Fluctuations and Irreversibility: An Experimental Demonstration of a Second-Law-Like Theorem Using a Colloidal Particle Held in an Optical Trap

D. M. Carberry,[1] J. C. Reid,[1] G. M. Wang,[1] E. M. Sevick,[1] Debra J. Searles,[2] and Denis J. Evans[1]

[1]*Research School of Chemistry, The Australian National University, Canberra ACT 0200 Australia*
[2]*School of Science, Griffith University, Brisbane QLD 4111 Australia*
(Received 9 September 2003; published 6 April 2004)

The puzzle of how time-reversible microscopic equations of mechanics lead to the time-irreversible macroscopic equations of thermodynamics has been a paradox since the days of Boltzmann. Boltzmann simply sidestepped this enigma by stating "as soon as one looks at bodies of such small dimension that they contain only very few molecules, the validity of this theorem [the second law of thermodynamics and its description of irreversibility] must cease." Today we can state that the transient fluctuation theorem (TFT) of Evans and Searles is a generalized, second-law-like theorem that bridges the microscopic and macroscopic domains and links the time-reversible and irreversible descriptions. We apply this theorem to a colloidal particle in an optical trap. For the first time, we demonstrate the TFT in an experiment and show quantitative agreement with Langevin dynamics.

    PACS numbers: 05.70.Ln, 05.40.–a

The second law of thermodynamics [1,2] states that for systems in the thermodynamic limit, the entropy production must be greater than or equal to zero. Although the underlying equations of motion are time reversible, the second law predicts an irreversible entropy production. This law applies to systems that are of infinite size and persist over long times. Despite this strict limitation, the second law is often treated as being universal in application because the size of most systems can be considered infinite when compared to atomic length and time scales. However, several systems of current scientific interest, such as nanomachines and protein motors, operate at length and time scales where the system cannot be considered infinite. At the nano- and microscales the thermal energy available per degree of freedom can be comparable to the work performed by the system. Classical thermodynamics does not apply to these small systems.

In 1993 the first quantitative description of entropy production in finite systems was given by the fluctuation theorem (FT) of Evans *et al.* [3]. In its most general form, the theorem provides an analytic expression for the probability that a dissipative flux flows in the direction opposite to that required by the second law of thermodynamics. For thermostated dissipative systems [4] the theorem relates the probability of observing a process of duration *t* with entropy production $\Omega_t = A \pm dA$, $P(\Omega_t = A)$, to that of a process with the same magnitude of entropy change, but where the entropy is consumed, rather than produced:

$$\frac{P(\Omega_t = A)}{P(\Omega_t = -A)} = \exp(A). \qquad (1)$$

Since entropy production is extensive and the total entropy production increases with time, the FT predicts that positive entropy production will be overwhelmingly likely as either the system size or the observation time increases. In this way the FT can be viewed as a generalization of the second law since the FT applies to finite systems observed over finite time and trivially recovers the second law in the thermodynamic limit. When applied to the transient response of a system, the theorem is referred to as the transient fluctuation theorem or TFT [2].

The FT has been verified in a number of molecular dynamics (MD) simulations, and it is possible to demonstrate the predictions of the theorem in an experiment involving a colloidal particle and an optical trap. An optical trap is formed when a transparent, micron-sized particle, whose index of refraction is greater than that of the surrounding medium, is located within a focused laser beam. The refracted rays differ in intensity over the volume of the sphere and exert a sub-picoNewton force ($pN = 10^{-12}$ N) on the particle, drawing it towards the region of highest light intensity. The optical trap is harmonic near the focal point; i.e., the potential energy of the system is $\frac{1}{2}k(\mathbf{r} \cdot \mathbf{r})$, where $\mathbf{r}$ is the vector position of the particle relative to the origin-centered trap, and k is the trapping constant which can be tuned by adjusting the laser power. Wang *et al.* [5] recorded the trajectories of a colloidal particle localized in a translating optical trap and evaluated the entropy production for a large number of trajectories. They were unable to demonstrate the TFT, but their results agreed quantitatively with an integrated form of the TFT (ITFT). This ITFT provides a prediction of the ratio of the frequency of entropy-consuming trajectories to that of entropy-generating trajectories:

$$\frac{P(\Omega_t < 0)}{P(\Omega_t > 0)} = \langle \exp(-\Omega_t) \rangle_{\Omega_t > 0}, \qquad (2)$$

where the angular brackets on the right-hand side (rhs) denote an average accumulated over entropy-producing trajectories.

In this Letter, we demonstrate the TFT, Eq. (1), directly by observing the time-dependent relaxation of a colloidal

particle under a step change in the strength of a stationary optical trap. For this experiment, a particle is localized in a stationary trap of strength k_0 over a sufficiently long time so that its position is described by an equilibrium distribution. At $t = 0$ the optical trap strength is increased discontinuously from k_0 to k_1 so that we more tightly confine or "capture" the particle. The particle's position is recorded as it relaxes to its new equilibrium distribution, and we evaluate the function Ω_t over an ensemble of experiments. We demonstrate for the first time that Ω_t, derived consistently from Langevin dynamics, evolves in time according to the TFT, as well as the ITFT, Eqs. (1) and (2). In addition, we show for the first time that the experimental results and TFT predictions are in quantitative agreement with Langevin dynamics.

The TFT is a simple but general relation that is derived exactly from the laws of mechanics. For thermostated dissipative systems such as that of Wang *et al.* [5], the argument of the theorem, Ω_t, is identified with the entropy production and its average, $\langle\Omega_t\rangle$, is identical to the irreversible entropy production [6]. The relationship between the fluctuation theorem and entropy has been examined by several researchers [7–9]. For more general systems, as in the experiment described here, Ω_t is identified as the dissipation function, having properties that are strikingly similar to the entropy production. Motion of a system which is consistent with the second law is associated with a positive dissipation function, analogous to entropy generation. A negative dissipation function is indicative of motion contrary to the second law and is similar to entropy consumption. And finally, like entropy production, the average of the dissipation function is a positive quantity. Regardless of whether Ω_t is a generalized dissipation function or the entropy production, Ω_t quantifies the irreversibility of the system. The TFT is a relation describing how Ω_t evolves in time, and in this way the theorem describes the transition from time-reversible, microscopic equations of motion to irreversible macroscopic behavior. The puzzle of how the time-reversible microscopic equations of quantum and classical mechanics lead to the time-irreversible macroscopic equations of thermodynamics and hydrodynamics has been a paradox since the days of Boltzmann. Boltzmann simply sidestepped this enigma by stating "as soon as one looks at bodies of such small dimension that they contain only very few molecules, the validity of this theorem [the second law and its description of irreversibility] must cease" [1]. Today we can state that the TFT is a generalized, second-law-like theorem that bridges the microscopic and macroscopic domains and links the time-reversible and irreversible descriptions.

A system's trajectory can be expressed in terms of the coordinates $\mathbf{q}$, and momenta, $\mathbf{p}$, of all constituent molecules, including solvent, as well as the thermostat employed to describe the system's heat exchange with its surroundings. Figure 1 illustrates these sets of trajectories in the $(\mathbf{q}, \mathbf{p})$ space. A system's trajectory $\mathbf{\Gamma}(t)$ from an initial state $(\mathbf{q}_0, \mathbf{p}_0)$ at $t = 0$ to a state $(\mathbf{q}_t, \mathbf{p}_t)$ some time t later is a solution of Newton's equations of motion. As Loschmidt [10] pointed out in 1876, since these equations are time reversible, for every trajectory $\mathbf{\Gamma}(t)$ that satisfies the equations of motion, there is a time-reversed trajectory or antitrajectory which is also a solution to the equations. This antitrajectory, represented by $\mathbf{\Gamma}^*(t)$, evolves from an initial state $(\mathbf{q}_t, -\mathbf{p}_t)$ to a final state $(\mathbf{q}_0, -\mathbf{p}_0)$. Thus, $\mathbf{\Gamma}(t)$ and $\mathbf{\Gamma}^*(t)$ represent a pair of conjugate trajectories that are related by time reversal. Consider a set of trajectories that start inside an infinitesimal volume $\delta V \equiv \delta\mathbf{q}_0\delta\mathbf{p}_0$ about an initial state $(\mathbf{q}_0, \mathbf{p}_0)$, and the set of conjugate trajectories that start inside the volume element $\delta V^* \equiv \delta\mathbf{q}_t\delta\mathbf{p}_t$ about $(\mathbf{q}_t, -\mathbf{p}_t)$. Every trajectory initiated within δV has a conjugate trajectory initiated in δV^*. The ratio of the probability of observing trajectories initiated within δV, $P(\delta V)$, to those within δV^* is the dissipation function,

$$\Omega_t(\mathbf{q}_0, \mathbf{p}_0) \equiv \ln\left[\frac{P(\delta V)}{P(\delta V^*)}\right]. \qquad (3)$$

If the probability of observing a trajectory in δV is equal to the probability of observing one in δV^*, then the system is thermodynamically reversible and $\Omega_t = 0$. A recent review [11] rigorously derives the probability densities of trajectories and the resulting dissipation functions from reversible equations of motion under various statistical ensembles.
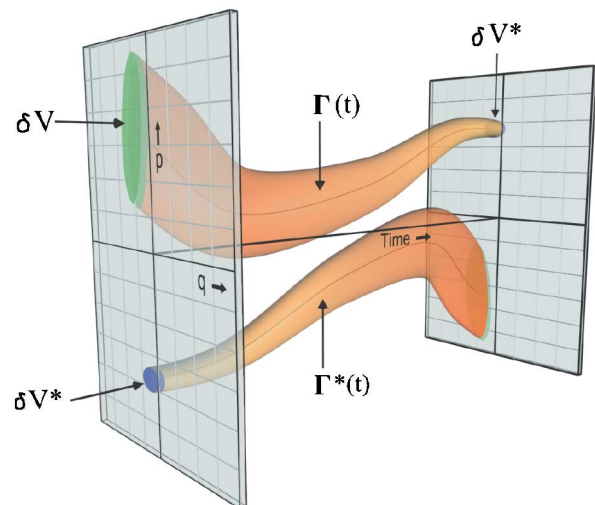

FIG. 1 (color online). An illustration of a set of neighboring trajectories initiated in a volume element δV (top tube) and the corresponding set of antitrajectories initiated in δV^* (lower tube) in coordinate momentum $(\mathbf{q}, \mathbf{p})$ and time, t, space. The ratio of the probability of observing trajectories initiated within δV to those in δV^* is a measure of the system's irreversibility. Every trajectory in the $(\mathbf{q}, \mathbf{p})$ space corresponds to a solution of Newton's time-reversible equations of motion. Thus, for every trajectory $\mathbf{\Gamma}(t)$ that starts at $(\mathbf{q}_0, \mathbf{p}_0)$ and ends at $(\mathbf{q}_t, \mathbf{p}_t)$ some time t later, there exists its time-reversed or antitrajectory $\mathbf{\Gamma}^*(t)$. This antitrajectory starts at $(\mathbf{q}_t, -\mathbf{p}_t)$ and ends at $(\mathbf{q}_0, -\mathbf{p}_0)$ at time t.

For colloidal systems, Newtonian MD simulations do not accurately portray the experiment due to the large disparity in the time scales of motion for solvent molecules and for colloidal particles. However, the motion of a single colloidal particle is well approximated by the inertialess Langevin equation [12]. This is a first-order stochastic differential equation describing the particle's coordinates over time, where the many degrees of freedom associated with the solvent molecules are recast into the macroscopic material properties of viscosity and temperature and the buffeting action of the solvent molecules on the particle is represented by a fluctuating random force. A state of the system is no longer described by the set of coordinates and momenta of all constituent molecules but is reduced to only the coordinates of the colloidal particle, $\mathbf{r}$. Unlike Newtonian dynamics, such stochastically described trajectories are not time reversible. Moreover, there exist a large subset of stochastic trajectories that evolve from $\mathbf{r}_0$ to $\mathbf{r}_t$ in time t, represented by $\{\mathbf{r}_0, \mathbf{r}_t\}$ as well as a conjugate set of "anti-" trajectories that evolve from $\mathbf{r}_t$ to $\mathbf{r}_0$, represented by $\{\mathbf{r}_t, \mathbf{r}_0\}$. For each forward trajectory in the set $\{\mathbf{r}_0, \mathbf{r}_t\}$, there exists a corresponding antitrajectory in $\{\mathbf{r}_t, \mathbf{r}_0\}$. The probabilities of observing these sets of stochastic trajectories, $P(\{\mathbf{r}_0, \mathbf{r}_t\})$ and $P(\{\mathbf{r}_t, \mathbf{r}_0\})$, can be constructed using Langevin dynamics. The measure of a system's irreversibility applies to trajectories, irrespective of whether these trajectories are described using Newtonian mechanics or Langevin stochastics [13–18]. Thus, by analogy to Eq. (3), the dissipation function for the stochastic trajectories is

$$\Omega_t(\mathbf{r}_0, \mathbf{r}_t) = \ln\left[\frac{P(\{\mathbf{r}_0, \mathbf{r}_t\})}{P(\{\mathbf{r}_t, \mathbf{r}_0\})}\right]. \quad (4)$$

For the case of a single colloidal particle in a two-dimensional harmonic potential, Langevin dynamics provides simple analytic formulas for the time-dependent probability densities [19]. The Green's function provides the probability of observing a particle at $\mathbf{r}'$ in a trap of strength k, given its position $\mathbf{r}$ at time t earlier, and is represented by $G(\mathbf{r}', k; \mathbf{r}, t)$. In the limit of long time, $t \rightarrow \infty$, the probability density of observing the particle at $\mathbf{r}'$ is independent of time and the Green's function reduces to the equilibrium Boltzmann's distribution, $P_B(\mathbf{r}', k)$. Thus, the probability density of forward trajectories, $P(\{\mathbf{r}_0, \mathbf{r}_t\})$, is $P_B(\mathbf{r}_0, k_0)G(\mathbf{r}_t, k_1; \mathbf{r}_0, t)$. Likewise, the probability density of the corresponding antitrajectories is $P(\{\mathbf{r}_t, \mathbf{r}_0\}) \equiv P_B(\mathbf{r}_t, k_0)G(\mathbf{r}_0, k_1; \mathbf{r}_t, t)$, and the dissipation function, Eq. (4), simplifies to

$$\Omega_t = \frac{1}{2k_B T}(k_0 - k_1)(\mathbf{r}_t^2 - \mathbf{r}_0^2). \quad (5)$$

This expression is identical to that derived using Newton's (deterministic) equations of motion [20].

The experimental setup consists of a Nikon DIAPHOT 300 inverted microscope equipped with a $100\times$ (N.A. = 1.3) oil-immersion objective lens and a 1 W infrared laser ($\lambda = 980$ nm, Cell Robotics, USA) to create the optical trap. We detect the position of a

trapped particle by projecting its image onto a quadrant photodiode with resolution 15 nm (S4349 Hamamatsu, Japan). The optical trap strength is controlled by adjusting the laser intensity. The change in the laser intensity, recorded by a photodiode (Honeywell, SDP86001) located above the sample cell, occurs over 2 to 3 ms. Labview (National Instruments, USA) and a PCI data acquisition card (PCI-6014 National Instruments, USA) are used for 1 kHz automatic data collection.

An 8.0 ml transparent sample cell is filled with an aqueous solvent containing 4.0 ml of 10 mM Tris-HCl with 1 mM EDTA maintained at a pH of 7.5, and 1.0 ml of glycerol, added to maximize the viscosity and time scale of particle motion without disrupting the neutral buoyancy of the colloidal particles. A single latex particle (6.3 μm diameter, Interfacial Dynamics Co., USA) is isolated in the cell and is used to calibrate the quadrant photodiode detector, to determine trap strength, and to collect an ensemble of trajectories. To determine the value of the trapping constant, k, and to verify the harmonicity, we sample the particle positions under laser power at 200 Hz for 120 s and apply the equipartition theorem. $k^{x,y} = k_B T/\langle r_{x,y}^2 \rangle$. The strength of the optical trap is then cycled discontinuously and evenly between the initial trap strength, $k_0 = 1.22$ pN/μm, and a final trap strength, $k_1^x = 2.90$ and $k_1^y = 2.70$ pN/μm with a period of 20 s. This allows us to record hundreds of trajectories in one experiment. A "trajectory" corresponds to the particle's position over 10 s in the weak trap ($-10 < t < 0$ s) and a further 10 s in the strong trap ($0 \geq t > 10$ s). The first half of the cycle provides ample time for the particle to equilibrate in the weak trap. The second half of the cycle
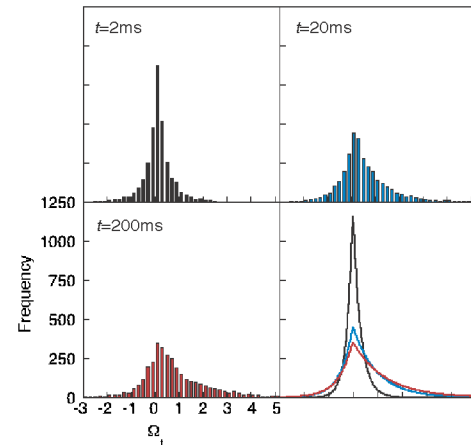

FIG. 2 (color online). Three histograms showing the distribution of the dissipation function, Ω_t, evaluated over 3300 experimental trajectories at times $t = 0.002$ s, $t = 0.02$ s, and $t = 0.2$ s after the trap strength is increased from $k_0 = 1.22$ pN/μm to $k_1 \equiv (k_1^x, k_1^y) = (2.90, 2.70$ pN/μm). The width of each bin, $\Delta = 0.2$, corresponds to $1/3$ of the standard deviation of the Ω_t distribution sampled at $t = 1$ ms. The curves shown in the bottom right show the Langevin-predicted distributions, $P(\Omega_t)$, for similar t/τ and trap strengths. The range of all plots in the figure are identical.
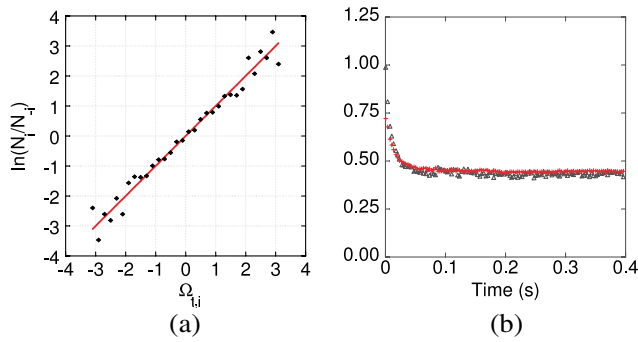
FIG. 3 (color online). (a) Logarithm of the ratio of number of trajectories in histogram bins, $\ln(N_i/N_{-i})$, versus the value of the dissipation function associated with the ith bin, $\Omega_{t,i}$, evaluated for trajectories of time duration $t = 26$ ms. The line of best fit has a slope of 0.98 ± 0.04 which agrees with the TFT predicted slope of 1 (shown as a line). This is the first direct demonstration of the TFT. (b) The lhs and rhs of the ITFT, $N_{i<0}/N_{i>0}$ ($\triangle$) and $\langle \exp(-\Omega_t) \rangle_{\Omega_t>0}$ ($+$) versus time, t.

provides the data necessary for analysis of the TFT. The relaxation time of the particle in the strong trap, $\tau \equiv \xi/k_1$, is 37.5 ms, where the friction coefficient was determined by Stokes drag as $\xi = 1.05 \times 10^{-7}$ Ns/m.

Figure 2 is a histogram of the values of the dissipation function, Ω_t, evaluated over 3300 experimental trajectories where the trap strength is increased from k_0 to k_1 at $t = 0$. The distribution of Ω_t is strongly peaked at $\Omega_t \approx 0$ and is clearly non-Gaussian. At early times, the distribution of Ω_t is roughly symmetric, showing an equal number of trajectories with positive and negative Ω_t. As time increases, the distribution of Ω_t becomes skewed towards positive values, showing that the system is increasingly characterized as irreversible. The average value of the dissipation function is always greater than zero and at long times reaches a plateau value.

This histogram data can be recast to *directly* demonstrate the TFT, Eq. (1). Let N_i be the number of experimental trajectories with dissipation function evaluated to be between $\Omega_{t,i} - \Delta/2$ and $\Omega_{t,i} + \Delta/2$, where Δ is the size of the histogram bin in dimensionless units of Ω_t and $\Omega_{t,i} = i\Delta$. The experimental analog of the left-hand side (lhs) of the TFT then becomes $\ln(N_i/N_{-i})$, where the argument of the logarithm is the ratio of the number of trajectories in a pair of histogram bins. Figure 3(a) is the logarithm of the ratio of populations of pairs of bins, $\ln(N_i/N_{-i})$, versus the value the dissipation function for the ith bin, $\Omega_{t,i}$, evaluated for trajectories of duration $t = 26$ ms. Also shown in the figure is a line of slope unity, representing the TFT prediction. The previous experimental work [5] only confirmed the *integrated* form of the TFT.

The ITFT, Eq. (2), is also demonstrated from the experiments and the result is shown in Fig. 3(b). Over the complete 10 s duration of the trajectories, the experimental values of the lhs and rhs of the ITFT are nearly equivalent, thereby demonstrating the ITFT [20]. Both show that the relaxation of $\langle \Omega_t \rangle$ is roughly on the time scale of the motion of the colloidal particle; i.e., $\tau = 37.5$ ms.

The TFT quantitatively describes how irreversible macroscopic behavior evolves from time-reversible microscopic dynamics as either the system size or the observation time increases. This experiment confirms, for the first time, the predictions of this theorem. Furthermore, these experiments demonstrate that the effects predicted occur over colloidal length and time scales and, consequently, show that the TFT is relevant to nanotechnological applications.

[1] E. Broda, *Ludwig Boltzmann: Man—Physicist—Philosopher* (Ox Bow Press, Woodbridge, 1983).
[2] D. J. Evans and D. J. Searles, Phys. Rev. E **50**, 1645 (1994).
[3] D. J. Evans, E. G. D. Cohen, and G. P. Morris, Phys. Rev. Lett. **71**, 2401 (1993).
[4] A thermostated dissipative system is a system in which work is turned into heat, in the presence of a heat bath that maintains the system at a constant temperature.
[5] G. M. Wang, E. M. Sevick, E. Mittag, D. J. Searles, and D. J. Evans, Phys. Rev. Lett. **89**, 050601 (2002).
[6] S. R. de Groot and P. Mazur, *Non-equilibrium Thermodynamics* (Dover Publications, Mineola, NY, 1984).
[7] G. Crooks, Phys. Rev. E **60**, 2721 (1999).
[8] C. Maes, F. Redig, and A. V. Moffaert, J. Math. Phys. (N.Y.) **41**, 1528 (2000).
[9] C. Maes and F. Redig, J. Stat. Phys. **101**, 3 (2000).
[10] J. Loschmidt, J. Sitzungsber. der kais. Akad. d. W. Math. Naturw. II **73**, 128 (1876).
[11] D. J. Evans and D. J. Searles, Adv. Phys. **51**, 1529 (2002).
[12] In this system the inertia time scale is less than 1 ms. This is smaller than our sampling rate and the inertialess Langevin equation can therefore be used.
[13] J. L. Lebowitz and H. Spohn, J. Stat. Phys. **95**, 333 (1999).
[14] G. Crooks, Phys. Rev. E **61**, 2361 (2000).
[15] C. Maes, J. Stat. Phys. **95**, 367 (1999).
[16] J. Kurchan, J. Phys. A **31**, 3719 (1998).
[17] R. van Zon and E. G. D. Cohen, Phys. Rev. E **67**, 046102 (2003).
[18] S. Tasaki, I. Terasaki, and T. Monnai, cond-mat/0208154.
[19] M. Doi and S. F. Edwards, *The Theory of Polymer Dynamics* (Oxford University Press, Oxford, 1988).
[20] J. Reid, D. M. Carberry, G. M. Wang, E. M. Sevick, D. J. Searles, and D. J. Evans (to be published).