lambda| < 1 \\ -\infty & \text{otherwise.} \end{cases} \tag{19}$$

This $e(\lambda)$ does not satisfy the symmetry relation in Eq. (15), e.g., for $\lambda = -1/2$, $e(\lambda) = -3/4$, whereas $e(1-\lambda) = -\infty$. The fact that Eq. (15) is not satisfied means that the conventional SSFT does not hold. To make this more precise, we use Eqs. (14) and (19) to find

$$\hat{e}(p) = \begin{cases} -p & \text{for } p < -1 \\ (p-1)^2/4 & \text{for } -1 \leq p \leq 3 \\ p-2 & \text{for } p > 3. \end{cases} \tag{20}$$

Note that via Eq. (13), the large $|p|$ behavior is indeed exponential [18]. Using Eqs. (11), (13), and (20), we find

$$\lim_{\tau\to\infty} f_\tau(p) = \begin{cases} p & \text{for } 0 \leq p < 1 \\ p - (p-1)^2/4 & \text{for } 1 \leq p < 3 \\ 2 & \text{for } p \geq 3. \end{cases} \tag{21}$$

For negative p, we have $f_\tau(-p) = -f_\tau(p)$. Equation (21) is an extension of the conventional SSFT. It coincides with it for the middle region $-1 < p < 1$ [19], but differs from it for other p values. Most notably, for $p \geq 3$, it attains a constant value of 2.

If we compare the exact prediction of Eq. (21) (plotted as the bold solid line) with the numerical results (bold dashed line) in Fig. 1, a clear discrepancy emerges: The curve of f_τ keeps increasing with increasing p, whereas Eq. (21) predicts that it should level off to a value of 2. This turns out to be a finite τ effect. To prove this, we need a better treatment for large but not infinite τ. This can be obtained from a saddle-point method applied to $e(\lambda)$, which we will present in a future publication [20]. The saddle-point method gives reliable results for sufficiently large τ, as can be verified by a comparison to our numerical results [20]. The asymptotic behavior for large τ is then given by

$$f_\tau(p) = \begin{cases} p + h(p)/\tau + O(\tau^{-2}) & \text{for } p < 1 \\ 2 + \sqrt{8(p-3)/\tau} + O(\tau^{-1}) & \text{for } p > 3, \end{cases} \tag{22}$$

where $h(p) = [(8p)/(9 - p^2)] - \frac{3}{2w}\ln\{[(3 - p)(1 + p)]/[(3 + p)(1 - p)]\}$. Being mainly interested in large and small p, we left out the behavior in between $p = 1$ and $p = 3$. Equation (22) shows that, as a function of p, $f_\tau(p)$ increases $\sim\sqrt{p - 3}$ for fixed τ, while as a function of τ it decreases for fixed p. As expected, for $\tau \to \infty$, it approaches the large deviation result Eq. (21) as τ^{-1} for small p and as $\tau^{-1/2}$ for $p > 3$.

Whether the new features beyond $p = 1$ are observable depends on the values of $P_\tau(pw\tau)$ and $P_\tau(-pw\tau)$. If they are too small, the corresponding $f_\tau(p)$ will not be observed in an experiment. The value of the distribution function as plotted in Fig. 1 is non-negligible for values for p (and $-p$) at which f_τ bends away from the conventional SSFT. Furthermore, the values $v^* = 1.5$ and $\tau = 1.3$ used in Fig. 1 are realistic, as in the experiment of Wang *et al.* $v^* \approx 2.5$ and τ goes up to ≈ 6. So this behavior should be experimentally detectable.

In summary, we have shown that the behavior of heat fluctuations in a dissipative system with a deterministic component (the potential), and a stochastic component (the heat bath, i.e., the water), differs from that known from previous studies, in two respects. (i) For infinite τ, the behavior of the conventional SSFT is seen only for the scaled heat fluctuation p between -1 and 1. For $p > 1$, after a parabolic region between $p = 1$ and 3, the quantity f_τ no longer increases, but stays at a plateau value of 2 (similarly, for $p < -1$, by antisymmetry of f_τ). (ii) The finite τ behavior of the conventional SSFT is in general unknown, but in our case we find that f_τ keeps increasing with p. However, f_τ stays well below the conventional SSFT, implying a larger ratio of the probabilities of the particle to absorb rather than supply heat. These features are observable.

One of the striking features of the extension of the SSFT, the plateau value of 2 for large (infinite) τ and large p, can be understood physically. For large τ and large Q_τ (i.e., p) the exponentially distributed ΔU_τ far outweighs the Gaussian distributed W_τ in Eq. (6). The distribution of ΔU_τ is exponential for large values ($\propto e^{-\beta|\Delta U_\tau|}$), since it is the difference of the potential energies of the particle at two times [cf. Equation (7)], which are both Boltzmann-like distributed (due to the presence of the water) and independent of each other for large τ. As $\langle Q_\tau\rangle = w\tau$, this leads to $P_\tau(Q_\tau) \propto e^{-\beta|Q_\tau - w\tau|}$, which yields $P_\tau(Q_\tau)/P_\tau(-Q_\tau) \approx e^{2\beta w\tau}$ (if $Q_\tau > 0$). Then Eq. (11) and $\beta = 1$ give $f_\tau \approx 2$ of Eq. (21).

We only considered an extension of the SSFT here. An extension of the TFT can also be obtained. While for $\tau \to \infty$, one gets again Eq. (21), for finite times, the extensions of TFT and SSFT differ [20]. Furthermore, the extension of the TFT differs fundamentally from the conventional TFT, in that the latter holds as an identity for all τ [6], but the former holds only in the $\tau \to \infty$ limit.

One may wonder how general our extension of the SSFT is. Our Langevin-based theory is applicable only near equilibrium. The arguments above suggest that the extension could also hold for other potentials. Perhaps it even holds for a larger class of systems, not near equilibrium, with deterministic and stochastic components, as these are the main physical ingredients in our theory.

We thank M. J. Feigenbaum, F. Bonetto, S. Ciliberto, and L. Rondoni for helpful discussions. This work was supported by the Office of Basic Engineering of the U.S. Department of Energy, under Grant No. DE-FG-02-88-ER13847.

[1] D. J. Evans, E. G. D. Cohen, and G. P. Morriss, Phys. Rev. Lett. **71**, 2401 (1993).
[2] D. J. Evans and D. J. Searles, Phys. Rev. E **50**, 1645 (1994); D. J. Searles and D. J. Evans, J. Chem. Phys. **113**, 3503 (2000).
[3] G. Gallavotti and E. G. D. Cohen, Phys. Rev. Lett. **74**, 2694 (1995); J. Stat. Phys. **80**, 931 (1995).
[4] J. Kurchan, J. Phys. A **31**, 3719 (1998).
[5] J. L. Lebowitz and H. Spohn, J. Stat. Phys. **95**, 333 (1999).
[6] E. G. D. Cohen and G. Gallavotti, J. Stat. Phys. **96**, 1343 (1999).
[7] C. Maes, J. Stat. Phys. **95**, 367 (1999); C. Maes and E. Verbitskiy, Commun. Math. Phys. **233**, 137 (2003).
[8] G. M. Wang, E. M. Sevick, E. Mittag, D. J. Searles, and D. J. Evans, Phys. Rev. Lett. **89**, 050601 (2002).
[9] L. Rey-Bellet and L. E. Thomas, Ann. Henri Poincaré **3**, 483 (2002).
[10] R. van Zon and E. G. D. Cohen, Phys. Rev. E **67**, 046102 (2003).
[11] We call Eq. (1) the conventional SSFT to distinguish it from the extension of the SSFT presented in this Letter.
[12] O. Mazonka and C. Jarzynski, cond-mat/9912121.
[13] E. M. Sevick, D. R. M. Williams, D. J. Searles, and D. J. Evans (private communication).
[14] Compared to Ref. [10], $k_B T = 1$, $k = 1$, and $\alpha = 1$. So the energy unit is $k_B T$, the time unit is the relaxation time α/k, and the length unit is the thermal width $\sqrt{k_B T/k}$.
[15] A force on the particle implies an opposite force on U, which must be overcome by an external force to keep U moving. The work done by this force is W_τ in Eq. (3) [10].
[16] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in Fortran, The Art of Scientific Computing* (Cambridge University Press, Cambridge, England, 1992), 2nd ed.
[17] H. Spohn, *Large Scale Dynamics of Interacting Particles* (Springer-Verlag, Berlin, 1991), p. 241; references ad 3.7 on p. 251.
[18] Singularities affected the distribution of a different quantity as well in J. Farago, J. Stat. Phys. **107**, 781 (2002).
[19] A restriction to the SSFT (i.e., $|p| < 1$) was also found in a model of heat conduction in Ref. [9].
[20] R. van Zon and E. G. D. Cohen (unpublished).