

Fluctuation Theorems

E.M. Sevick,¹ R. Prabhakar,¹
Stephen R. Williams,¹ and Debra J. Searles²

¹Research School of Chemistry, Australian National University, Canberra, Australian Capital Territory 0200 Australia; email: sevick@rsc.anu.edu.au, prabhakar.ranganathan@eng.monash.edu.au, swilliams@rsc.anu.edu.au

²Nanoscale Science and Technology Centre, School of Biomolecular and Physical Sciences, Griffith University, Brisbane, QLD 4111 Australia; email: D.Bernhardt@griffith.edu.au

Annu. Rev. Phys. Chem. 2008. 59:603–33

First published online as a Review in Advance on December 11, 2007

The *Annual Review of Physical Chemistry* is online at <http://physchem.annualreviews.org>

This article's doi:
10.1146/annurev.physchem.58.032806.104555

Copyright © 2008 by Annual Reviews.
All rights reserved

0066-426X/08/0505-0603\$20.00

Key Words

nonequilibrium statistical mechanics, far-from-equilibrium processes, second law of thermodynamics, reversibility, free energy change

Abstract

Fluctuation theorems, developed over the past 15 years, have resulted in fundamental breakthroughs in our understanding of how irreversibility emerges from reversible dynamics and have provided new statistical mechanical relationships for free-energy changes. They describe the statistical fluctuations in time-averaged properties of many-particle systems such as fluids driven to nonequilibrium states and provide some of the few analytical expressions that describe nonequilibrium states. Quantitative predictions on fluctuations in small systems that are monitored over short periods can also be made, and therefore the fluctuation theorems allow thermodynamic concepts to be extended to apply to finite systems. For this reason, we anticipate an important role for fluctuation theorems in the design of nanotechnological devices and in the understanding of biological processes. This review discusses these theorems, their physical significance, and results for experimental and model systems.

Macroscopic irreversibility:

macroscopic processes have a direction associated with them; the opposing direction is prohibited by the second law

1. INTRODUCTION

Our understanding of thermodynamics, which is the study of the flow of heat and the transformation of work into heat, is largely confined to equilibrium states. Linear irreversible thermodynamics is an extension of the nineteenth-century concepts of equilibrium thermodynamics to systems that are close to equilibrium (1, 2). These traditional concepts are limited in application to large systems or averages over an ensemble of states, referred to as the thermodynamic limit. In contrast, modern inventors and engineers endeavor to scale down machines and devices to nanometer sizes for a wide range of technological purposes. However, there is a fundamental limit to miniaturization because small engines are not simply rescaled versions of their larger counterparts. If the work performed during the duty cycle of any machine is comparable to its thermal energy per degree of freedom, then one can expect that the machine will operate in reverse over short timescales. That is, heat energy from the surroundings is converted into useful work, allowing the engine to run backwards. For larger engines, we would describe this as a violation of the second law of thermodynamics. Until recently, this received little attention in the nanotechnology literature, as there was no second law–like relation for finite-sized systems outside the thermodynamic limit.

In the past 15 years, researchers have proposed several fluctuation theorems that revolutionized our understanding and use of thermodynamics. First, these new theorems lift the requirement of the thermodynamic limit. This allows thermodynamic concepts to be applied to finite, even small systems. Second, these new theorems can be applied to systems that are arbitrarily far from equilibrium. Third, for the first time, these theorems explain how macroscopic irreversibility appears naturally in systems that obey time-reversible microscopic dynamics. One of these fluctuation theorems (FTs), the Evans-Searles FT (3–5), results in a generalization of the second law of thermodynamics so that it applies to small systems, including those that evolve far from equilibrium. Another, the Crooks FT (6, 7), provides a method of predicting equilibrium free-energy differences from nonequilibrium paths that connect two equilibrium states. Both FTs are at odds with the traditional understanding of nineteenth-century thermodynamics. Nevertheless, these theorems are essential for the application of thermodynamic concepts to nanotechnology systems that are currently of such interest to biologists, physical scientists, and engineers.

1.1. The Evans-Searles Fluctuation Theorem

In many areas of physical chemistry, researchers strive to understand new systems through deterministic equations of motion. They seek to quantify microscopic forces and understand how a system responds to external perturbations, using techniques such as molecular dynamics simulation. At the heart of this endeavor is the notion that if the equations of motion or trajectories of the system are known, then any question about that system may be answered. However, such deterministic equations (such as Newton's equations) are time reversible, so for every trajectory, there exists a conjugate, time-reversed trajectory or antitrajectory that is also a solution to the

equations. One can use the relative probabilities of observing bundles of conjugate trajectories to quantify the macroscopic irreversibility of the system: If the probability of observing all trajectories and their respective antitrajectories is equal, the system is reversible; conversely, if the probability of observing antitrajectories is vanishingly small, the system is irreversible. The second law of thermodynamics stipulates that a system evolves irreversibly in one time-forward direction (i.e., the probability of all antitrajectories is zero). However, the second law strictly applies to large systems or over long timescales and does not describe the reversibility of the small systems that are of current scientific interest, such as protein motors and nanomachines. In 1876 Loschmidt first noted this long-standing question of how irreversible macroscopic equations, as summarized by the second law, can be derived from reversible microscopic equations of motion, and it is referred to today as Loschmidt's paradox (8, 9). Resolution of this irreversibility paradox has defied our efforts for more than 100 years. Boltzmann and his successors simply sidestepped this paradox, with Boltzmann 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] must cease" (10).

The Evans-Searles FT (3–5) describes how a system's irreversibility develops in time from a completely time-reversible system at short observation times, to an irreversible one at long times. It also shows how irreversibility emerges as the system size increases. That is, it bridges the microscopic and macroscopic descriptions, relating a system's time-reversible equations of motion to the second law, and provides a quantitative resolution to the long-standing irreversibility paradox. Specifically, the FT relates the relative probabilities, p , of observing trajectories of duration t characterized by the dissipation function, Ω_t , taking on arbitrary values \mathcal{A} and $-\mathcal{A}$, respectively:

$$\frac{p(\Omega_t = \mathcal{A})}{p(\Omega_t = -\mathcal{A})} = \exp(\mathcal{A}). \quad (1)$$

It is an expression that describes the asymmetry in the distribution of Ω_t over a particular ensemble of trajectories. The dissipation function, Ω_t , is, in general, a dimensionless dissipated energy, accumulated along the system's trajectory; expressions for Ω_t differ from system to system. However, any trajectory of the system characterized by a particular value $\Omega_t = \mathcal{A}$ has, under time-reversible mechanics, a conjugate or time-reversed antitrajectory with $\Omega_t = -\mathcal{A}$. In this way, the left-hand side of Equation 1 has also been interpreted as a ratio of the probabilities of observing trajectories to their respective antitrajectories. The dissipation function, Ω_t , is an extensive property (i.e., its magnitude scales with system size, and it also scales with the observation time, t). Thus, Equation 1 also shows that as the system size gets larger or the observation time gets longer, antitrajectories become rare, and it becomes overwhelmingly likely that the system appears time irreversible, in accordance with the second law. That is, the evolution of a large macroscopic system proceeds preferentially in one direction. Equation 1 also shows that the ensemble average of the dissipation function is positive for all t , for all nonequilibrium systems, and for any system size; i.e., $\langle \Omega_t \rangle \geq 0$ (5), which is referred to as the second law inequality.

1.2. The Crooks Fluctuation Theorem and Jarzynski's Equality

From classical thermodynamics, the work required to drive a system from one equilibrium state to another equilibrium state is equivalent to the change of free energy, ΔF , between the states, only in the special case in which the path is traversed quasistatically. That is, the path between the two states must be traversed so slowly that intermediate states, as well as the initial and final states of the system, are all in thermodynamic equilibrium. The Crooks FT (6, 7) states something quite remarkable. In the case of paths that are traversed at an arbitrary rate, ranging from quasistatic to far from equilibrium, the distribution of trajectories, characterized by the work done over the duration of the trajectory, follows

$$\frac{p_f(W = \mathcal{A})}{p_r(W = -\mathcal{A})} = \exp[\beta(\mathcal{A} - \Delta F)], \quad (2)$$

where $\beta = 1/(k_B T)$; k_B is Boltzmann's constant; and T is the initial temperature of the system, or equivalently the temperature of the surroundings with which the system is initially at equilibrium. This expression is similar to Evans-Searles FT in that it relates distributions of trajectories, characterized by an energy, specifically to the work, W . Whereas the Evans-Searles FT (Equation 1) describes the asymmetry in the distribution of trajectories starting from the same initial distribution, the Crooks FT (Equation 2) relates trajectories initiated from two different equilibrium states, A and B . That is, it considers (a) a distribution, p_f , of forward trajectories, $A \rightarrow B$, where the free-energy change between equilibrium states A and B is $\Delta F = F_B - F_A$, and (b) the distribution, p_r , of reverse trajectories, $B \rightarrow A$, where the respective equilibrium free-energy change is $-\Delta F$. Similar to the Evans-Searles FT, the Crooks FT also quantifies how irreversibility evolves out of reversible equations of motion. A perfectly reversible (quasistatic) system is one in which the work required to traverse $B \rightarrow A$ is equal but opposite in sign to the work required in the time-reversed trajectory, $A \rightarrow B$. Thus the right-hand side of Equation 2 is unity for these reversible paths and $W = \Delta F$, in agreement with classical thermodynamics.

Taking the ensemble average of $\exp(-\beta W)$ and using the Crooks FT give

$$\exp(-\beta \Delta F) = \langle \exp(-\beta W) \rangle_f. \quad (3)$$

Here the notation $\langle \dots \rangle_f$ denotes an ensemble average using the distribution function of state A , and the work is measured over forward trajectories, $A \rightarrow B$. Jarzynski (11, 12) first posed this expression in 1997, before the derivation of the Crooks FT, and it is known as Jarzynski's equality. It states that one can determine the free energy by measuring the work done along dynamical paths that connect the two states. These forward paths may be traversed at arbitrary rates, so that the intervening states might not be in thermodynamic equilibrium. This provides a completely new way of treating thermodynamics. If, instead of averaging the work, you average the exponential of the work, then you can calculate the equilibrium free-energy difference from information obtained along nonequilibrium paths. On the practical side, Equation 3 suggests that measuring work on small microscopic processes could yield thermodynamic quantities ΔF that are traditionally inferred by calorimetric measurements. The importance here is that to understand molecular-scale processes, one must probe

them using molecular time/length scales: The FTs and Jarzynski's equality are necessary for interpreting measurements taken over these small time/length scales.

2. BACKGROUND CONCEPTS

The degrees of freedom of a system of particles are represented by the vectors of time-dependent coordinates, \mathbf{q} , and momenta, \mathbf{p} , or a point in the system's phase space, denoted $\Gamma \equiv (\mathbf{q}, \mathbf{p})$. Let us consider a closed system in thermal equilibrium with a reservoir. From equilibrium statistical mechanics, the equilibrium probability distribution of the system is given by the canonical distribution function,

$$f_{eq}(\Gamma) = \frac{e^{-\beta\mathcal{H}(\Gamma)}}{Z}, \quad (4)$$

where $Z = \int e^{-\beta\mathcal{H}} d\Gamma$ is the equilibrium partition function, and \mathcal{H} is the internal energy,¹ which is the sum of the kinetic and potential energies, $K(\mathbf{p})$ and $\phi(\mathbf{q})$, of the system.

A closed adiabatic system can exchange energy with its environment in the form of work: Work is the form of energy exchange that is directly controllable by the environment. For example, it might be desirable to change the mean internal energy, $U = \langle \mathcal{H} \rangle$, of the system, which can be achieved by externally controlling some parameter λ in the potential energy function of the system, ϕ . Examples of such λ parameters include the switch on an externally applied electric field in a crystalline salt, the trapping constant in an optical trap holding a colloidal particle, or a mathematical agent that changes the size of Lennard-Jones spheres in a computer simulation. We emphasize this mode of external control by formally making λ time dependent and by writing the internal energy as

$$\mathcal{H}(\Gamma, s) = K(\mathbf{p}) + \phi(\mathbf{q}, \lambda(s)). \quad (5)$$

When an external agent does work on a system without changing its underlying equilibrium state, we refer to that agent as a purely dissipative field, denoted \mathbf{F}_e . This dissipative field does not figure in the underlying equilibrium distribution or partition function, but drives the system away from equilibrium according to the equations of motion, in which \mathbf{F}_e is explicit. Although it may be possible to represent an external agent using either \mathbf{F}_e or λ , we choose the convention that if $\dot{\lambda} \neq 0$ and $\mathbf{F}_e = 0$, the system always relaxes to a (nondissipative) equilibrium state, and if $\mathbf{F}_e \neq 0$, the system never relaxes to a nondissipative state. This distinction may depend on the state of system (e.g., fluid or solid). Examples of such dissipative fields include a fluid under a shear flow, a colloidal particle being dragged in a fluid, and an electric field acting on a molten salt. Under adiabatic conditions (i.e., the rate of heat exchange with the reservoir is $\dot{Q} = 0$), the combined action of both kinds of external agents

¹More exactly, \mathcal{H} is the phase variable corresponding to the internal energy. For simplicity, we use the term "internal energy" to refer to \mathcal{H} and refer to the thermodynamic internal energy $U = \langle \mathcal{H} \rangle$ as the "mean internal energy."

(i.e., a time-dependent potential represented by a λ parameter and a dissipative field \mathbf{F}_e) results in the equations of motion,

$$\begin{aligned}\dot{\mathbf{q}} &= \frac{\partial \mathcal{H}(\Gamma, s)}{\partial \mathbf{p}} + \mathbf{C}(\Gamma) \cdot \mathbf{F}_e(s), \\ \dot{\mathbf{p}} &= -\frac{\partial \mathcal{H}(\Gamma, s)}{\partial \mathbf{q}} + \mathbf{D}(\Gamma) \cdot \mathbf{F}_e(s),\end{aligned}\quad (6)$$

with $\mathcal{H}(\Gamma, s)$ given by Equation 5.² For an externally driven adiabatic system, the rate of increase of \mathcal{H} must be identically equal to the rate of work \dot{W} done on the system by the environment. Thus,

$$\dot{W}(\Gamma, s) = \dot{\mathcal{H}}^{ad}(\Gamma, s) = \dot{\lambda} \frac{\partial \mathcal{H}(\Gamma, s)}{\partial \lambda} + \dot{\mathbf{q}} \cdot \frac{\partial \mathcal{H}(\Gamma, s)}{\partial \mathbf{q}} + \dot{\mathbf{p}} \cdot \frac{\partial \mathcal{H}(\Gamma, s)}{\partial \mathbf{p}}, \quad (7)$$

where the superscript *ad* emphasizes adiabatic conditions. Using Equation 6, we obtain

$$\dot{W}(\Gamma, s) = \dot{\lambda} \frac{\partial \phi(\mathbf{q}, \lambda(s))}{\partial \lambda} - V \mathbf{J}(\Gamma) \cdot \mathbf{F}_e(s), \quad (8)$$

where V is the volume of the system; and $\mathbf{J}(\Gamma)$, the dissipative flux due to the field $\mathbf{F}_e(s)$, is formally defined through the equation

$$V \mathbf{J} \cdot \mathbf{F}_e \equiv - \left(\frac{\partial \mathcal{H}}{\partial \mathbf{q}} \cdot \mathbf{C} \cdot \mathbf{F}_e + \frac{\partial \mathcal{H}}{\partial \mathbf{p}} \cdot \mathbf{D} \cdot \mathbf{F}_e \right). \quad (9)$$

In the case of small systems such as protein motors or artificial nanomachines, it is quite difficult to thermally isolate the system to achieve perfectly adiabatic conditions. Moreover, in most applications of interest, such systems typically function in an environment of constant temperature. Molecular dynamics simulations of small systems have employed thermostats that involve appending Equation 6 with a mathematical constraint to fix the temperature T . For example, with a Gaussian isokinetic thermostat, Equation 6 takes the form (13)

$$\begin{aligned}\dot{\mathbf{q}} &= \frac{\partial \mathcal{H}(\Gamma, s)}{\partial \mathbf{p}} + \mathbf{C}(\Gamma) \cdot \mathbf{F}_e(s), \\ \dot{\mathbf{p}} &= -\frac{\partial \mathcal{H}(\Gamma, s)}{\partial \mathbf{q}} + \mathbf{D}(\Gamma) \cdot \mathbf{F}_e(s) - \alpha(\Gamma) \mathbf{S} \cdot \mathbf{p}.\end{aligned}\quad (10)$$

Here, α is a thermostat multiplier,³ and \mathbf{S} is a diagonal matrix (with ones and zeros on the diagonal) that describes which components of the system are thermostatted.⁴

²We must point out here that the notation does not imply that \mathbf{F}_e is a force, nor does it have to always be a vector. For example, it could be a second-order tensor, such as the velocity gradient tensor in a fluid. The coupling tensors \mathbf{C} and \mathbf{D} are functions of Γ , have no explicit time dependence, and are formally one tensorial order higher than \mathbf{F}_e .

³In the case of a Nosé-Hoover thermostat (14), α becomes an additional independent variable and is a function of time governed by an additional equation rather than a direct function of Γ (13).

⁴When the external field is removed, $\mathbf{F}_e = 0$, and the potential's parameter is held fixed, $\dot{\lambda} = 0$, the equilibrium Equation 10 relaxes to is given by $f_{eq}(\Gamma) = \frac{\exp(-\beta \mathcal{H}(\Gamma, \lambda))}{Z} \delta(\mathbf{p} \cdot \mathbf{S} \cdot \mathbf{p} - 2mK)$, where the Dirac delta function accounts for the kinetic energy of the thermostatted particles being held fixed to the value K , and m is the particle mass. The partition function is given by $Z = \int d\Gamma \exp(-\beta \mathcal{H}(\Gamma, \lambda)) \delta(\mathbf{p} \cdot \mathbf{S} \cdot \mathbf{p} - 2mK)$.

Several other mathematical constraints can be constructed, which are artificial but irrelevant to the theory (for a brief discussion, see the **Supplemental Appendix**; follow the Supplemental Material link from the Annual Reviews home page at <http://www.annualreviews.org>). An increase in the internal energy must equal the sum of work done on the system by the environment and the heat added to the system by the thermostat. However, the rate of work must still be given by Equation 8 regardless of whether the system is thermostatted. From the first law of thermodynamics, the expression for the rate of heat exchange is $\dot{Q} = \dot{\mathcal{H}} - \dot{W}$, with \dot{W} given by Equation 8 and \dot{Q} depending on the thermostat employed in the equations of motion.⁵

Let us consider a particular trajectory initiated at time $s = 0$ at $\Gamma_0 \equiv (\mathbf{q}_0, \mathbf{p}_0)$ that terminates after time t at $\Gamma_t \equiv (\mathbf{q}_t, \mathbf{p}_t)$. We let $d\Gamma_s$ represent an infinitesimal volume of phase space at time s about the point Γ_s . As the dynamics is deterministic, the trajectory is completely determined by the phase-space coordinates at any time s along the trajectory and the duration or observation time, t , of the trajectory. Consequently, for every initial state within a volume element $d\Gamma_0$, there exists a unique destination point within volume element $d\Gamma_t$. As the trajectories in an infinitesimal bundle around the initial state, $d\Gamma_0$, form the later bundle $d\Gamma_t$, the ratio of the volumes of the infinitesimal volume elements varies as

$$\left| \frac{d\Gamma_t}{d\Gamma_0} \right| = \frac{\delta V(\Gamma_t)}{\delta V(\Gamma_0)} = \exp \left(\int_0^t \Lambda(\Gamma_s) ds \right), \quad (11)$$

where the right-hand side is a path integral. Here $d\Gamma_t/d\Gamma_0$ is the Jacobian of the transformation of the initial Γ_0 to the final Γ_t , and $\Lambda(\Gamma_s)$ is the phase-space compression factor:

$$\Lambda(\Gamma_s) \equiv \frac{\partial}{\partial \Gamma_s} \cdot \dot{\Gamma}_s = \left(\frac{\partial}{\partial \mathbf{q}} \cdot \dot{\mathbf{q}} + \frac{\partial}{\partial \mathbf{p}} \cdot \dot{\mathbf{p}} \right)_s. \quad (12)$$

Equations 11 and 12 describe how the volume of a small region of phase space evolves in time. For adiabatic systems, there is no change of phase-space volume along a trajectory or $\Lambda^{ad} = 0$. From the equations for an adiabatic system (Equation 6), and the definition of Λ , the field \mathbf{F}_e and the coupling tensors \mathbf{C} and \mathbf{D} must satisfy

$$\frac{\partial}{\partial \mathbf{q}} \cdot \mathbf{C} \cdot \mathbf{F}_e + \frac{\partial}{\partial \mathbf{p}} \cdot \mathbf{D} \cdot \mathbf{F}_e = 0, \quad (13)$$

irrespective of whether the system is thermostatted. This condition is known as the adiabatic incompressibility of phase space (AI Γ). However, for thermostatted systems in a driven steady state, a contraction of phase space occurs continually, as the initial phase volume shrinks to a fractal attractor of lower dimension than the ostensible phase space. For appropriately selected thermostats,⁶ the phase-space contraction factor is directly proportional to the rate of heat exchange with the thermostat (15, 16),

$$\dot{Q}(\Gamma) = k_B T \Lambda(\Gamma). \quad (14)$$

⁵For the thermostat represented in Equation 10, $\dot{Q}(\Gamma) = -\alpha(\Gamma) \frac{\partial \mathcal{H}}{\partial \mathbf{p}} \cdot \mathbf{S} \cdot \mathbf{p}$.

⁶These include the Gaussian isokinetic thermostat and the Nosé-Hoover thermostat. However, we note that Equation 12 has to be extended in the case of the Nosé-Hoover thermostat as detailed in Reference 13.

Because the same exclusive set of trajectories passes through both the phase volumes $d\Gamma_0$ and $d\Gamma_t$, the differential probability measures of the two infinitesimal volumes must be identical:

$$dP(d\Gamma_0, 0) = dP(d\Gamma_t, t). \quad (15)$$

We can express the probability measure P in terms of the probability density f as

$$dP(d\Gamma_s, s) = f(\Gamma_s, s) \delta V(\Gamma_s), \quad (16)$$

where $f(\Gamma_s, s)$ is the time-dependent phase-space probability density. The observation that the probability measure is conserved in phase space also leads to the Liouville equation for the probability density:

$$\frac{\partial f(\Gamma, s)}{\partial s} + \dot{\mathbf{q}} \frac{\partial f(\Gamma, s)}{\partial \mathbf{q}} + \dot{\mathbf{p}} \frac{\partial f(\Gamma, s)}{\partial \mathbf{p}} = -f(\Gamma, s) \Lambda(\Gamma). \quad (17)$$

We can recast Equation 17 into the following Lagrangian form,

$$\frac{d \ln f(\Gamma, s)}{ds} = -\Lambda(\Gamma), \quad (18)$$

from which it can be shown that

$$f(\Gamma_t, t) = f(\Gamma_0, 0) \exp\left(-\int_0^t \Lambda(\Gamma_s) ds\right). \quad (19)$$

This equation is also obtained directly from Equations 11, 15, and 16.

3. FLUCTUATION THEOREMS FROM DETERMINISTIC DYNAMICS

3.1. Evans-Searles Fluctuation Theorem

The Evans-Searles FT shows how irreversibility emerges naturally in systems whose equations of motion are time reversible. To fully appreciate the substance of this FT, we define two fundamental concepts: microscopic time reversibility and macroscopic irreversibility.

Microscopic time reversibility. The equations of motion in Section 2 describe the time evolution of a point, Γ , and may depend explicitly on the time owing to the possible time dependencies of λ and \mathbf{F}_e (that is, the equations of motion may be nonautonomous). If the equations of motion are reversible, then there exists a time-reversal mapping that transforms the point $\Gamma \equiv (\mathbf{q}, \mathbf{p})$ to Γ^* such that if we generate a trajectory starting at Γ_0 and terminating at Γ_t , then under the same dynamics, we start at $\Gamma_t^* \equiv (\Gamma_t)^*$ and arrive back at $\Gamma_0^* \equiv (\Gamma_0)^*$ after time t . We refer to a trajectory and its antitrajectory as a conjugate pair of trajectories. The time average of properties that are even under the mapping has equal values for the trajectory and its conjugate, whereas the time average of properties that are odd under the mapping has values with equal magnitude, but opposite signs for the trajectory and its conjugate. For many dynamics (e.g., Newtonian dynamics), the appropriate mapping gives $\Gamma^* = (\mathbf{q}, -\mathbf{p})$.

For the equations of motion⁷ (Equations 6 or 10) to satisfy this condition, we must have⁸

$$\begin{aligned}\phi(\mathbf{q}, \lambda(s)) &= \phi(\mathbf{q}, \lambda(t-s)), \\ \mathbf{C}(\Gamma) \cdot \mathbf{F}_e(s) &= -\mathbf{C}(\Gamma^*) \cdot \mathbf{F}_e(t-s), \\ \mathbf{D}(\Gamma) \cdot \mathbf{F}_e(s) &= \mathbf{D}(\Gamma^*) \cdot \mathbf{F}_e(t-s).\end{aligned}\quad (20)$$

Let us now consider a system whose overall equations of motion are time reversible. For every trajectory that is initiated at Γ_0 and terminates at Γ_t in a system with microscopically time-reversible dynamics, there exists a unique antitrajectory that starts at the phase-space point Γ_t^* at $s = 0$ and ends at Γ_0^* at $s = t$. The bundle of antitrajectories at time t passes through the volume element $d\Gamma_0^*$ centered about the point Γ_0^* . However, the size of the volume element $d\Gamma_t^*$ is equal to that of $d\Gamma_t$. Moreover, if there is a volume contraction from $d\Gamma_0$ to $d\Gamma_t$ (as shown in **Figure 1**), then there is an equivalent volume expansion associated with the bundle of antitrajectories.

The question of how microscopically time-reversible dynamics gives rise to observable macroscopic irreversibility is Loschmidt's paradox. To resolve this, we require an unambiguous measure of macroscopic irreversibility that is consistent with classical thermodynamics in the thermodynamic limit and applies to microscopic time-reversible equations of motion.

Macroscopic irreversibility: the dissipation function. A system undergoes a macroscopically reversible process in the time interval $0 \leq s \leq t$ if two conditions are satisfied. (a) The system is ergodically consistent. That is, for every trajectory that initiates at Γ_0 , the starting coordinates of its respective antitrajectory, Γ_t^* , are represented in the phase space of the system at $s = 0$, or equivalently the probability density of the initial coordinates of antitrajectories at time $s = 0$ is nonzero: $f(\Gamma_t^*, 0) \neq 0$, for all Γ_0 . (b) The probability of observing any bundle of trajectories, occupying an infinitesimal volume, is equal to the probability of observing the conjugate bundle of antitrajectories, or

$$dP(d\Gamma_0, 0) = dP(d\Gamma_t^*, 0). \quad (21)$$

We can write the latter condition for macroscopic reversibility more conveniently in terms of the distribution function of the phase space: $f(\Gamma_0, 0)\delta V(\Gamma_0) = f(\Gamma_t^*, 0)\delta V(\Gamma_t^*)$ or

$$\frac{f(\Gamma_0, 0)}{f(\Gamma_t^*, 0)} \left| \frac{d\Gamma_0}{d\Gamma_t^*} \right| = 1.$$

⁷In the case of Nosé-Hoover equations of motion, we have an extra degree of freedom owing to the thermostat multiplier, $\alpha(t)$. This phase-space variable must be reversed along with the momentum upon applying the time-reversal mapping Γ^* .

⁸Often there is an external symmetry; for example, for a fluid we may be able to drive a process in the opposite direction and consider it equivalent to the original direction, such that we may consider the conditions $\mathbf{C}(\Gamma) \cdot \mathbf{F}_e(s) = \mathbf{C}(\Gamma^*) \cdot \mathbf{F}_e(t-s)$ and $\mathbf{D}(\Gamma) \cdot \mathbf{F}_e(s) = -\mathbf{D}(\Gamma^*) \cdot \mathbf{F}_e(t-s)$ to in effect provide time-reversal symmetry. This is why the Evans-Searles FT can often allow protocols which have an odd time parity.

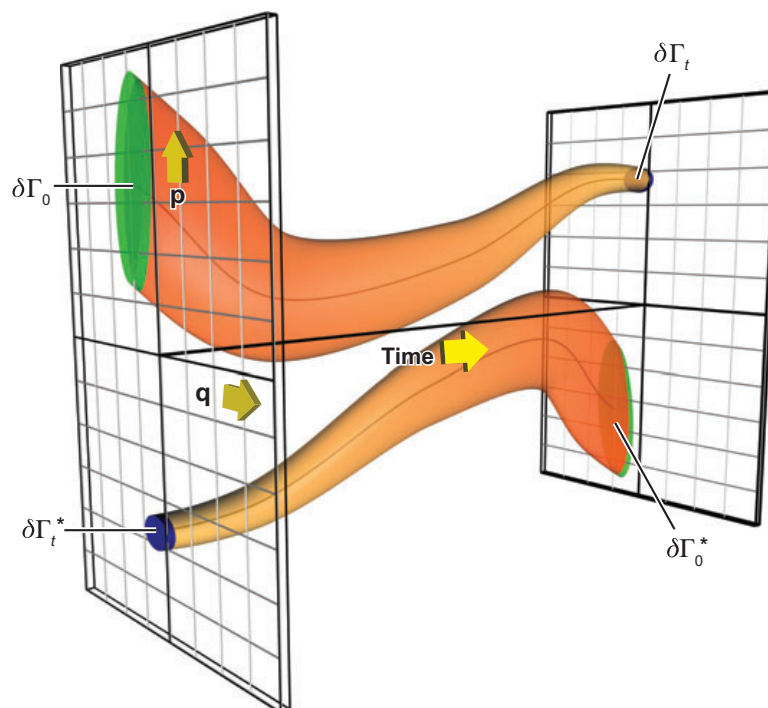


Figure 1

A set of neighboring deterministic trajectories of duration time, t , (*top tube*) and their corresponding set of time reverse or antitrajectories (*bottom tube*) in coordinate, momentum $\Gamma \equiv \{\mathbf{q}, \mathbf{p}\}$ and time, s , space in which an external agent does work on the system. This external agent (either λ parameter in the potential energy or a purely dissipative field, \mathbf{F}_e incorporated in the equations of motion) must have even time parity for the dynamics to remain time reversible. For every trajectory that starts at $\Gamma_0 = \{\mathbf{q}_0, \mathbf{p}_0\}$ in the volume element $\delta\Gamma_0$ and ends at $\Gamma_t = \{\mathbf{q}_t, \mathbf{p}_t\}$ in volume element $\delta\Gamma_t$ at some time t later, there exists the antitrajectory, whose coordinates, at any time s along the trajectory starting at $s = 0$, are given by $\Gamma_{t-s}^* = \{\mathbf{q}_{t-s}, -\mathbf{p}_{t-s}\}$. Thus, the antitrajectory starts at $\Gamma_t^* \equiv \{\mathbf{q}_t, -\mathbf{p}_t\}$ in volume element $\delta\Gamma_t^*$ and terminates after a time $s = t$ at $\Gamma_0^* \equiv \{\mathbf{q}_0, -\mathbf{p}_0\}$ in $\delta\Gamma_0^*$. For thermostatted systems, if the action of the external agent does work on the system, there is a contraction of phase-space volume in time (i.e., $\delta\Gamma_t < \delta\Gamma_0$ as represented in the figure by the shrinking of the tube's cross-sectional area in time). As the equations of motion are time reversible, the phase-space volume increases from $s = 0$ to $s = t$, and the size of the volume elements $\delta\Gamma_t$ and $\delta\Gamma_t^*$ is equal.

The volume of $d\Gamma_t^*$ is the same as $d\Gamma_t$. Thus, from Equation 11, we see that the condition for macroscopic reversibility (Equation 21) becomes

$$\ln \left[\frac{f(\Gamma_0, 0)}{f(\Gamma_t^*, 0)} \right] - \int_0^t \Lambda(\Gamma_s) ds = 0 \quad (22)$$

for any initial coordinate Γ_0 . Indeed, a quantitative measure of irreversibility may be defined as the inequivalence of Equation 22, which we denote Ω_t or the dissipation function

$$\begin{aligned} \Omega_t(\Gamma_0) &= \ln \left[\frac{dP(d\Gamma_0, 0)}{dP(d\Gamma_t^*, 0)} \right] \\ &= \ln \left[\frac{f(\Gamma_0, 0)}{f(\Gamma_t^*, 0)} \right] - \int_0^t \Lambda(\Gamma_s) ds. \end{aligned} \quad (23)$$

The dissipation function Ω_t is completely determined for a deterministic trajectory by the initial coordinate, Γ_0 , and the duration of the trajectory, t . We note here that the time reversibility of the dynamics dictates that conjugate pairs of trajectories are characterized by the same magnitude of Ω_t , but of opposite sign:

$$\Omega_t(\Gamma_t^*) = -\Omega_t(\Gamma_0). \quad (24)$$

Furthermore, if $\Omega_t = 0$ for all trajectories initiated anywhere in phase space, then the system is in equilibrium, and the probabilities of observing any trajectory and its corresponding antitrajectory are equal. If $\Omega_t > 0$ for a trajectory, then the corresponding antitrajectory is less likely to be seen, and if the ensemble average is greater than zero, $\langle \Omega_t \rangle > 0$, we have macroscopic dynamics moving in the forward direction. If $\langle \Omega_t \rangle < 0$, then we have macroscopic dynamics in the reverse direction. Thus $\langle \Omega_t \rangle \neq 0$ is the condition for macroscopic irreversibility. Our knowledge of the second law, however, suggests that the arrow of time points unambiguously in one firm direction; accordingly,

$$\langle \Omega_t \rangle \geq 0. \quad (25)$$

We explain how this comes about below.

The Evans-Searles fluctuation theorem. We consider trajectories of duration t in phase space by selecting all those initial coordinates Γ_0 for which Ω_t takes on some value \mathcal{A} between $\mathcal{A} \pm d\mathcal{A}$ and thus obtain the probability density

$$p(\Omega_t = \mathcal{A}) = \int d\Gamma_0 \delta[\Omega_t(\Gamma_0) - \mathcal{A}] f(\Gamma_0, 0). \quad (26)$$

Recognizing that Γ_0 is merely a dummy variable of integration, the conjugate probability density is

$$p(\Omega_t = -\mathcal{A}) = \int d\Gamma_t^* \delta[\Omega_t(\Gamma_t^*) + \mathcal{A}] f(\Gamma_t^*, 0). \quad (27)$$

Using the definition of Ω_t in Equation 23 along with Equation 24 and Equation 11, we have

$$p(\Omega_t = -\mathcal{A}) = \exp[-\mathcal{A}] \int d\Gamma_0 \delta[\Omega_t(\Gamma_0) - \mathcal{A}] f(\Gamma_0, 0), \quad (28)$$

which leads to the Evans-Searles FT:

$$\frac{p(\Omega_t = \mathcal{A})}{p(\Omega_t = -\mathcal{A})} = \exp[\mathcal{A}]. \quad (29)$$

Using this, we can average over all values of \mathcal{A} to give the second law inequality or $\langle \Omega_t \rangle \geq 0$ (5). In this derivation of the Evans-Searles FT, it was assumed that (a) the dynamics is ergodically consistent with the initial distribution function, (b) $f(\Gamma, 0) = f(\Gamma^*, 0)$, and (c) the dynamics are deterministic and microscopically reversible.

For systems of particles, the third condition above implies that the time-dependent λ and \mathbf{F}_t must have an even time parity $0 \leq s \leq t$. (If symmetry permits, an odd time parity is also acceptable.) These are sufficient conditions for the Evans-Searles FT to be valid, but the condition of microscopic reversibility can be relaxed to some degree, and stochastic versions of the Evans-Searles FT (7, 17) exist.

Microscopic reversibility:
the equations of motion are reversible

Dissipation function for systems initially in a canonical ensemble. Similar to the work $W(\Gamma_0, t)$, we can express the dissipation function in terms of a trajectory of duration t with an initial coordinate Γ_0 [i.e., $\Omega_t(\Gamma_0)$]. Analogous to W , we restrict ourselves to trajectories initiated at equilibrium in the canonical ensemble and consider the action of both a time-dependent λ -controlled potential where $\lambda(s=0) = A$ initially and $\lambda(s=t) = B$ finally and a time-dependent dissipative field, \mathbf{F}_e . For the definition of the dissipation function, Equation 23 becomes

$$\Omega_t(\Gamma_0) = \ln \frac{f_{eq}(\Gamma_0, \lambda = A)}{f_{eq}(\Gamma_t^*, \lambda = A)} - \int_0^t ds \Lambda(\Gamma_s) \quad (30)$$

$$= \beta [\mathcal{H}(\Gamma_t^*, \lambda = A) - \mathcal{H}(\Gamma_0, \lambda = A)] - \beta \int_0^t ds \dot{Q}(\Gamma_s). \quad (31)$$

Noting that the coordinates of the trajectory and antitrajectory, Γ_s and Γ_s^* , differ only in the direction of momenta, \mathbf{p} , and that \mathcal{H} is even in momenta, the left-hand side can be cast as a time integral over the trajectory

$$\Omega_t(\Gamma_0) = \beta \int_0^t ds [\dot{\mathcal{H}}(\Gamma_s, \lambda(s)) - \dot{Q}(\Gamma_s, s) - \dot{\phi}(\mathbf{q}, \lambda(s)) + \dot{\phi}(\mathbf{q}, \lambda = A)],$$

so that

$$\Omega_t(\Gamma_0) = \beta \int_0^t ds [\dot{\mathcal{H}}^{ad}(\Gamma_s, \lambda(s)) - \dot{\phi}(\mathbf{q}, \lambda_s) + \dot{\phi}(\mathbf{q}, \lambda = A)] \quad (32)$$

$$= \beta W(\Gamma_0, t) - \beta \int_0^t ds [\dot{\phi}(\mathbf{q}, \lambda_s) + \dot{\phi}(\mathbf{q}, \lambda = A)], \quad (33)$$

or explicitly in terms of the potential and the external field as

$$\Omega_t(\Gamma_0) = \beta \int_0^t ds \left[-\mathbf{J} \cdot \mathbf{F}_e V + \dot{\mathbf{q}} \left(\frac{\partial \phi(\mathbf{q}, \lambda(s))}{\partial \mathbf{q}} - \frac{\partial \phi(\mathbf{q}, \lambda = A)}{\partial \mathbf{q}} \right) \right]. \quad (34)$$

3.2. Crooks Fluctuation Theorem for Deterministic Dynamics

We can generate probability distribution functions for W , the work done on the system, in terms of a λ -dependent potential and/or a dissipative field, \mathbf{F}_e , in the same way we generate probability distributions in Section 2 for Ω_t . In contrast to the Evans-Searles FT, the Crooks FT considers the probability of observing trajectories from two different equilibrium states. The probability density for a trajectory of duration t , initiated at equilibrium with $\lambda = A$, is

$$p_f(W = \mathcal{A}) = \int d\Gamma_0 \delta[W(\Gamma_0, t) - \mathcal{A}] f_{eq}(\Gamma_0, \lambda = A), \quad (35)$$

where $W(\Gamma_0, t)$ denotes the work done over a trajectory of duration t , initiated at Γ_0 ; and $f_{eq}(\Gamma_0, \lambda = A)$ is the equilibrium distribution in the canonical ensemble. Now the reverse trajectory or antitrajectory starts at coordinates Γ_t^* and is guaranteed under deterministic dynamics to give a value of work equal and opposite that of the forward trajectory (see **Figures 2** and **3**). In this case, the reverse trajectory must also initiate under equilibrium conditions, however, with $\lambda = B$, and the time dependence of the parameter λ and field \mathbf{F}_e must be reversed. Thus, the probability density for the

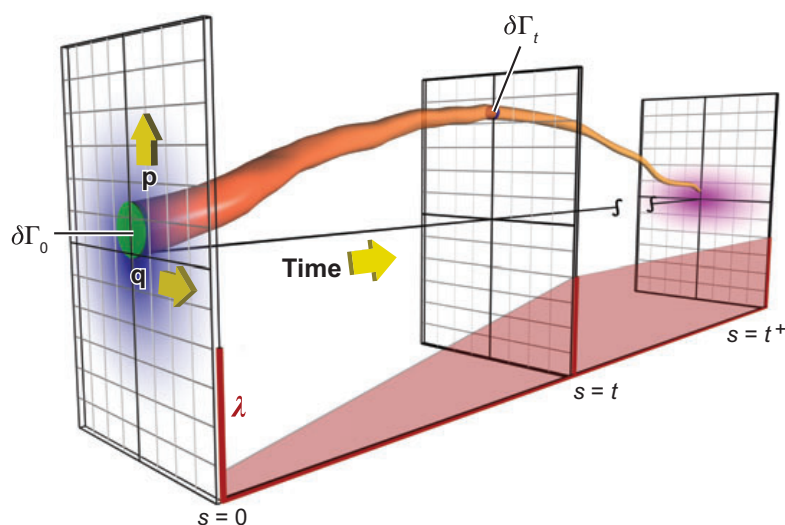


Figure 2

A set of deterministic trajectories generated under the action of an external agent, in which the agent does work on the system of magnitude $W \pm dW$. The external agent is represented as a time-dependent λ parameter that controls the system's potential energy and equilibrium state. The trajectories are initiated under equilibrium conditions under a constant $\lambda(s = 0)$, for example, equal to A ; the shading in the $s = 0$ phase-space plane represents the equilibrium distribution, $f_{eq}(\Gamma, \lambda = A)$. Here we represent a linearly time-dependent λ that varies over the time interval $0 \leq s \leq t$, with $W = \int_0^t ds \lambda \partial \mathcal{H} / \partial \lambda$. In this case, λ is not applied quasistatically; in other words, at time t when the field has attained its final value of $\lambda = B$, the system is far from equilibrium and relaxes over $t \leq s \leq t^+$ toward a new equilibrium state characterized by $\lambda = B$. The magnitude of $f_{eq}(\Gamma, \lambda = B)$, the equilibrium distribution for $\lambda = B$, is represented by the shading in the $s = t^+$ plane. The change in free energy brought about by $\lambda = A \rightarrow \lambda = B$ is determined by the ratio of the integrals over Γ of f_{eq} at $\lambda = B$ and $\lambda = A$.

time-reverse trajectory is

$$p_r(W = -A) = \int d\Gamma_t^* \delta[W(\Gamma_t^*, t) + A] f_{eq}(\Gamma_t^*, \lambda = B). \quad (36)$$

At this point it is useful to note that if the system is driven strongly (i.e., far from equilibrium), the destination coordinate in the forward trajectory, Γ_t , may not be significantly weighted in the equilibrium distribution associated with the initial coordinates of the reverse trajectory; in other words,

$$f_{eq}(\Gamma_t, \lambda = B) = \frac{\exp[-\beta \mathcal{H}_{\lambda=B}(\Gamma_t)]}{Z_B} \quad (37)$$

may be very small. That is, the reverse trajectory can be rare, creating a difficult challenge in sampling the distribution $p_r(W = -A)$ in the Crooks FT, causing the convergence of the ensemble average in Jarzynski's equality to become very slow (18–22). Recasting $\mathcal{H}_{\lambda=B}$ in terms of the work done on the system, using the first law,

$$\begin{aligned} W(\Gamma_t^*, t) &= -W(\Gamma_0, t) = -\int_0^t ds \dot{\mathcal{H}}^{ad}(\Gamma_s, \lambda(s)) \\ &= [\mathcal{H}_{\lambda=A}(\Gamma_0) - \mathcal{H}_{\lambda=B}(\Gamma_t)] + k_B T \int_0^t ds \Lambda(\Gamma_s), \end{aligned}$$

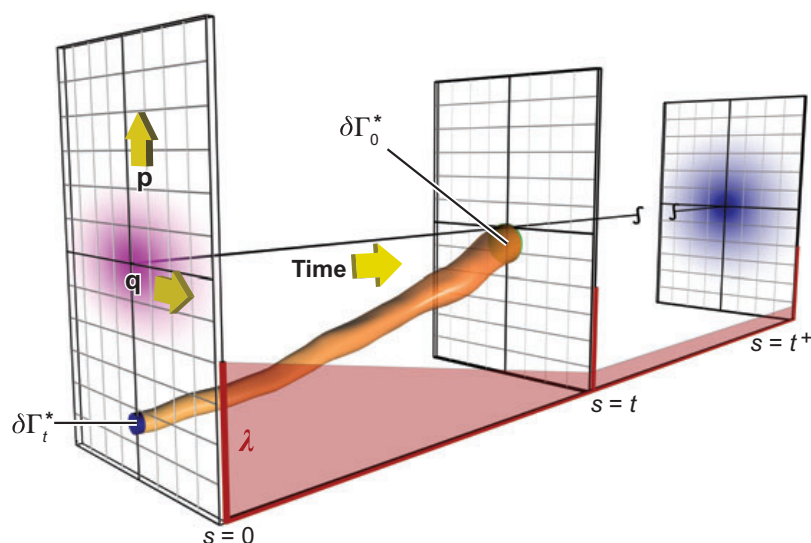


Figure 3

A set of time-reversed trajectories, conjugate to the trajectory segments in the time interval $0 \leq s \leq t$ in **Figure 2**. These trajectories are constructed under a time-reverse mapping as explained in **Figure 1**, in which the field is also time reversed, $\lambda(s = 0) = B \rightarrow \lambda(s = t) = A$. As these are generated under deterministic dynamics, this bundle of reverse trajectories is characterized by the same magnitude, but with the opposite sign of W as in **Figure 2** (i.e., $-W \mp W$). We note that the initial coordinates of these reverse coordinates are not significantly weighted in Z_B , the partition function of the equilibrium state at $\lambda = B$; in other words, $f_{eq}(\Gamma_t^*, \lambda = B)$ is small, and these reverse trajectories, sampled from an initial equilibrium state, are rare. These reverse trajectories become less rare as the difference in A and B becomes small, and $\dot{\lambda}(s)$ over the interval $0 \leq s \leq t$ is reduced. This provides a practical challenge when sampling the distribution $p_r(W = -\mathcal{A})$ in the Crooks fluctuation theorem (FT) or ensuring convergence in the average $\langle \exp[-\beta W] \rangle_f$ in Jarzynski's equality. However, it does not negate the validity of either relation. This same sampling problem can occur with Evans-Searles FT as well. However, in Evans-Searles FT, the external agents must have even time parity; it is the fast application of external agents that renders rare trajectories with $\Omega_t < 0$ that can be difficult to sample in experiment or simulation.

and noting that \mathcal{H} is an even function of the momenta, i.e., $\mathcal{H}_\lambda(\Gamma_s^*) = \mathcal{H}_\lambda(\Gamma_s)$,

$$f_{eq}(\Gamma_t^*, \lambda = B) = \frac{Z_A}{Z_B} f_{eq}(\Gamma_0, \lambda = A) \exp[\beta W(\Gamma_t^*, t)] \frac{\delta V(\Gamma_0)}{\delta V(\Gamma_t^*)},$$

where we use the phase-space compression factor given above. Now $\delta V(\Gamma_s^*) = \delta V(\Gamma_s)$, and the Jacobian is such that

$$p_r(W = -\mathcal{A}) = \frac{Z_A}{Z_B} \int d\Gamma_0 \delta[W(\Gamma_t^*, t) + \mathcal{A}] \exp[\beta W(\Gamma_t^*, t)] f_{eq}(\Gamma_0, \lambda = A). \quad (38)$$

Furthermore, as forward and reverse trajectories have equal but opposite values of W under time-reversible dynamics, $W(\Gamma_t^*, t) = -W(\Gamma_0, t)$,

$$p_r(W = -\mathcal{A}) = \exp[\beta(\Delta F - \mathcal{A})] \int d\Gamma_0 \delta[W(\Gamma_0, t) - \mathcal{A}] f_{eq}(\Gamma_0, \lambda = A).$$

We can identify the integral on the right-hand side as $p_f(W = \mathcal{A})$, resulting in the Crooks FT:

$$\frac{p_f(W = \mathcal{A})}{p_r(W = -\mathcal{A})} = \exp[\beta(\mathcal{A} - \Delta F)]. \quad (39)$$

In the above derivation of the Crooks FT for the deterministic system, it was assumed that (a) the dynamics is such that for any phase point Γ for which $f(\Gamma, \lambda = \mathcal{A}) \neq 0$, then $f(\Gamma^*, \lambda = B) \neq 0$; (b) $f(\Gamma, 0) = f(\Gamma^*, 0)$; and (c) when the time evolution of λ and \mathbf{F}_e is reversed, the dynamics remains deterministic and microscopically time reversible. These are sufficient conditions for the Crooks FT to be valid, but the condition of reversibility can be relaxed to some degree, and stochastic versions of this FT exist.

3.3. Steady-State Fluctuation Theorems

Above we focus on FTs that apply to a system driven out of an initial equilibrium state by an external field, characterized by \mathbf{F}_e , or a parametric change in the potential characterized by $\dot{\lambda} \neq 0$. Indeed, the Evans-Searles FT applied to systems driven from a known initial state over transient trajectories is often referred to as the transient fluctuation theorem. However, according to the derivations of the Evans-Searles FT, the initial phase-space distribution is not restricted to time-invariant or even equilibrium distributions. The only requirement the Evans-Searles FT places on the initial distribution function is that it is known and expressible in the ostensible dimension of the equations of motion (this is not the case for the Crooks FT). Here we consider the FTs applied to trajectories under a steady state (i.e., the system is acted upon by a purely dissipative, constant external field, \mathbf{F}_e).

There are two steady-state fluctuation theorems (SSFTs) that appear in the literature. We can trace both back to the original paper on FTs (23, 24) that focused on isoenergetic equations of motion, but it is only later that the two separate theorems were distinguished: the steady-state version of the Evans-Searles FT (4) and the Gallavotti-Cohen FT (25).

The Evans-Searles steady-state fluctuation theorem. In its simplest formulation, the Evans-Searles SSFT involves a rearranged form of Equation 1 applied in the long time limit to trajectories of a system wholly in a nonequilibrium steady state (i.e., the distribution function is time invariant). A more complete derivation of the SSFT is available (26); however, here we provide a simpler presentation that is physically compelling and suitable for those primarily interested in a scientific justification.

The argument of the Evans-Searles FT applied to the steady state is

$$\Omega_t^{\text{ss}}(\Gamma_0) = \ln \left[\frac{f^{\text{ss}}(\Gamma_0, 0)}{f^{\text{ss}}(\Gamma_t^*, 0)} \right] - \int_0^t \Lambda(\Gamma_s) ds, \quad (40)$$

where $f^{\text{ss}}(\Gamma_0, 0)$ is now the phase-space distribution function associated with a steady state, rather than with an equilibrium state (as in the Evans-Searles FT applied to transient trajectories). However, typically this definition of Ω_t^{ss} is difficult to implement. Steady-state distribution functions for the types of deterministic dynamics under

consideration are not generally known. What is known is that, in the steady state, the dynamics approaches a strange attractor that has a different fractal dimension to the ostensible phase space. Even if we knew the details of this attractor, it would still be difficult to apply the Evans-Searles FT as it describes bundles of phase-space trajectories in the phase-space dimension and not in the dimension of the strange attractor. However, there are special cases in which these steady-state distribution functions can be expressed simply and exactly under stochastic equations of motion, and Equation 1 may be applied (27).

In general these steady-state distribution functions are not known, and consequently, it is not possible to construct exact expressions for Ω_t^{ss} for deterministic trajectory segments of duration t that are wholly at a nonequilibrium steady state. However, one can construct an approximate steady-state dissipation function from trajectories initiated at a known equilibrium, in the absence of the dissipative field, \mathbf{F}_t . [This distribution function is often referred to as the Kawasaki distribution function (13) and can be considered to form the basis of the formal proof (26).] At time $t = 0$, the dissipative field is introduced, and we express Ω_t in terms of its instantaneous rate of change, $\Omega(s)$, at time s :

$$\Omega_t = \int_0^\tau ds \Omega(s) + \int_\tau^t ds \Omega(s). \quad (41)$$

Here, τ is some arbitrary long time (e.g., several Maxwell times), so that the fluid has completely relaxed into a steady state. Thus Ω_t is cast as a sum of transient and steady-state contributions with the steady-state contribution, identified as the steady-state dissipation function, Ω_t^{ss} , used to approximate Ω_t with an error or order $\mathcal{O}(\tau)$. It is instructive to express these dissipation functions as time averages, $\bar{\Omega}_t = \Omega_t/t$, such that

$$\bar{\Omega}_t^{\text{ss}} \approx \bar{\Omega}_t + \mathcal{O}\left(\frac{\tau}{t}\right). \quad (42)$$

We make the physically compelling argument that, in the long time limit, the distribution function for steady-state trajectories asymptotically converges to that for the full transient trajectories:

$$\lim_{t \rightarrow \infty} p^{\text{ss}}(\bar{\Omega}_t^{\text{ss}}) = p(\bar{\Omega}_t). \quad (43)$$

Finally, the fluctuations in $\bar{\Omega}_t^{\text{ss}}$ vanish in the long time limit, and for the SSFT to be of any importance, it is necessary that these fluctuations vanish more slowly than $\mathcal{O}(\tau/t)$, the error in the $\bar{\Omega}_t^{\text{ss}}$ approximation, as detailed in Reference 27. Consequently, we can approximate $\bar{\Omega}_t$ in the FT (Equation 1) with the steady-state dissipation function $\bar{\Omega}_t^{\text{ss}}$, leading to the SSFT or

$$\lim_{t \rightarrow \infty} \frac{1}{t} \ln \frac{p(\bar{\Omega}_t = \mathcal{A})}{p(\bar{\Omega}_t = -\mathcal{A})} = \mathcal{A}. \quad (44)$$

Gallavotti-Cohen fluctuation theorem. We can write the fluctuation relation of the Gallavotti-Cohen FT as

$$\lim_{t \rightarrow \infty} \frac{1}{t} \ln \frac{p(-\bar{\Lambda}_t = \mathcal{A})}{p(-\bar{\Lambda}_t = -\mathcal{A})} = \mathcal{A}, \quad (45)$$

where the time-average phase-space compression factor (or the divergence of the flow), measured in the steady state, is given as

$$\bar{\Lambda}_t = \frac{1}{t} \int_0^t ds \Lambda(s).$$

This FT (23) was originally proposed for the special case of isoenergetic dynamics, for which $\Omega(\Gamma) = -\Lambda(\Gamma)$. However, subsequently, Gallavotti & Cohen's (25, 28) work strongly suggests that under appropriate conditions, Equation 45 [with a restriction on the values of \mathcal{A} (29)⁹] can be applied to a larger class of dynamical systems (e.g., constant temperature systems). They arrived at Equation 45 through a formal derivation (25, 28) that drew upon the Sinai-Ruelle-Bowen measure (for a discussion, see 31), which requires the dynamics to be an Anosov diffeomorphism (32). (An axiom A diffeomorphism also suffices for the Gallavotti-Cohen FT.) In general, the equations of motion (Equation 10) do not form an Anosov diffeomorphism. To address this, Gallavotti & Cohen introduced a new hypothesis, termed the chaotic hypothesis (33),¹⁰ which, for the purposes of the Gallavotti-Cohen FT, allows many-body dynamics to be treated as an Anosov diffeomorphism. Unfortunately, as yet, there is no way to independently ascertain if a physical system may be treated as Anosov diffeomorphic. The requirements for the valid application of the Gallavotti-Cohen FT to physical systems are therefore extremely difficult to establish.

A necessary but insufficient condition for Anosov diffeomorphism is that the dynamical system must be hyperbolic (31). This means that the number of expanding and contracting directions on the attractive set must be equal, or in other words, the number of positive and negative finite-time Lyapunov exponents must be equal and no zero exponents are allowed.

There is a large body of computer simulation results, for various processes, that has tested the SSFTs (e.g., 17, 23, 24, 34–46). (There are results in the literature that test the Evans-Searles FT while erroneously claiming to test the Gallavotti-Cohen FT.) We know of no case for which the Gallavotti-Cohen FT converges faster than the Evans-Searles FT. For temperature-regulated dynamics, when the dissipative field strength is very small, the Gallavotti-Cohen FT can take extremely long times to converge. Indeed as the dissipative field strength approaches zero, the amount of time it takes the Gallavotti-Cohen FT to converge diverges (30, 47). To understand this, we consider the arguments of the Evans-Searles FT and the Gallavotti-Cohen FT. When the field strength approaches zero, so does the instantaneous dissipation function. More precisely, the average value of the instantaneous dissipation function, to leading order, is $\Omega(\Gamma) = 0 + \mathcal{O}(F_e^2)$, and the standard deviation is $\sigma = 0 + \mathcal{O}(F_e)$. For the phase-space compression factor, the mean is $\Lambda = 0 + \mathcal{O}(F_e^2)$ and the standard deviation is $\sigma = \sigma_0 + \mathcal{O}(F_e^2)$, where σ_0 is the amplitude of the standard deviation at

⁹Equation 45 is restricted to values of \mathcal{A} bounded by a value \mathcal{A}^* : $\mathcal{A} \in (-\mathcal{A}^*, \mathcal{A}^*)$. In the small field limit, this value is given to leading order as $\mathcal{A}^* = 0 + \mathcal{O}(F_e^2)$ (30).

¹⁰Quoting from Reference 33, "Chaotic hypothesis: for the purpose of studying macroscopic properties, the time evolution map S of a many-particle system can be regarded as a mixing Anosov map." In References 25 and 28, the term "transitive Anosov map" was used to mean "mixing Anosov map."

equilibrium. The difference between the behavior in the amplitude of the fluctuations, σ , for these two quantities is crucial. As F_e approaches zero, so too does the amplitude of the fluctuations in Ω but not those in Λ . Now the form of the fluctuation formulae is asymmetric. In the limit $F_e \rightarrow 0$, Equation 44 (the SSFT) remains consistent with a given trajectory segment being equally likely to occur as its antitrajectory segment. This is a necessary condition for equilibrium. In contrast, Equation 45 (the Gallavotti-Cohen FT) is not consistent with this, owing to the fluctuations in Λ remaining finite when $F_e \rightarrow 0$. One resolution is for the time averaging, or the time for which the Gallavotti-Cohen FT is given to converge, to be so long that there are no significant fluctuations remaining. However, theory (30) specifies that the largest fluctuations for which the Gallavotti-Cohen FT may be validly applied vanish in the small field limit.

3.4. The Einstein Relation and Green Kubo Theory

Above we review the Evans-Searles FT as well as the Crooks FT as recent theorems in nonequilibrium statistical mechanics. Here, we show that the FTs, and in particular the Evans-Searles FT, are consistent with the long-standing and well-known relations in the field, namely the Einstein-Sutherland relation (48–50) and the Green-Kubo relations. The Einstein-Sutherland relation dates back to the early days of nonequilibrium statistical mechanics; we can write it as

$$\langle \mathbf{v} \rangle_{F_e} = \beta D \mathbf{F}_e. \quad (46)$$

This important relation describes the average steady-state velocity of a particle, $\langle \mathbf{v} \rangle_{F_e}$, under an applied field, \mathbf{F}_e , to the variance in the particle's displacement over time in the absence of the field, which is commonly referred to as the diffusion constant, $D = \lim_{t \rightarrow \infty} \langle \Delta x(t)^2 \rangle_0 / (2t)$, where $\Delta x(t) = \int_0^t ds \dot{x}(s)$. Starting from the Evans-Searles FT, we reformulate a generalized form of this Einstein-Sutherland relation, and, from that, the Green-Kubo relations. Although this does not produce new results, it demonstrates the FTs' consistency with important existing theorems in nonequilibrium statistical mechanics, and it also emphasizes and clarifies the conditions necessary for the application of these theories, as we show below, in the case of supercooled liquids.

To derive the more generalized version of the Einstein-Sutherland relation (Equation 46) from the FT, we identify the product of the particles' drift velocity and the applied field, $-\mathbf{F}_e \cdot \mathbf{v}(t)$, as a specific example of a dissipative field flux, which we represent by $JV \cdot \mathbf{F}_e$, which, in the case of the flux and constant field being in the same direction, we write more simply as JVF_e . Under steady state, the time-averaged dissipative flux is defined as

$$\bar{J}_t \equiv \frac{1}{t} \int_{\tau}^{\tau+t} ds J(s), \quad (47)$$

where τ is a time long enough after the application of the field so that the system is at a steady state. The SSFT (Equation 44) may then be written as

$$\lim_{t \rightarrow \infty} \frac{1}{t} \ln \frac{P(\bar{J}_t = \mathcal{A})}{P(\bar{J}_t = -\mathcal{A})} = -\beta VF_e \mathcal{A}. \quad (48)$$

In the limit of long time, we may invoke the central limit theorem, which states that close to the mean, the distribution of \bar{J}_t is Gaussian. Additionally, in the limit of small field strength, values of \bar{J}_t close to the mean dominate:

$$P(\bar{J}_t = \mathcal{A}) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(\mathcal{A} - \langle\bar{J}_t\rangle_{F_e})^2}{2\sigma^2}\right). \quad (49)$$

Now as the variance in the distribution of \bar{J}_t is independent of the field direction or sign of F_e , then, to leading order in F_e , the variance behaves as

$$\sigma_{F_e}^2 = \sigma_0^2 + \mathcal{O}(F_e^2) = \langle\bar{J}_t^2\rangle_0 + \mathcal{O}(F_e^2). \quad (50)$$

The SSFT (Equation 48), the central limit theorem (Equation 49), and Equation 50 combine to give a generalized Einstein-Sutherland relation:

$$\langle J \rangle_{F_e} = \lim_{t \rightarrow \infty} -\frac{1}{2} \beta V F_e \langle \bar{J}_t^2 \rangle_0 t + \mathcal{O}(F_e^2). \quad (51)$$

The Einstein-Sutherland relation is only valid to linear order in the field, F_e . However, the FT provides more detailed understanding of how this relation fails under large fields. When the field is increased, the mean dissipative flux, $\langle\bar{J}_t\rangle$, also increases (**Figure 4**). When the mean is large relative to the standard deviation, then for every typical value of the flux $\bar{J}_t = \mathcal{A}$, its conjugate value $\bar{J}_t = -\mathcal{A}$ in the SSFT is represented in the wings of the distribution in which the central limit theorem no longer applies. In this instance, the generalized Einstein-Sutherland relation is invalid, even if the equilibrium variance $\langle\bar{J}_t^2\rangle_0$ is replaced with the variance under steady state, or $\langle\bar{J}_t^2\rangle_{F_e}$. Molecular dynamics simulations of planar shear have shown that the

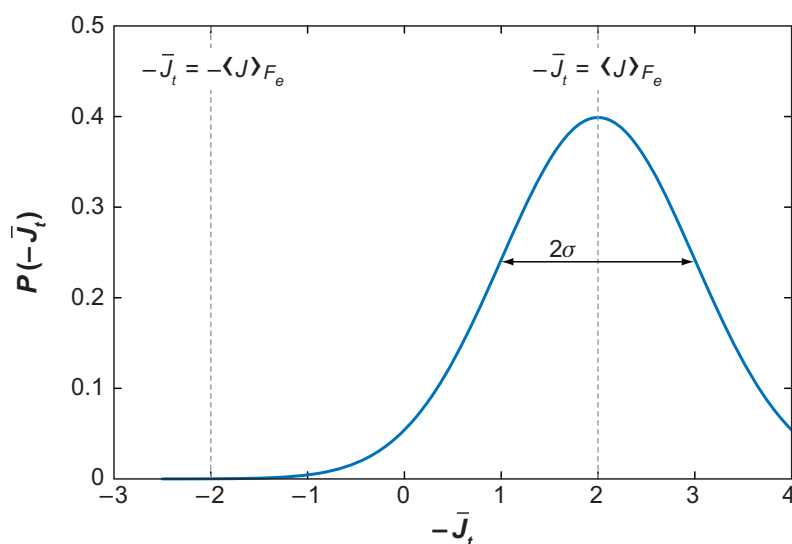


Figure 4

A typical Gaussian distribution, of unit variance $\sigma^2 = 1$, for the time-averaged flux \bar{J}_t with the ensemble average $\langle J \rangle_{F_e} = -2$ denoted by the dashed line. The other dashed line is at $\bar{J}_t = -\langle J \rangle_{F_e} = 2$, which is the value the fluctuation theorem compares to the mean. As time proceeds, in the limit $t \rightarrow \infty$, the variance decreases like $\sim 1/t$.

breakdown of the central limit theorem dominates the approximation of ignoring terms of $\mathcal{O}(F_e^2)$ or the variance that describes the distribution near its mean (36). In the case of a single, tagged particle, interacting with a constant field, and embedded in a supercooled liquid, the amount of time it takes for the steady-state FT to converge, as well as the time it takes for the distribution to become Gaussian, increases rapidly upon approach to the glass transition (51). Moreover, the variance decreases with time. As this time increases, the variance decreases inversely proportionally. As the nominal glass transition is approached, the strongest field for which a linear response may be observed in the steady state vanishes.

The variance of the flux may be expressed in terms of the integral,

$$\langle \bar{J}_t^2 \rangle_0 = \frac{1}{t^2} \int_0^t ds \int_0^t du \langle J(s)J(u) \rangle_0 \quad (52)$$

$$= \frac{2}{t} \int_0^t ds \langle J(0)J(s) \rangle_0 - \frac{2}{t^2} \int_0^t ds s \langle J(0)J(s) \rangle_0. \quad (53)$$

In the long time limit, the second term on the right-hand side vanishes and

$$\lim_{t \rightarrow \infty} \langle \bar{J}_t^2 \rangle_0 = \frac{2}{t} \int_0^t ds \langle J(0)J(s) \rangle_0. \quad (54)$$

Combining this with the generalized Einstein-Sutherland relation (Equation 51) gives the celebrated Green-Kubo theory for the steady state:

$$\langle J \rangle_{F_e} = -\beta V F_e \lim_{t \rightarrow \infty} \int_0^t ds \langle J(0)J(s) \rangle_0. \quad (55)$$

We can use this to obtain a transport coefficient in terms of equilibrium fluctuations in the form of an autocorrelation function. One might wonder why the Green-Kubo theory holds such important status given that the presentation here shows it to be equivalent to the Einstein-Sutherland relation. In contrast to the Einstein-Sutherland relation, the Green-Kubo theory is also applicable to time-dependent phenomena. A time-dependent version of the Green-Kubo theory cannot be obtained from the FT, which must satisfy definite time parity conditions.

The FT has been used in Couette flow or planar shear (23, 36) employing the SLLOD equations of motion (13), which, in the absence of a fictitious thermostat and for the case of constant shear rate $\dot{\gamma}$, are equivalent to Newton's equations of motion. As we are controlling the shear rate externally, we identify it as the external field $F_e = \dot{\gamma}$ and the flux as $J = P_{xy}$, where P_{xy} is the xy element of the pressure tensor. The dissipative field or entropy production for Couette flow is then

$$JV F_e = \dot{\gamma} V P_{xy}. \quad (56)$$

The shear viscosity η is the rate at which work is being done on the fluid divided by the product of the volume and the shear rate squared. The Green-Kubo expression for the viscosity is thus

$$\eta = \frac{-\langle P_{xy} \rangle}{\dot{\gamma}} = \beta V \int_0^\infty dt \langle P_{xy}(0)P_{xy}(t) \rangle_0, \quad (57)$$

and the Einstein-Sutherland expression is

$$\eta = \frac{1}{2} \beta V t \lim_{t \rightarrow \infty} \langle \bar{P}_{xy, t}^2 \rangle. \quad (58)$$

If the system is very viscous, the Green-Kubo expression (Equation 57) requires a long time to converge, and consequently, the generalized Einstein-Sutherland expression (Equation 58) is probably the better method to extract the viscosity in such a situation (52). In contrast, if we wish to calculate the self-diffusion coefficient for a very viscous system, the Einstein-Sutherland expression for the diffusion, $\langle \Delta x(t)^2 \rangle_0 / (2t)$, can be very slow to converge, whereas the Green-Kubo expression, $D = \int_0^\infty ds \langle v(0)v(s) \rangle_0$, usually converges quite rapidly (53).

4. APPLICATIONS OF FLUCTUATION THEOREMS TO EXPERIMENTAL AND MODEL SYSTEMS

Theoreticians and mathematicians interested in nonequilibrium statistical mechanics accomplished much of the work done in developing and extending the FTs. Until 2002, demonstrations of the theorems were limited to computer simulations, and there were no practical experimental demonstrations of the theorems, despite the range of interests in nano/micromachines or molecular devices that impose nanometer-scale displacements with pico-Newton scale forces. Such small machines include single biomolecules that act as molecular motors and whose experimental observation highlights the nonequilibrium phenomena described by the FTs. Linear motors, such as the actin-myosin or the kinesin-microtubule motor, are fueled by proton currents or ATP hydrolysis and function as integral parts of cellular metabolism, and consequently, they work under inherently nonequilibrium conditions (54). Over time, on average, these molecular engines must not violate the second law; however, occasionally they run backwards, converting heat from the surroundings to generate useful mechanical/chemical energy. This work, done on molecular time and length scales, has a natural variation or spread of values that is governed by the FTs.

In 2002, two independent groups experimentally demonstrated the FTs, each with a unique focus and both using optical tweezers. Wang et al. (55) demonstrated the Evans-Searles FT by monitoring the transient trajectory of a single colloidal bead in a translating optical trap. Simultaneously, Liphardt et al. (56) used optical tweezers to pull the ends of a DNA-RNA hybrid chain, measuring the work required to unravel or unfold a specific domain in the chain. These experiments had complementary aims: The colloidal experiment was a classical model system constructed to clearly demonstrate, as rigorously as possible in experiment, the Evans-Searles FT. In contrast, Liphardt's RNA-unfolding experiment importantly demonstrated the application of the Crooks FT to a complex biomolecular system, highlighting the potential practical use of FTs to a wider range of scientists. In this section, we review both experiments in some detail before more briefly mentioning other more recent experimental applications of the FTs, as well as other proposed experimental systems and implications.

Optical tweezers, or an optical trap, are 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 subpico-Newton force on the particle, drawing it toward the region of highest intensity, the focal point or trap center. The optical trap is harmonic; a particle located a distance \mathbf{r} from the center of the trap has an optical force, $\mathbf{f}_{opt} = -k\mathbf{r}$, acting to restore its position to the trap center. k is the trapping constant, which is determined by the distribution of particle positions at equilibrium and is tuned by adjusting the intensity of the laser. Using an objective lens of high numerical aperture, the optical trapping is strongest in the direction perpendicular to the focal plane, such that the particle remains localized entirely within the focal plane, fluctuating about the focal point. As the particle position \mathbf{r} is measured at kilohertz frequency (over timescales significantly large that inertia of the colloidal particle is negligible), the measured optical force, \mathbf{f}_{opt} , balances any applied forces, either forces arising from the surrounding solvent, such as Brownian or drag forces, or the tension associated with a tethered chain molecule such as DNA or RNA.

4.1. Single Colloidal Particle in an Optical Trap

Wang et al. (55) carried out the first experiment demonstrating the FTs. They monitored the trajectory of a single colloidal particle, weakly held in a stationary optical trap that was translated uniformly with constant, vanishingly small velocity \mathbf{v}_{opt} starting at time $s = 0$. Initially, the particle's position in the trap is distributed according to an equilibrium distribution with an average particle velocity of 0. With trap translation, the particle is displaced from its equilibrium position until, at some time later, the average velocity of the particle is equal to the velocity of the trap. From this point, the system is in a nonequilibrium steady state. The external agent is purely dissipative (i.e., $\dot{\lambda} = 0$ and $\mathbf{F}_e \sim \mathbf{v}_{opt}$), so the dissipation function is

$$\Omega_t = -\beta \int_0^t ds \mathbf{J}V \cdot \mathbf{F}_e = \beta \int_0^t ds \mathbf{f}_{opt} \cdot \mathbf{v}_{opt}. \quad (59)$$

With the ability to resolve nanometer-scale particle displacements and femto-Newton scale optical forces, they determined $\Omega(s)$ with sub- $k_B T$ resolution. As there is no change in the underlying state of the system, and the field has even time parity, the Evans-Searles FT and Crooks FT reduce to the same expression as $\Omega_t \equiv W$. We expect, from the second law, that work is done to translate the particle-filled optical trap, or $W > 0$, but according to the FTs, there should also be a nonvanishing probability of observing short trajectories where $W < 0$; that is, thermal fluctuations provide the work. Indeed, Wang et al. (55) showed that trajectories with $W = \Omega_t < 0$ persisted for 2–3 s, far longer than had been demonstrated by simulation. However, in this initial experiment, there was an insufficient number of trajectories to properly sample the distribution, $p(\Omega_t)$, and the authors instead tested a coarse-grained form of the FT, the integrated Evans-Searles FT:

$$\frac{p(\Omega_t < 0)}{p(\Omega_t > 0)} = \langle \exp[-\Omega_t] \rangle_{\Omega_t > 0}, \quad (60)$$

where the brackets on the right-hand side denote an average over that part of the distribution for which $\Omega_t > 0$. Later, Wang et al. (27) revised this experiment and sampled a larger number of trajectories, enabling a direct demonstration of the Evans-Searles FT. Moreover, they also translated the particle-filled optical trap in a circular or race-course pattern, producing one long single trajectory, which, outside the initial short time interval, was at steady state. Using contiguous segments of this single steady-state trajectory, they demonstrated the steady-state version of the Evans-Searles FT (27).

A particle in an optical trap was also used to demonstrate the distinction between the Evans-Searles FT and Crooks FT in the so-called capture experiment (57, 58). For this experiment, a particle is equilibrated in a stationary trap of strength k_0 . At time $s = 0$, the optical strength is increased discontinuously from k_0 to k_1 , $k_1 > k_0$, so the particle is more tightly confined or captured. Alternatively, the trap strength can be decreased from k_1 to k_0 , to release the particle. The external parameter, λ , is the time-dependent trap strength, $k(s)$, which varies discontinuously, $\dot{\lambda}(s) = (k_1 - k_0)\delta(s)$, and in the absence of a purely dissipative field, or $\mathbf{F}_e = 0$. The particle's position is recorded as it relaxes to its new equilibrium distribution, and W and Ω_t are evaluated over an ensemble of nonequilibrium trajectories. Work is the change in the internal energy that occurs with the change in the trapping constant:

$$W = \int_0^t ds \dot{\lambda} \frac{d\phi_{ext}}{d\lambda} = 1/2(k_1 - k_0)\mathbf{r}_0^2.$$

As all trajectories must initiate under equilibrium conditions under k_0 , the probability distribution W is then

$$p_{k_0 \rightarrow k_1}(W) = \sqrt{\frac{k_0}{\pi(k_1 - k_0)W}} \exp\left[-\frac{k_0 W}{k_1 - k_0}\right].$$

The ensemble average, $\langle \exp[-\beta W] \rangle$, confirms Jarzynski's equality, or $\langle \exp[-\beta W] \rangle = \exp[-\beta \Delta F]$, where $\Delta F = k_B T \ln[\sqrt{k_1/k_0}]$ from classical thermodynamics. Furthermore, if we consider the probability distribution of W for both the forward or capture direction and the reverse or release direction (i.e., $W_{k_1 \rightarrow k_0}$), then it is straightforward to show that the Crooks FT is obeyed. We note that in the context of the capture experiment, the Crooks FT depends only on the equilibrium distribution of initial particle positions: These equilibrium distributions are independent of the viscoelastic response of the suspending fluid, for example. In contrast, the dissipation function, Ω_t , depends sensitively on the nonequilibrium trajectory and on the material properties of the surrounding fluid (59). The dissipation function for capture is

$$\Omega_t = \beta \int_0^t ds \dot{\mathbf{r}} \left[\frac{d\phi_{ext}(\lambda = k_0)}{d\mathbf{r}} - \frac{d\phi_{ext}(\lambda(s))}{d\mathbf{r}} \right] = \frac{\beta(k_0 - k_1)}{2} (\mathbf{r}_t^2 - \mathbf{r}_0^2).$$

If we consider a cyclic protocol in which the λ parameter had even time parity—i.e., $\dot{\lambda}(s) = (k_1 - k_0)[\delta(s) - \delta(t - s)]$, that is, we would capture at $s = 0$ and release at $s = t$ —then $\Omega_t = W_{k_0 \rightarrow k_1} + W_{k_1 \rightarrow k_0}$, and the Evans-Searles and Crooks FTs would again reduce to the same expression.

These experiments were simple, but important demonstrations of the FTs. They also demonstrate FTs derived under stochastic equations of motion, a topic not covered in this review. The motion of a single colloidal particle in a purely viscous solvent is accurately described by a stochastic Langevin equation; the Evans-Searles FT can be constructed from the stochastic, inertialess Langevin equation of motion, in much the same manner as was done for deterministic dynamics (27). This is important evidence against the notion that macroscopic irreversibility results from coarse graining or the separation of timescales of the system's degrees of freedom: For the optically trapped particles, both Langevin equations of motion as well as fictitious thermostatted deterministic equations (molecular dynamics) show a monotonic decrease in Ω_t with time, t . Conversely, if a system's dynamics are well described by stochastic Langevin dynamics, with uncorrelated Gaussian noise, then we are assured that the FTs hold. Indeed, several simple model experiments, including optically trapped colloidal particles in purely viscous media, can be described using stochastic Langevin dynamics, and consequently one could argue that these experiments confirm the Langevin dynamics rather than the FTs. However more recently, Carberry et al. (59) confirmed the FT using an optically trapped bead in a viscoelastic solvent, in which the stochastic equations of motion require a dissipative term with memory, from which the FTs have yet to be directly derived. This experimental confirmation is indeed a confirmation of the FTs, rather than a confirmation of the dynamics that satisfy the FTs.

4.2. Stretching of Biopolymers

In 2002 Liphardt et al. (56) measured the tension-induced, unfolding transition of a P5abs domain in a single RNA molecule. To do this, they tethered the ends of a single DNA-RNA hybrid molecule containing the P5abc domain to micron-sized colloidal beads whose surfaces were chemically functionalized. These beads act as handles to grab and manipulate the single molecule in an optical trap. One bead is weakly held in an optical trap while the other bead is held in a micropipette whose position/translational speed is controlled by a piezoelectric actuator. As the micropipette is translated relative to the optical trap, the chain is stretched and the stretching force is measured. In this way, Liphardt et al. constructed force-extension profiles of a single biomolecule, focusing specifically on a window of extensions over which the P5abs domain unfolded. Over a $\sim 30\text{-}\mu\text{m}$ extension range, they found that typically the force increased monotonically against the entropic elasticity of the chain, but at $\sim 10\text{ pN}$, the force was either constant or decreased slightly for a further $\sim 10\text{ }\mu\text{m}$ indicative of unfolding of the domain, before increasing monotonically again. The time-reverse path corresponds to retraction of the chain ends, creating a force-extension profile for the refolding of the domain. When the chain was unfolded slowly (corresponding to small $\dot{\lambda}$), the folding-unfolding process was reversible; in other words, the work done (force integrated over extension) of the forward (unfolding) and reverse (refolding) paths was roughly equal but of opposite sign. When the chain was unfolded quickly, a hysteresis loop appeared in the forward-reverse force profile, and the W to fold and unfold differed in magnitude owing to macroscopic irreversibility. Theoretically, the distributions $p_f(W)$ and $p_r(W)$ at any given protocol or stretching

rate, $\dot{\lambda}$, should obey the Crooks FT: However, variation in W owing to experimental error or approximations in analysis clearly must be minimal in comparison to the inherent variation in W that arises from the irreversibility of the process. Furthermore, it is important to sample a sufficient number of trajectories, N : As the protocol rate, $\dot{\lambda}$, increases and the process is driven further away from a quasistatic process, the required number N grows as rare trajectories in the distribution become important (**Figure 3**). Successive single-molecule manipulations in an optical trap can be difficult, limiting the maximum possible number of trajectories. However, Liphardt et al. (56) constructed $N = \mathcal{O}(300)$ stretching profiles, unfortunately an insufficient number to describe the distributions, but fortunately sufficient to show that ΔF obtained from Jarzynski's equality could be determined to within $k_B T$. In later work from the same group, Collin et al. (60) similarly used optical tweezers to construct experimental work distributions and demonstrate the Crooks FT for the folding/unfolding of an RNA hairpin and an RNA three-helix junction. These results demonstrate that nonequilibrium single-molecule measurements, when analyzed in conjunction with the FTs, can provide thermodynamic information, even though these single-molecule events may not be at equilibrium. More importantly, these papers introduced the proper analysis/interpretation of force measurements necessary in single-molecule force spectroscopy.

4.3. Other Model Systems

Several other more recent experimental demonstrations of the FTs have appeared in the literature in the past couple of years. Garnier & Ciliberto (61) demonstrated the FTs by measuring the fluctuating voltage of a resistor in parallel with a capacitor, driven out of equilibrium by a constant current flow; Schuler et al. (62) excited a single defect center in a diamond using an intensity-modulated laser, forming a two-state system; and Douarche et al. (63, 64) experimentally checked Jarzynski's equality and the Crooks FT against the thermal fluctuations of a mechanical oscillator in contact with a heat reservoir. Applications of FTs to model systems need not be experimental. To date FTs have been used in theoretical descriptions/computer simulations of sheared liquids, (23), chemical reactions (42, 65–70), molecular motors (71, 72), granular systems (73, 74), and glasses (51, 75).

Finally, through applications to time-dependent shear of viscoelastic fluids, FTs have been used to resolve a long-standing problem in linear irreversible thermodynamics (76). The reference of JVF_e to the entropy production is taken from linear irreversible thermodynamics, which asserts that, in local equilibrium, the entropy source strength is the sum of products of irreversible thermodynamic fluxes and forces (1, 2). By the second law, we then expect that $\langle J \rangle F_e$ is always negative. However for time-dependent processes in viscoelastic fluids, this is not always the case. In the steady state, the free energy F , the mean internal energy U , the temperature T , and the entropy S are all constant. In local equilibrium, at which the entropy transported to the reservoir is given by $\dot{S}_{tr} = \dot{Q}/T$, we have $\dot{F} = -JVF_e + \dot{Q} - T\dot{S} = 0$. This leads us directly to the conclusion that $JVF_e = -T\dot{S}_{sp}$, where $\dot{S}_{sp} = \dot{S} - \dot{S}_{tr}$ is the rate at which entropy is being spontaneously produced, owing to the external field

F_e . If we are not in the steady state or are not in local equilibrium, we cannot show that $JVF_e = -T\dot{S}_p$. Indeed for time-dependent processes in viscoelastic fluids, $\langle J \rangle F_e$ may be positive for periods of time. For a system initially in equilibrium, subject to a purely dissipative process (i.e., $\dot{\lambda} = 0 \forall t$), the time-averaged value of the entropy production, rather than the more traditional instantaneous entropy production, can be shown from Jarzynski's equality to form the inequality that specifies the time direction. This was demonstrated in simulations of time-dependent planar shear of viscoelastic fluids (76) in which the instantaneous value of $\langle J \rangle F_e$ was indeed negative for periods. Furthermore, this inequality is valid for time-dependent processes that may be arbitrarily far from equilibrium.

5. CONCLUSIONS

This brief review focuses on two new and fundamental relations in the field of nonequilibrium statistical mechanics: the Evans-Searles FT and Crooks FT. We provide here a unifying description of these FTs under deterministic dynamics and show that these provide a quantitative resolution of Loschmidt's paradox and new insight into the treatment of experimental trajectories of systems that are small, outside the thermodynamic limit, and evolve under nonequilibrium conditions. Recent experimental systems have demonstrated the use of these FTs, providing confirmation in a simple colloid system and showing potential practical use in single-molecule force spectroscopy. Among many challenges that remain, foremost is the application of the FTs in the development of nanotechnology and our understanding of biological systems.

SUMMARY POINTS

1. The Evans-Searles FT is a quantitative description of the emergence of thermodynamic irreversibility from time-reversible equations of motion. The theorem can be understood as a generalized second law that applies outside the thermodynamic limit, that is, where the system is small and observation time short.
2. From classical thermodynamics, the work done on a system over a quasi-static trajectory is equal to the equilibrium free-energy change. Crooks FT is an important extension of this to nonequilibrium thermodynamics. It relates the distribution of work, evaluated over ensembles of nonequilibrium trajectories, to an equilibrium free-energy change.
3. A simplified derivation of FTs starts from the system's equations of motion, which can include thermostat or heat exchange with a surrounding, and a definition of macroscopic reversibility.
4. Recent experiments that, for example, measure the work required to displace a colloidal bead or stretch a biopolymer have demonstrated applications of FTs to small systems under nonequilibrium conditions.

FUTURE ISSUES

1. Although recent interest has been focused on experimental demonstrations of FTs, FTs should be useful in gaining an improved understanding of molecular-scale processes, particularly in the biomolecular sciences and nanotechnology fields.
2. As small systems and their processes deviate further from equilibrium, the energy distributions described by FTs become broad and difficult to sample in experiment or simulation. New techniques, such as importance sampling in simulation or repetitive experimental sampling, will be required to fully elucidate these distributions for small systems driven far from equilibrium.

DISCLOSURE STATEMENT

The authors are not aware of any biases that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

The authors thank Prof. Denis J. Evans for his collaborative efforts over many years, as well as Dr. David M. Carberry for his assistance in generating illustrations. We also acknowledge financial support from the Australian Research Council (ARC).

LITERATURE CITED

1. de Groot SR, Mazur P. 1984. *Non-Equilibrium Thermodynamics*. New York: Dover
2. Kondepudi D, Prigogine I. 1998. *Modern Thermodynamics*. New York: Wiley
3. Evans DJ, Searles DJ. 1994. Equilibrium microstates which generate second law violating steady states. *Phys. Rev. E* 50:1645–48
4. Evans DJ, Searles DJ. 2002. The fluctuation theorem. *Adv. Phys.* 51:1529–85
5. Searles DJ, Evans DJ. 2004. Fluctuations relations for nonequilibrium systems. *Aust. J. Chem.* 57:1119–23
6. Crooks GE. 1998. Nonequilibrium measurements of free energy differences for microscopically reversible Markovian systems. *J. Stat. Phys.* 90:1481–87
7. Crooks GE. 1999. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Phys. Rev. E* 60:2721–26
8. Loschmidt J. 1876. Über den zustand des wärmeleichgewichtes eines systems von körpern mit rücksicht auf die schwerkraft. *J. Sitz. Kais. Akad. Wiss. Wien. Math. Nat. II. Abt.* 73:128
9. Cercignani C. 1998. *Ludwig Boltzmann: The Man Who Trusted Atoms*. Oxford: Oxford Univ. Press
10. Broda E. 1983. *Ludwig Boltzmann: Man, Physicist, Philosopher: L. Boltzmann, Rejoinder to the Heat Theoretical Considerations of Mr E. Zermelo (1896)*. Woodbridge: Ox Bow

11. Jarzynski C. 1997. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.* 78:2690–93
12. Jarzynski C. 1997. Equilibrium free-energy differences from nonequilibrium measurements: a master-equation approach. *Phys. Rev. E* 56:5018–35
13. Evans DJ, Morriss GP. 1990. *Statistical Mechanics of Nonequilibrium Liquids*. London: Academic
14. Hoover WG. 1985. Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A* 31:1695–97
15. Williams SR, Searles DJ, Evans DJ. 2004. Independence of the transient fluctuation theorem to thermostatting details. *Phys. Rev. E* 70:066113
16. Bright JN, Evans DJ, Searles DJ. 2005. New observations regarding deterministic, time-reversible thermostats and Gauss's principle of least constraint. *J. Chem. Phys.* 122:194106
17. Searles DJ, Evans DJ. 1999. Fluctuation theorem for stochastic systems. *Phys. Rev. E* 60:159–64
18. Jarzynski C. 2002. Targeted free energy perturbation. *Phys. Rev. E* 65:046122
19. Kofke DA. 2005. Free energy methods in molecular simulation. *Fluid Phase Equilib.* 228:41–48
20. Kofke DA. 2006. On the sampling requirements for exponential-work free-energy calculations. *Mol. Phys.* 104:3701–8
21. Kosztin I, Barz B, Janosi L. 2006. Calculating potentials of mean force and diffusion coefficients from nonequilibrium processes without Jarzynski's equality. *J. Chem. Phys.* 124:064106
22. Jarzynski C. 2006. Rare events and the convergence of exponentially averaged work values. *Phys. Rev. E* 73:046105
23. Evans DJ, Cohen EGD, Morriss GP. 1993. Probability of second law violations in shearing steady states. *Phys. Rev. Lett.* 71:2401–4
24. Evans DJ, Cohen EGD, Morriss GP. 1993. Errata: Probability of second law violations in shearing steady states. *Phys. Rev. Lett.* 71:3616
25. Gallavotti G, Cohen EGD. 1995. Dynamical ensembles in nonequilibrium statistical mechanics. *Phys. Rev. Lett.* 74:2694–97
26. Searles DJ, Rondoni L, Evans DJ. 2007. The steady state fluctuation relation for the dissipation function. *J. Stat. Phys.* 128:1337–63
27. Wang GM, Reid JC, Carberry DM, Williams DRM, Sevick EM, et al. 2005. Experimental study of the fluctuation theorem in a nonequilibrium steady state. *Phys. Rev. E* 71:046142
28. Gallavotti G, Cohen EGD. 1995. Dynamical ensembles in stationary states. *J. Stat. Phys.* 80:931–70
29. Gallavotti G. 1995. Reversible Anosov diffeomorphisms and large deviations. *Math. Phys. Electron. J.* 1:1
30. Bonetto F, Gallavotti G, Giuliani A, Zamponi F. 2006. Chaotic hypothesis, fluctuation theorem and singularities. *J. Stat. Phys.* 123:39–54
31. Young LS. 2002. What are SRB measures, and which dynamical systems have them? *J. Stat. Phys.* 108:733–54
32. Ruelle D. 2004. *Thermodynamic Formalism*. Cambridge: Cambridge Univ. Press

33. Gallavotti G. 1999. *Statistical Mechanics: A Short Treatise*. Berlin: Springer
34. Ayton G, Evans DJ, Searles DJ. 2001. A local fluctuation theorem. *J. Chem. Phys.* 115:2033–37
35. Evans DJ, Searles DJ. 1995. Steady states, invariant measures, and response theory. *Phys. Rev. E* 52:5839–48
36. Searles DJ, Evans DJ. 2000. The fluctuation theorem and Green-Kubo relations. *J. Chem. Phys.* 112:9727–35
37. Searles DJ, Evans DJ. 2001. Fluctuation theorem for heat flow. *Int. J. Thermophys.* 22:123–34
38. Bonetto F, Gallavotti G, Garrido PL. 1997. Chaotic principle: an experimental test. *Phys. D* 105:226–52
39. Bonetto F, Chernov NI, Lebowitz JL. 1998. (Global and local) fluctuations of phase space contraction in deterministic stationary nonequilibrium. *Chaos* 8:823–33
40. Bonetto F, Lebowitz JL. 2001. Thermodynamic entropy production fluctuation in a two-dimensional shear flow model. *Phys. Rev. E* 64:056129
41. Lepri S, Livi R, Politi A. 1997. Heat conduction in chains of nonlinear oscillators. *Phys. Rev. Lett.* 78:1896–99
42. Baranyai A. 2003. The kinetics of mixing and the fluctuation theorem in ideal mixtures of two-component model fluids. *J. Chem. Phys.* 119:2144–46
43. Schmick M, Markus M. 2004. Fluctuation theorem for a deterministic one-particle system. *Phys. Rev. E* 70:065101
44. Dolowschiak M, Kovacs Z. 2005. Fluctuation formula in the Nosé-Hoover thermostated Lorentz gas. *Phys. Rev. E* 71:025202
45. Zamponi F, Ruocco G, Angelani L. 2004. Fluctuations of entropy production in the isokinetic ensemble. *J. Stat. Phys.* 115:1655–68
46. Williams SR, Searles DJ, Evans DJ. 2006. Numerical study of the steady state fluctuation relations far from equilibrium. *J. Chem. Phys.* 124:194102
47. Evans DJ, Searles DJ, Rondoni L. 2005. Application of the Gallavotti-Cohen fluctuation relation to thermostated steady states near equilibrium. *Phys. Rev. E* 71:056120
48. Einstein A. 1905. Über die von der molekularkinetischen theorie der wärme geforderte bewegung von in ruhenden flüssigkeiten suspendierten teilchen. *Ann. Phys.* 17:549–60
49. Sutherland W. 1905. A dynamical theory of diffusion for nonelectrolytes and the molecular mass of albumin. *Philos. Mag.* 9:781–85
50. Einstein A. 1956. *Investigations on the Theory of Brownian Movement*. New York: Dover
51. Williams SR, Evans DJ. 2006. Linear response domain in glassy systems. *Phys. Rev. Lett.* 96:015701
52. Hess S, Evans DJ. 2001. Computation of the viscosity of a liquid from time averages of stress fluctuations. *Phys. Rev. E* 64:011207
53. Williams SR, McGlynn P, Bryant G, Snook IK, van Megen W. 2006. Dynamical signatures of freezing: stable fluids, metastable fluids, and crystals. *Phys. Rev. E* 74:031204

54. Howard J. 2001. *Mechanics of Motor Proteins and the Cytoskeleton*. Sunderland, MA: Sinauer Assoc.
55. Wang GM, Sevick EM, Mittag E, Searles DJ, Evans DJ. 2002. Experimental demonstration of violations of the second law of thermodynamics for small systems and short time scales. *Phys. Rev. Lett.* 89:050601
56. Liphardt J, Dumont S, Smith SB, Tinoco I, Bustamante C. 2002. Equilibrium information from nonequilibrium measurements in an experimental test of Jarzynski's equality. *Science* 296:1832–35
57. Carberry DM, Reid JC, Wang GM, Sevick EM, Searles DJ, et al. 2004. Fluctuations and irreversibility: an experimental demonstration of a second-law-like theorem using a colloidal particle held in an optical trap. *Phys. Rev. Lett.* 92:140601
58. Reid JC, Carberry DM, Wang GM, Sevick EM, Evans DJ, et al. 2004. Reversibility in nonequilibrium trajectories of an optically trapped particle. *Phys. Rev. E* 70:016111
59. Carberry DM, Baker MAB, Wang GM, Sevick EM, Evans DJ. 2007. An optical trap experiment to demonstrate fluctuation theorems in viscoelastic media. *J. Opt. A* 9:S204–14
60. Collin D, Ritort F, Jarzynski C, Smith SB, Tinoco I, et al. 2005. Verification of the Crooks fluctuation theorem and recovery of RNA folding free energies. *Nature* 437:231–34
61. Garnier N, Ciliberto S. 2005. Nonequilibrium fluctuations in a resistor. *Phys. Rev. E* 71:060101
62. Schuler S, Speck T, Tietz C, Wrachtrup J, Seifert U. 2005. Experimental test of the fluctuation theorem for a driven two-level system with time-dependent rates. *Phys. Rev. Lett.* 94:180602
63. Douarche F, Ciliberto S, Petrosyan A, Rabbiosi I. 2005. An experimental test of the Jarzynski equality in a mechanical experiment. *Europhys. Lett.* 70:593–99
64. Douarche F, Joubaud S, Garnier NB, Petrosyan A, Ciliberto S. 2006. Work fluctuation theorems for harmonic oscillators. *Phys. Rev. Lett.* 97:140603
65. Gaspard P. 2004. Fluctuation theorem for nonequilibrium reactions. *J. Chem. Phys.* 120:8898–905
66. Seifert U. 2004. Fluctuation theorem for birth-death or chemical master equations with time-dependent rates. *J. Phys. A* 37:L517–21
67. Andrieux D, Gaspard P. 2004. Fluctuation theorem and Onsager reciprocity relations. *J. Chem. Phys.* 121:6167–74
68. Andrieux D, Gaspard P. 2007. Fluctuation theorem for currents and Schnakenberg network theory. *J. Stat. Phys.* 127:107–31
69. Paramore S, Ayton GS, Voth GA. 2007. Extending the fluctuation theorem to describe reaction coordinates. *J. Chem. Phys.* 126:051102
70. Schmiedl T, Seifert U. 2007. Stochastic thermodynamics of chemical reaction networks. *J. Chem. Phys.* 126:044101
71. Seifert U. 2005. Fluctuation theorem for a single enzyme or molecular motor. *Europhys. Lett.* 70:36–41
72. Andrieux D, Gaspard P. 2006. Fluctuation theorems and the nonequilibrium thermodynamics of molecular motors. *Phys. Rev. E* 74:011906

73. Visco P, Puglisi A, Barrat A, Trizac E, van Wijland F. 2006. Fluctuations of power injection in randomly driven granular gases. *J. Stat. Phys.* 125:533–68
74. Puglisi A, Rondoni L, Vulpiani A. 2006. Relevance of initial and final conditions for the fluctuation relation in Markov processes. *J. Stat. Mech.* 2006:P08010
75. Williams SR, Evans DJ. 2007. Statistical mechanics of time independent nondissipative nonequilibrium states. *J. Chem. Phys.* 127:184101
76. Williams SR, Evans DJ, Mittag E. 2007. Negative entropy production in oscillatory processes. *C.R. Phys.* 8:620–24



Contents

A Fortunate Life in Physical Chemistry <i>Stuart A. Rice</i>	1
Chemistry and Photochemistry of Mineral Dust Aerosol <i>David M. Cwiertny, Mark A. Young, and Vicki H. Grassian</i>	27
Femtobiology <i>Villy Sundström</i>	53
Structures, Kinetics, Thermodynamics, and Biological Functions of RNA Hairpins <i>Philip C. Bevilacqua and Joshua M. Bloise</i>	79
Understanding Protein Evolution: From Protein Physics to Darwinian Selection <i>Konstantin B. Zeldovich and Eugene I. Shakhnovich</i>	105
Quasicrystal Surfaces <i>Patricia A. Thiel</i>	129
Molecular Ordering and Phase Behavior of Surfactants at Water-Oil Interfaces as Probed by X-Ray Surface Scattering <i>Mark L. Schlossman and Aleksey M. Tikhonov</i>	153
Extraordinary Transmission of Metal Films with Arrays of Subwavelength Holes <i>James V. Coe, Joseph M. Heer, Shannon Teeters-Kennedy, Hong Tian, and Kenneth R. Rodriguez</i>	179
The Ultrafast Dynamics of Photodetachment <i>Xiyi Chen and Stephen E. Bradforth</i>	203
Energy Flow in Proteins <i>David M. Leitner</i>	233
Advances in Correlated Electronic Structure Methods for Solids, Surfaces, and Nanostructures <i>Patrick Huang and Emily A. Carter</i>	261
Two-Dimensional Infrared Spectroscopy of Photoswitchable Peptides <i>Peter Hamm, Jan Helbing, and Jens Bredenbeck</i>	291

Wave-Packet Interferometry and Molecular State Reconstruction: Spectroscopic Adventures on the Left-Hand Side of the Schrödinger Equation <i>Jeffrey A. Cina</i>	319
Ions at Aqueous Interfaces: From Water Surface to Hydrated Proteins <i>Pavel Jungwirth and Bernd Winter</i>	343
Nanografting for Surface Physical Chemistry <i>Maozi Liu, Nabil A. Amro, and Gang-yu Liu</i>	367
Extending X-Ray Crystallography to Allow the Imaging of Noncrystalline Materials, Cells, and Single Protein Complexes <i>Jianwei Miao, Tetsuya Ishikawa, Qun Shen, and Thomas Earnest</i>	387
Patterning Fluid and Elastomeric Surfaces Using Short-Wavelength UV Radiation and Photogenerated Reactive Oxygen Species <i>Babak Sanii and Atul N. Parikh</i>	411
Equation-of-Motion Coupled-Cluster Methods for Open-Shell and Electronically Excited Species: The Hitchhiker's Guide to Fock Space <i>Anna I. Krylov</i>	433
Attosecond Electron Dynamics <i>Matthias F. Kling and Marc J. Vrakking</i>	463
Functional Polymer Brushes in Aqueous Media from Self-Assembled and Surface-Initiated Polymers <i>Ryan Toomey and Matthew Tirrell</i>	493
Electronic Spectroscopy of Carbon Chains <i>Evan B. Jochowitz and John P. Maier</i>	519
Multiscale Simulation of Soft Matter: From Scale Bridging to Adaptive Resolution <i>Matej Praprotnik, Luigi Delle Site, and Kurt Kremer</i>	545
Free Energies of Chemical Reactions in Solution and in Enzymes with Ab Initio Quantum Mechanics/Molecular Mechanics Methods <i>Hao Hu and Weitao Yang</i>	573
Fluctuation Theorems <i>E.M. Sevick, R. Prabhakar, Stephen R. Williams, and Debra J. Searles</i>	603
Structure, Dynamics, and Assembly of Filamentous Bacteriophages by Nuclear Magnetic Resonance Spectroscopy <i>Stanley J. Opella, Ana Carolina Zeri, and Sang Ho Park</i>	635
Inside a Collapsing Bubble: Sonoluminescence and the Conditions During Cavitation <i>Kenneth S. Suslick and David J. Flannigan</i>	659

Elastic Modeling of Biomembranes and Lipid Bilayers <i>Frank L.H. Brown</i>	685
Water in Nonpolar Confinement: From Nanotubes to Proteins and Beyond <i>Jayendran C. Rasaiah, Shekhar Garde, and Gerhard Hummer</i>	713
High-Resolution Spectroscopic Studies and Theory of Parity Violation in Chiral Molecules <i>Martin Quack, Jürgen Stohner, and Martin Willeke</i>	741
Collapse Mechanisms of Langmuir Monolayers <i>Ka Yee C. Lee</i>	771

Indexes

Cumulative Index of Contributing Authors, Volumes 55–59	793
Cumulative Index of Chapter Titles, Volumes 55–59	796

Errata

An online log of corrections to *Annual Review of Physical Chemistry* articles may be found at <http://physchem.annualreviews.org/errata.shtml>