Theoretical Foundations

1
Fluctuation Relations: A Pedagogical Overview

Richard Spinney and Ian Ford

1.1
Preliminaries

Ours is a harsh and unforgiving universe, and not just in the little matters that conspire against us. Its complicated rules of evolution seem unfairly biased against those who seek to predict the future. Of course, if the rules were simple, then there might be no universe of any complexity worth considering. Perhaps richness of behavior emerges only because each component of the universe interacts with many others and in ways that are very sensitive to details: this is the harsh and unforgiving nature. In order to predict the future, we have to take into account all the connections between the components, since they might be crucial to the evolution; furthermore, we need to know everything about the present in order to predict the future: both of these requirements are in most cases impossible. Estimates and guesses are not enough: unforgiving sensitivity to the details very soon leads to loss of predictability. We see this in the workings of a weather system. The approximations that meteorological services make in order to fill gaps in understanding, or initial data, eventually make the forecasts inaccurate.

So a description of the dynamics of a complex system is likely to be incomplete and we have to accept that predictions will be uncertain. If we are careful in the modeling of the system, the uncertainty will grow only slowly. If we are sloppy in our model building or initial data collection, it will grow quickly. We may expect the predictions of any incomplete model to tend toward a state of general ignorance, whereby we cannot be sure about anything: rain, snow, heat wave, or hurricane. We must expect there to be a spread, or fluctuations, in the outcomes of such a model.

This discussion of the growth of uncertainty in predictions has a bearing on another matter: the apparent *irreversibility* of all but the most simple physical processes. This refers to our inability to drive a system exactly backward by reversing the external forces that guide its evolution. Consider the mechanical work required to compress a gas by a piston in a cylinder. We might hope to see the expended energy returned when we stop pushing and allow the gas to drive the piston all the way back to the starting point: but not all will be returned. The system

Nonequilibrium Statistical Physics of Small Systems: Fluctuation Relations and Beyond, First Edition.
Edited by Rainer Klages, Wolfram Just, and Christopher Jarzynski.

seems to mislay some energy to the benefit of the wider environment. This is the familiar process of friction. The one-way dissipation of energy during mechanical processing is an example of the famous second law of thermodynamics. But the process is actually rather mysterious: What about the underlying reversibility of Newton's equations of motion? Why is the leakage of energy one way?

We may suspect that a failure to engineer the exact reversal of a compression is simply a consequence of a lack of control over all components of the gas and its environment: the difficulty in setting things up properly for the return leg implies the virtual impossibility of retracing the behavior. So we might not expect to be able to retrace exactly. But why do we not sometimes see "antifriction?" A clue might be seen in the relative size and complexity of the system and its environment. The smaller system is likely to evolve in a more complicated fashion as a result of the coupling, while we may expect the larger environment to be much less affected. There is a disparity in the effect of the coupling on each participant, and it is believed that this is responsible for the apparent one-way nature of friction. It is possible to implement these ideas by modeling the behavior of a system using uncertain or stochastic dynamics. The probability of observing a reversal of the behavior on the return leg can be calculated explicitly and it turns out that the difference between probabilities of observing a particular compression and seeing its reverse on the return leg leads to a measure of the irreversibility of natural processes. The second law is then a rather simple consequence of the dynamics. A similar asymmetric treatment of the effect on a system of coupling to a large environment is possible using deterministic and reversible nonlinear dynamics. In both cases, Loschmidt's paradox, the apparent breakage of time reversal symmetry for thermally constrained systems, is evaded, although for different reasons.

This chapter describes the so-called *fluctuation relations*, or *theorems* [1–5], that emerge from the analysis of a physical system interacting with its environment and that provide the structure that leads to the conclusion just outlined. They can quantify unexpected outcomes in terms of the expected. They apply on microscopic as well as macroscopic scales, and indeed their consequences are most apparent when applied to small systems. They can be derived on the basis of a rather natural measure of irreversibility, just alluded to, that offers an interpretation of the second law and the associated concept of entropy production. The dynamical rules that control the universe might seem harsh and unforgiving, but they can also be charitable and from them have emerged fluctuation relations that seem to provide a better understanding of entropy, uncertainty, and the limits of predictability.

This chapter is structured as follows. In order to provide a context for the fluctuation relations suitable for newcomers to the field, we begin with a brief summary of thermodynamic irreversibility and then describe how stochastic dynamics might be modeled. We use a framework based on stochastic rather than deterministic dynamics, since developing both themes here might not provide the most succinct pedagogical introduction. Nevertheless, we refer to the deterministic framework briefly later on to emphasize its equivalence. We discuss the identification of entropy production with the degree of departure from dynamical reversibility and then take a careful look at the developments that follow, which include the various fluctuation relations, and consider how the second law might not operate as we expect. We illustrate the fluctuation relations using simple analytical models as an aid to understanding. We conclude with some final remarks, but the broader implications are to be found elsewhere in this book, for which we hope this chapter will serve as a helpful background.

1.2
Entropy and the Second Law

Ignorance and uncertainty has never been an unusual state of affairs in human perception. In mechanics, Newton's laws of motion provided tools that seemed to dispel some of the haze: here were mathematical models that enabled the future to be foretold! They inspired attempts to predict future behavior in other fields, particularly in thermodynamics, the study of systems through which matter and energy can flow. The particular focus in the early days of the field was the heat engine, a device whereby fuel and the heat it can generate can be converted into mechanical work. Its operation was discovered to produce a quantity called entropy that could characterize the efficiency with which energy in the fuel could be converted into motion. Indeed, entropy seemed to be generated whenever heat or matter flowed. The second law of thermodynamics famously states that the total entropy of the evolving universe is always increasing. But this statement still attracts discussion, more than 150 years after its introduction. We do not debate the meaning of Newton's second law anymore, so why is the second law of thermodynamics so controversial?

Well, it is hard to understand how there can be a physical quantity that never decreases. Such a statement demands the breakage of the principle of time reversal symmetry, a difficulty referred to as Loschmidt's paradox. Newton's equations of motion do not specify a preferred direction in which time evolves. Time is a coordinate in a description of the universe and it is just a convention that real-world events take place while this coordinate increases. Given that we cannot actually run time backward, we can demonstrate this symmetry in the following way. A sequence of events that take place according to time reversal symmetric equations can be inverted by instantaneously reversing all the velocities of all the participating components and then proceeding forward in time once again, suitably reversing any external protocol of driving forces, if necessary. The point is that any evolution can be imagined in reverse, according to Newton. We therefore do not expect to observe any quantity ever-increasing with time. This is the essence of Loschmidt's objection to Boltzmann's [6] mechanical interpretation of the second law.

Nobody, however, has been able to initiate a heat engine such that it sucks exhaust gases back into its furnace and combines them into fuel. The denial of such a spectacle is empirical evidence for the operation of the second law, but it is also an expression of Loschmidt's paradox. Time reversal symmetry is broken by the apparent illegality of entropy-consuming processes and that seems unacceptable. Perhaps we should not blindly accept the second law in the sense that has traditionally been ascribed to it. Or perhaps there is something deeper going on. Furthermore, a law that only specifies the sign of a rate of change sounds rather incomplete.

But what has emerged in the past two decades or so is the realization that Newton's laws of motion, when supplemented by the acceptance of uncertainty in the way systems behave, brought about by roughly specified interactions with the environment, can lead quite naturally to a quantity that grows with time, that is, uncertainty itself. It is reasonable to presume that incomplete models of the evolution of a physical system will generate additional uncertainty in the reliability of the description of the system as they are evolved. If the velocities were all instantaneously reversed, in the hope that a previous sequence of events might be reversed, uncertainty would continue to grow within such a model. We shall, of course, need to quantify this vague notion of uncertainty. Newton's laws on their own are time reversal symmetric, but intuition suggests that the injection and evolution of configurational uncertainty would break the symmetry. Entropy production might therefore be equivalent to the leakage of our confidence in the predictions of an incomplete model: an interpretation that ties in with prevalent ideas of entropy as a measure of information.

Before we proceed further, we need to remind ourselves about the phenomenology of irreversible classical thermodynamic processes [7]. A system possesses energy E and can receive additional incremental contributions in the form of heat dQ from a heat bath at temperature T and work dW from an external mechanical device that might drag, squeeze, or stretch the system. It helps perhaps to view dQ and dW roughly as increments in kinetic and in potential energy, respectively. We write the first law of thermodynamics (energy conservation) in the form $dE = dQ + dW$. The second law is then traditionally given as Clausius' inequality:

$$\oint \frac{dQ}{T} \leq 0, \tag{1.1}$$

where the integration symbol means that the system is taken around a cycle of heat and work transfers, starting and ending in thermal equilibrium with the same macroscopic system parameters, such as temperature and volume. The temperature of the heat bath might change with time, though by definition and in recognition of its presumed large size it always remains in thermal equilibrium, and the volume and shape imposed upon the system during the process might also be time dependent. We can also write the second law for an incremental thermodynamic process as

$$dS_{\text{tot}} = dS + dS_{\text{med}}, \tag{1.2}$$

where each term is an incremental entropy change, the system again starting and ending in equilibrium. The change in system entropy is denoted dS and the change in entropy of the heat bath, or the surrounding medium, is defined as

$$dS_{\text{med}} = -\frac{dQ}{T}, \tag{1.3}$$

such that dS_{tot} is the total entropy change of the two combined (the "universe"). We see that Eq. (1.1) corresponds to the condition $\oint dS_{\text{tot}} \geq 0$, since $\oint dS = 0$. A more powerful reading of the second law is that

$$dS_{\text{tot}} \geq 0, \tag{1.4}$$

for any incremental segment of a thermodynamic process, as long as it starts and ends in equilibrium. An equivalent expression of the law would be to combine these statements to write $\mathrm{d}W - \mathrm{d}E + T\mathrm{d}S \geq 0$, from which we conclude that the *dissipative* work (sometimes called irreversible work) in an isothermal process,

$$\mathrm{d}W_{\mathrm{d}} = \mathrm{d}W - \mathrm{d}F, \tag{1.5}$$

is always positive, where $\mathrm{d}F$ is a change in Helmholtz free energy. We may also write $\mathrm{d}S = \mathrm{d}S_{\mathrm{tot}} - \mathrm{d}S_{\mathrm{med}}$ and regard $\mathrm{d}S_{\mathrm{tot}}$ as a contribution to the change in entropy of a system that is not associated with a flow of entropy from the heat bath, the $\mathrm{d}Q/T$ term. For a thermally isolated system, where $\mathrm{d}Q = 0$, we have $\mathrm{d}S = \mathrm{d}S_{\mathrm{tot}}$ and the second law then says that the system entropy increase is due to "internal" generation; hence, $\mathrm{d}S_{\mathrm{tot}}$ is sometimes [7] denoted $\mathrm{d}S_i$.

Boltzmann tried to explain what this ever-increasing quantity might represent at a microscopic level [6]. He considered a thermally isolated gas of particles interacting through pairwise collisions within a framework of classical mechanics. The quantity

$$H(t) = \int f(\mathbf{v}, t) \ln f(\mathbf{v}, t) \mathrm{d}\mathbf{v}, \tag{1.6}$$

where $f(\mathbf{v}, t)\mathrm{d}\mathbf{v}$ is the population of particles with a velocity in the range of $\mathrm{d}\mathbf{v}$ about $\mathbf{v}$, can be shown to decrease with time, or remain constant if the population is in a Maxwell–Boltzmann distribution characteristic of thermal equilibrium. Boltzmann obtained this result by assuming that the collision rate between particles at velocities $\mathbf{v}_1$ and $\mathbf{v}_2$ is proportional to the product of populations at those velocities, that is, $f(\mathbf{v}_1, t)f(\mathbf{v}_2, t)$. He proposed that H was proportional to the negative of system entropy and that his so-called H-theorem provides a sound microscopic and mechanical justification for the second law. Unfortunately, this does not hold up. As Loschmidt pointed out, Newton's laws of motion cannot lead to a quantity that always decreases with time: $\mathrm{d}H/\mathrm{d}t \leq 0$ would be incompatible with the principle of time reversal symmetry that underlies the dynamics. The H-theorem does have a meaning, but it is statistical: the decrease in H is an expected, but not guaranteed, result. Alternatively, it is a correct result for a dynamical system that does not adhere to time reversal symmetric equations of motion. The neglect of correlation between the velocities of colliding particles, both in the past and in the future, is where the model departs from Newtonian dynamics.

The same difficulty emerges in another form when, following Gibbs, it is proposed that the entropy of a system might be viewed as a property of an ensemble of many systems, each sampled from a probability density $P(\{\mathbf{x}, \mathbf{v}\})$, where $\{\mathbf{x}, \mathbf{v}\}$ denotes the positions and velocities of all the particles in a system. Gibbs wrote

$$S_{\mathrm{Gibbs}} = -k_{\mathrm{B}} \int P(\{\mathbf{x}, \mathbf{v}\}) \ln P(\{\mathbf{x}, \mathbf{v}\}) \prod \mathrm{d}\mathbf{x}\mathrm{d}\mathbf{v}, \tag{1.7}$$

where k_{B} is Boltzmann's constant and the integration is over all phase space. The Gibbs representation of entropy is compatible with classical equilibrium thermodynamics. But the probability density P for an isolated system should evolve in time according to Liouville's theorem, in such a way that S_{Gibbs} is a *constant* of the motion. How, then, can the entropy of an isolated system, such as the universe, increase? Either equation (1.7) is valid only for equilibrium situations, something has been left out, or too much has been assumed.

The resolution of this problem is that Gibbs' expression can represent thermodynamic entropy, but only if P is not taken to provide an exact representation of the state of the universe or, if you wish, of an ensemble of universes. At the very least, practicality requires us to separate the universe into a system about which we may know and care a great deal and an environment with which the system interacts, which is much less precisely monitored. This indeed is one of the central principles of thermodynamics. We are obliged by this incompleteness to represent the probability of environmental details in a so-called coarse-grained fashion, which has the effect that the probability density appearing in Gibbs' representation of the *system* entropy evolves not according to Liouville's equations, but according to versions with additional terms that represent the effect of an uncertain environment upon an open system. This then allows S_{Gibbs} to change, the detailed nature of which will depend on exactly how the environmental forces are represented.

For an isolated system however, an increase in S_{Gibbs} will emerge only if we are obliged to coarse-grained aspects of the system itself. This line of development could be considered rather unsatisfactory, since it makes the entropy of an isolated system grain-size dependent, and alternatives may be imagined where the entropy of an isolated system is represented by something other than S_{Gibbs}. The reader is directed to the literature [8] for further consideration of this matter. However, in this chapter, we shall concern ourselves largely with entropy generation brought about by systems in contact with coarse-grained environments described using stochastic forces, and within such a framework the Gibbs' representation of system entropy will suffice.

We shall discuss a stochastic representation of the additional terms in the system's dynamical equations in the next section, but it is important to note that a deterministic description of environmental effects is also possible, and it might perhaps be thought more natural. On the other hand, the development using stochastic environmental forces is in some ways easier to present. But it should be appreciated that some of the early work on fluctuation relations was developed using deterministic so-called thermostats [1, 9], and that this theme is represented briefly in Section 1.9, and elsewhere in this book.

1.3 Stochastic Dynamics

1.3.1 Master Equations

We pursue the assertion that sense can be made of the second law, its realm of applicability and its failings, when Newton's laws are supplemented by the explicit inclusion of a developing configurational uncertainty. The deterministic rules of evolution of a system need to be replaced by rules for the evolution of the *probability* that the system should take a particular configuration. We must first discuss what we mean by probability. Traditionally, it is the limiting frequency that an event might occur among a large number of trials. But there is also a view that probability represents a distillation, in numerical form, of the best judgment or belief about the state of a system: our information [10]. It is a tool for the evaluation of *expectation* values of system properties, representing what we expect to observe based on information about a system. Fortunately, the two interpretations lead to laws for the evolution of probability that are of similar form.

So let us derive equations that describe the evolution of probability for a simple case. Consider a random walk in one dimension, where a step of variable size is taken at regular time intervals [11–13]. We write the *master equation* describing such a *stochastic process*:

$$\mathcal{P}_{n+1}(x_m) = \sum_{m'=-\infty}^{\infty} T_n(x_m - x_{m'}|x_{m'})\mathcal{P}_n(x_{m'}), \tag{1.8}$$

where $\mathcal{P}_n(x_m)$ is the probability that the walker is at position x_m at timestep n, and $T_n(\Delta x|x)$ is the transition probability for making a step of size Δx in timestep n given a starting position of x. The transition probability may be considered to represent the effect of the environment on the walker. We presume that Newtonian forces cause the move to be made, but we do not know enough about the environment to model the event any better than this. We have assumed the Markov property such that the transition probability does not depend on the previous history of the walker, but only on the position x prior to making the step. It is normalized such that

$$\sum_{m=-\infty}^{\infty} T_n(x_m - x_{m'}|x_{m'}) = 1, \tag{1.9}$$

since the total probability that *any* transition is made, starting from $x_{m'}$, is unity. The probability that the walker is at position m at time n is a sum of probabilities of all possible previous histories that lead to this situation. In the Markov case, the master equation shows that these *path* probabilities are products of transition probabilities and the probability of an initial situation, a simple viewpoint that we shall exploit later.

1.3.2
Kramers–Moyal and Fokker–Planck Equations

The Kramers–Moyal and Fokker–Planck equations describe the evolution of *probability density functions*, denoted P, which are continuous in space (KM) and additionally in time (FP). We start with the Chapman–Kolmogorov equation, an integral form of the master equation for the evolution of a probability density function that is continuous in space:

$$P(x, t+\tau) = \int T(\Delta x|x - \Delta x, t)P(x - \Delta x, t)\mathrm{d}\Delta x. \tag{1.10}$$

We have swapped the discrete time label n for a parameter t. The quantity $T(\Delta x|x,t)$ describes a jump from x through distance Δx in a period τ starting from time t. Note that T now has dimensions of inverse length (it is really a Markovian transition probability *density*), and is normalized according to $\int T(\Delta x|x,t)\mathrm{d}\Delta x = 1$.

We can turn this integral equation into a differential equation by expanding the integrand in Δx to get

$$P(x,t+\tau) = P(x,t) + \int \mathrm{d}\Delta x \sum_{n=1}^{\infty} \frac{1}{n!}(-\Delta x)^n \frac{\partial^n (T(\Delta x|x,t)P(x,t))}{\partial x^n} \tag{1.11}$$

and define the Kramers–Moyal coefficients, proportional to moments of T,

$$M_n(x,t) = \frac{1}{\tau}\int \mathrm{d}\Delta x (\Delta x)^n T(\Delta x|x,t), \tag{1.12}$$

to obtain the (discrete time) Kramers–Moyal equation:

$$\frac{1}{\tau}(P(x,t+\tau) - P(x,t)) = \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \frac{\partial^n (M_n(x,t)P(x,t))}{\partial x^n}. \tag{1.13}$$

Sometimes the Kramers–Moyal equation is defined with a time derivative of P on the left-hand side instead of a difference.

Equation (1.13) is rather intractable, due to the infinite number of higher derivatives on the right-hand side. However, we might wish to confine attention to evolution in continuous time and consider only stochastic processes that are continuous in space in this limit. This excludes processes that involve discontinuous jumps: the allowed step lengths must go to zero as the timestep goes to zero. In this limit, every KM coefficient vanishes except the first and second, consistent with the Pawula theorem. Furthermore, the difference on the left-hand side of Eq. (1.13) becomes a time derivative and we end up with the Fokker–Planck equation (FPE):

$$\frac{\partial P(x,t)}{\partial t} = -\frac{\partial (M_1(x,t)P(x,t))}{\partial x} + \frac{1}{2}\frac{\partial^2 (M_2(x,t)P(x,t))}{\partial x^2}. \tag{1.14}$$

We can define a probability current,

$$J = M_1(x,t)P(x,t) - \frac{1}{2}\frac{\partial (M_2(x,t)P(x,t))}{\partial x}, \tag{1.15}$$

and view the FPE as a continuity equation for probability density:

$$\frac{\partial P(x,t)}{\partial t} = -\frac{\partial}{\partial x}\left(M_1(x,t)P(x,t) - \frac{1}{2}\frac{\partial (M_2(x,t)P(x,t))}{\partial x}\right) = -\frac{\partial J}{\partial x}. \tag{1.16}$$

The FPE reduces to the familiar diffusion equation if we take M_1 and M_2 to be zero and $2D$, respectively. Note that it is probability that is diffusing, not a physical property like gas concentration. For example, consider the limit of the symmetric Markov random walk in one dimension as timestep and spatial step go to zero: the so-called Wiener process. The probability density $P(x, t)$ evolves according to

$$\frac{\partial P(x,t)}{\partial t} = D\frac{\partial^2 P(x,t)}{\partial x^2}, \tag{1.17}$$

with an initial condition $P(x, 0) = \delta(x)$, if the walker starts at the origin. The statistical properties of the process are represented by the probability density that satisfies this equation, that is,

$$P(x,t) = \frac{1}{(4\pi Dt)^{1/2}} \exp\left(-\frac{x^2}{4Dt}\right), \tag{1.18}$$

representing the increase in positional uncertainty of the walker as time progresses.

1.3.3 Ornstein–Uhlenbeck Process

We now consider a very important stochastic process describing the evolution of the velocity of a particle v. We shall approach this from a different viewpoint: a treatment of the dynamics where Newton's equations are supplemented by environmental forces, some of which are stochastic. It is proposed that the environment introduces a linear damping term together with random noise:

$$\dot{v} = -\gamma v + b\xi(t), \tag{1.19}$$

where γ is the friction coefficient, b is a constant, and ξ has statistical properties $\langle \xi(t) \rangle = 0$, where the angle brackets represent an expectation over the probability distribution of the noise, and $\langle \xi(t)\xi(t') \rangle = \delta(t - t')$, which states that the so-called "white" noise is sampled from a distribution with no autocorrelation in time. The singular variance of the noise might seem to present a problem, but it can be accommodated. This is the Langevin equation. We can demonstrate that it is equivalent to a description based on a Fokker–Planck equation by evaluating the KM coefficients, considering Eq. (1.12) in the form

$$M_n(v,t) = \frac{1}{\tau}\int \mathrm{d}\Delta v (\Delta v)^n T(\Delta v|v,t) = \frac{1}{\tau}\langle (v(t+\tau) - v(t))^n \rangle, \tag{1.20}$$

and in the continuum limit where $\tau \to 0$. This requires an equivalence between the average of $(\Delta v)^n$ over a transition probability density T and the average over the statistics of the noise ξ. We integrate Eq. (1.19) for small τ to get

$$v(t+\tau) - v(t) = -\gamma \int_t^{t+\tau} v \mathrm{d}t + b\int_t^{t+\tau} \xi(t')\mathrm{d}t' \approx -\gamma v(t)\tau + b\int_t^{t+\tau} \xi(t')\mathrm{d}t', \tag{1.21}$$

and according to the properties of the noise, this gives $\langle \mathrm{d}v \rangle = -\gamma v \tau$ with $\mathrm{d}v = v(t + \tau) - v(t)$, such that $M_1(v) = \langle \dot{v} \rangle = -\gamma v$. We also construct $(v(t+\tau) - v(t))^2$ and using the appropriate statistical properties and the continuum limit, we get $\left\langle (\mathrm{d}v)^2 \right\rangle = b^2\tau$ and $M_2 = b^2$. We have therefore established that the FPE equivalent to the Langevin equation (Eq. (1.19)) is

$$\frac{\partial P(v,t)}{\partial t} = \frac{\partial(\gamma v P(v,t))}{\partial v} + \frac{b^2}{2}\frac{\partial^2 P(v,t)}{\partial v^2}. \tag{1.22}$$

The stationary solution to this equation ought to be the Maxwell–Boltzmann velocity distribution $P(v) \propto \exp\left(-mv^2/2k_\mathrm{B}T\right)$ of a particle of mass m in thermal equilibrium with an environment at temperature T, so b must be related to T and γ in the form $b^2 = 2k_\mathrm{B}T\gamma/m$, where k_B is Boltzmann's constant. This is a connection known as a fluctuation–dissipation relation: b characterizes the fluctuations and γ the dissipation or damping in the Langevin equation. Furthermore, it may be shown that the time-dependent solution to Eq. (1.22), with initial condition $\delta(v - v_0)$ at time t_0, is

$$P^T_{\mathrm{OU}}[v,t|v_0,t_0] = \sqrt{\frac{m}{2\pi k_\mathrm{B}T(1-\mathrm{e}^{-2\gamma(t-t_0)})}}\exp\left(-\frac{m\left(v - v_0\mathrm{e}^{-\gamma(t-t_0)}\right)^2}{2k_\mathrm{B}T(1-\mathrm{e}^{-2\gamma(t-t_0)})}\right). \tag{1.23}$$

This is a Gaussian with time-dependent mean and variance. The notation $P^T_{\mathrm{OU}}[\cdots]$ characterizes this as a transition probability density for the so-called Ornstein–Uhlenbeck process starting from initial value v_0 at initial time t_0, and ending at the final value v at time t.

The same mathematics can be used to describe the motion of a particle in a harmonic potential $\phi(x) = \kappa x^2/2$, in the limit where the frictional damping coefficient γ is very large. The Langevin equations that describe the dynamics are $\dot{v} = -\gamma v - \kappa x/m + b\xi(t)$ and $\dot{x} = v$, which reduce in this so-called overdamped limit to

$$\dot{x} = -\frac{\kappa}{m\gamma}x + \frac{b}{\gamma}\xi(t), \tag{1.24}$$

which then has the same form as Eq. (1.19), but for position instead of velocity. The transition probability (1.23), recast in terms of x, can therefore be employed.

In summary, the evolution of a system interacting with a coarse-grained environment can be modeled using a stochastic treatment that includes time-dependent random external forces. However, these really represent the effect of uncertainty in the *initial* conditions for the system and its environment: indefiniteness in some of these initial environmental conditions might only have an impact upon the system at a later time. For example, the uncertainty in the velocity of a particle in a gas increases as particles that were initially far away, and that were poorly specified at the initial time, have the opportunity to move closer and interact. The evolution equations are not time reversal symmetric since the principle of causality is assumed: the probability of a system configuration depends upon events that precede it in time, and not on events in the future. The evolving probability density can capture the growth in configurational uncertainty with time. We can now explore how growth of uncertainty in system configuration might be related to entropy production and the irreversibility of macroscopic processes.

1.4 Entropy Generation and Stochastic Irreversibility

1.4.1 Reversibility of a Stochastic Trajectory

The usual statement of the second law in thermodynamics is that it is impossible to observe the reverse of an entropy-producing process. Let us immediately reject this version of the law and recognize that nothing is impossible. A ball might roll off a table and land at our feet. But there is never stillness at the microscopic level and, without breaking any law of mechanics, the molecular motion of the air, ground, and ball might conspire to reverse their macroscopic motion, bringing the ball back to rest on the table. This is not ridiculous: it is an inevitable consequence of the time reversal symmetry of Newton's laws. All we need for this event to occur is to create the right initial conditions. Of course, that is where the problem lies: it is virtually impossible to engineer such a situation, but virtually impossible is not absolutely impossible.

This of course highlights the point behind Loschmidt's paradox. If we were to time reverse the equations of motion of every atom that was involved in the motion of the ball at the end of such an event, we *would* observe the reverse behavior. Or rather more suggestively, we would observe both the forward *and* the reverse behavior with probability 1. This of course is such an overwhelmingly difficult task that one would never entertain the idea of its realization. Indeed, it is also not how one typically considers irreversibility in the real world, whether that be in the lab or through experience. What one may in principle be able to investigate is the explicit time reversal of just the motion of the particle(s) of interest to see whether the previous history can be reversed. Instead of reversing the motion of all the atoms of the ground, the air, and so on, we just attempt to roll the ball back toward the table at the same speed at which it landed at our feet. In this scenario, we would certainly not expect the reverse behavior. Now because the reverse motion is not inevitable, we have somehow, for the system we are considering, identified (or perhaps constructed) the concept of irreversibility albeit on a somewhat anthropic level: events do not easily run backward.

How have we evaded Loschmidt's paradox here? We failed to provide the initial conditions that would ensure reversibility: we left out the reversal of the motion of all other atoms. If they act upon the system differently under time reversal, then irreversibility is (virtually) inevitable. This is not so very profound, but what we have highlighted here is one of the principal paradigms of thermodynamics, the separation of the system of interest and its environment, or for our example the ball and the rest of the surroundings. Given then that we expect such irreversible behavior when we ignore the details of the environment in this way, we can ask what representation of that environment might be most suitable when establishing a measure of the irreversibility of the process? The answer to which is when the environment explicitly interacts with the system in such a way that time reversal is irrelevant. While never strictly true, this can hold as a limiting case that can be represented in a model, allowing us to determine the extent to which the reversal of just the velocities of the system components can lead to a retracing of the previous sequence of events.

Stochastic dynamics can provide an example of such a model. In the appropriate limits, we may consider the collective influence of all the atoms in the environment to act on the system in the same inherently unpredictable and dissipative way regardless of whether their coordinates are time reversed or not. In the Langevin equation, this is achieved by ignoring a quite startling number of degrees of freedom associated with the environment, idealizing their behavior as noise along with a frictional force that slows the particle regardless of which way it is traveling. If we consider now the motion of our system of interest according to this Langevin scheme, its forward and reverse motion both are no longer certain and we can attribute a probability to each path under the influence of the environmental effects. How can we measure irreversibility given these dynamics? We ask the question, what is the probability of observing some forward process compared to the probability of seeing that forward process undone? Or perhaps, to what extent has the introduction of stochastic behavior violated Loschmidt's expectation? This section is largely devoted to the formulation of such a quantity.

Intuitively, we understand that we should be comparing the probability of observing some forward and reverse behavior, but these ideas need to be made concrete. Let us proceed in a manner that allows us to make a more direct connection between irreversibility and our consideration of Loschmidt's paradox. First, let us imagine a system that evolves under some suitable stochastic dynamics. We specifically consider a realization or trajectory that runs from time $t = 0$ to $t = \tau$. Throughout this process, we imagine that any number of system parameters may be subject to change. This could be, for example under suitable Langevin dynamics, the temperature of the heat bath, or perhaps the nature of a confining potential. The changes in the parameters alter the probabilistic behavior of the system as time evolves. Following the literature, we assume that any such change in these system parameters occurs according to some protocol $\lambda(t)$ that itself is a function of time. We note that a particular realization is not guaranteed to take place, since the system is stochastic; so, consequently, we associate with it a probability of occurring that is entirely dependent on the exact trajectory taken, for example, $x(t)$, and the protocol $\lambda(t)$.

We can readily compare probabilities associated with different paths and protocols. To quantify an irreversibility in the sense of the breaking of Loschmidt's expectation however, we must consider one specific path and protocol. Recall now our definition of the paradox. In a deterministic system, a time reversal of all the variables at the end of a process of length τ leads to the observation of the reverse behavior with probability 1 over the same period τ. It is the probability of the trajectory that corresponds to this reverse behavior within a stochastic system that we must address. To do so, let us consider what we mean by time reversal. A time reversal can be thought of as the operation of the time reversal operator $\hat{T}$ on the system variables and distribution. Specifically, for position x, momentum p, and some protocol λ, we have $\hat{T}x = x$, $\hat{T}p = -p$, and $\hat{T}\lambda = \lambda$. If we were to do this after time τ for a set of Hamilton's equations of motion in which the protocol was time independent, the trajectory would be the exact time-reversed retracing of the forward trajectory. We shall call this trajectory the *reversed trajectory* and is phenomenologically the "running backward" of the forward behavior. Similarly, if we were to consider a motion in a deterministic system that was subject to some protocol (controlling perhaps some external field), we would observe the reversed trajectory only if the original protocol were performed symmetrically backward. This running of the protocol backward we shall call the *reversed protocol*.

We now are in a position to construct a measure of irreversibility in a stochastic system. We do so by comparing the probability of observing the forward trajectory under the forward protocol with the probability of observing the reversed trajectory under the reversed protocol following a time reversal at the end of the forward process. We literally attempt to undo the forward process and measure how likely that is. Since the quantities we have just defined here are crucial to this chapter, we shall make their nature absolutely clear before we proceed. To reiterate, we wish to consider the following:

- **Reversed trajectory:** Given a trajectory $X(t)$ that runs from time $t = 0$ to $t = \tau$, we define the reversed trajectory $\bar{X}(t)$ that runs forward in time explicitly such that $\bar{X}(t) = \hat{T}X(\tau - t)$. Examples are for position $\bar{x}(t) = x(\tau - t)$ and for momentum $\bar{p}(t) = -p(\tau - t)$.
- **Reversed protocol:** The protocol $\lambda(t)$ behaves in the same way as the position variable x under time reversal and so we define the reversed protocol $\bar{\lambda}(t)$ such that $\bar{\lambda}(t) = \lambda(\tau - t)$.

Given these definitions, we can construct the path probabilities we seek to compare. For notational clarity, we label path probabilities that depend upon the forward protocol $\lambda(t)$ with the superscript F to denote the forward process and probabilities that depend upon the reversed protocol $\bar{\lambda}(t)$ with the superscript R to denote the reverse process. The probability of observing a given trajectory X, $\mathcal{P}^F[X]$, has two components. First, the probability of the path given its starting point $X(0)$, which we shall write as $\mathcal{P}^F[X(\tau)|X(0)]$; second, the initial probability of being at the start of the path, which we write as $\mathcal{P}_{\text{start}}(X(0))$ since it concerns the distribution of variables at the start of the forward process. The probability of observing the forward path is then given as

$$\mathcal{P}^F[X] = \mathcal{P}_{\text{start}}(X(0))\mathcal{P}^F[X(\tau)|X(0)]. \tag{1.25}$$

It is intuitive to proceed if we imagine the path probability as being approximated by a sequence of jumps that occur at distinct times. Since continuous stochastic behavior can be readily approximated by jump processes, but not the other way round, this simultaneously allows us to generalize any statements for a wider class of Markov processes. We shall assume for brevity that the jump processes occur in discrete time. By repeated application of the Markov property for such a system, we can write

$$\mathcal{P}^F[X] = \mathcal{P}_{\text{start}}(X_0)\mathcal{P}(X_1|X_0, \lambda(t_1)) \times \mathcal{P}(X_2|X_1, \lambda(t_2)) \times \cdots \times \mathcal{P}(X_n|X_{n-1}, \lambda(t_n)). \tag{1.26}$$

Here, we consider a trajectory that is approximated by the jump sequence between $n+1$ points $X_0, X_1, \ldots, X_n$ such that there are n distinct transitions that occur at discrete times $t_1, t_2, \ldots, t_n$, and where $X_0 = X(0)$ and $X_n = X(\tau)$. $\mathcal{P}(X_i|X_{i-1}, \lambda(t_i))$ is the probability of a jump from X_{i-1} to X_i using the value of the protocol evaluated at time t_i.

Continuing with our description of irreversibility, we construct the probability of the reversed trajectory under the reversed protocol. Approximating as a sequence of jumps as before, we may write

$$\begin{aligned}\mathcal{P}^R[\bar{X}] &= \hat{T}\mathcal{P}_{\text{end}}(\bar{X}(0))\mathcal{P}^R[\bar{X}(\tau)|\bar{X}(0)] \\ &= \mathcal{P}^R_{\text{start}}(\bar{X}(0))\mathcal{P}^R[\bar{X}(\tau)|\bar{X}(0)] \\ &= \mathcal{P}^R_{\text{start}}(\bar{X}_0)\mathcal{P}(\bar{X}_1|\bar{X}_0, \bar{\lambda}(t_1)) \times \cdots \times \mathcal{P}(\bar{X}_n|\bar{X}_{n-1}, \bar{\lambda}(t_n)).\end{aligned} \tag{1.27}$$

There are two key concepts here. First, in accordance with our definition of irreversibility, we attempt to "undo" the motion from the end of the forward process and so the initial distribution is formed from the distribution to which $\mathcal{P}_{\text{start}}$ evolves under $\lambda(t)$, such that for continuous probability density distributions we have

$$P_{\text{end}}(X(\tau)) = \int \mathrm{d}X P_{\text{start}}(X(0)) P^F[X(\tau)|X(0)], \tag{1.28}$$

so named because it is the probability distribution at the end of the forward process. For our discrete model, the equivalent is given by

$$\mathcal{P}_{\text{end}}(X_n) = \sum_{X_0} \cdots \sum_{X_{n-1}} \prod_{i=0}^{n-1} \mathcal{P}(X_{i+1}|X_i, \lambda(t_{i+1}))\mathcal{P}_{\text{start}}(X_0). \tag{1.29}$$

Second, to attempt to observe the reverse trajectory starting from $X(\tau)$, we must perform a time reversal of our system to take advantage of the reversibility in Hamilton's equations. However, when we time reverse the variable X, we are obliged to transform the distribution $\mathcal{P}_{\text{end}}$ as well, since the likelihood of starting the reverse trajectory with variable $\hat{T}X$ after we time reverse X is required to be the same as the likelihood of arriving at X before the time reversal. This transformed $\mathcal{P}_{\text{end}}$, $\hat{T}\mathcal{P}_{\text{end}}$, is the initial distribution for the reverse process and is thus labeled $\mathcal{P}^R_{\text{start}}$. Analogously, evolution under $\bar{\lambda}(t)$ takes the system distribution to $\mathcal{P}^R_{\text{end}}$. The forward process and its relation to the reverse process are illustrated for both coordinates x and v, which do and do not change sign following time reversal, respectively, in Figure 1.1, along with illustrations of the reversed trajectories and protocols.

Let us form our prototypical measure of the irreversibility of the path X, which for now we denote I:

$$I[X] = \ln\left[\frac{\mathcal{P}^F[X]}{\mathcal{P}^R[\bar{X}]}\right]. \tag{1.30}$$

There are some key points to note about such a quantity. First, since $\bar{X}$ and X are simply related, I is a functional of the trajectory X and accordingly will

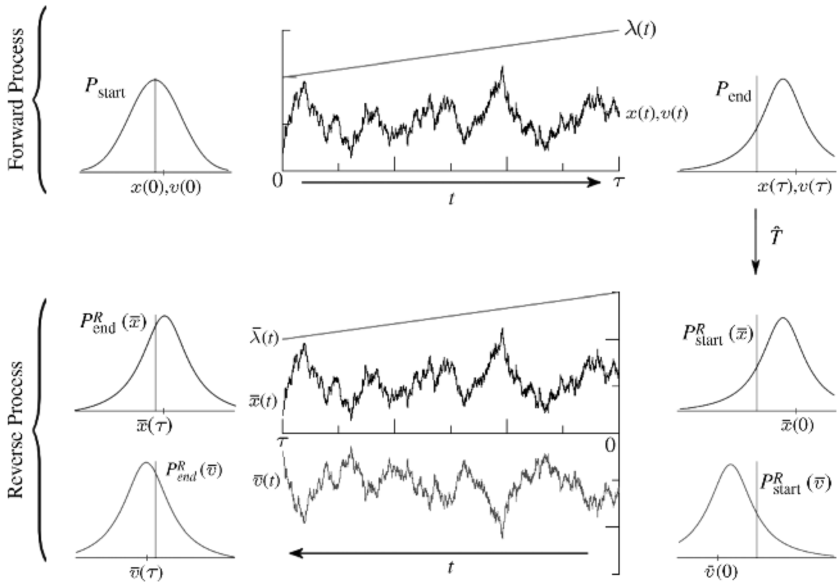

Figure 1.1 An illustration of the definition of the forward and reverse processes. The forward process consists of an initial probability *density* P_{start} that evolves forward in time under the forward protocol $\lambda(t)$ over a period τ, at the end of which the variable is distributed according to P_{end}. The reverse process consists of evolution from the distribution P^R_{start}, which is related to P_{end} by a time reversal, under the reversed protocol $\bar{\lambda}(t)$ over the same period τ, at the end of which the system will be distributed according to some final distribution P^R_{end}, which in general is not related to P_{start} and does not explicitly feature in assessment of the irreversibility of the forward process. A particular realization of the forward process is characterized by the forward trajectory $X(t)$, illustrated here as being $x(t)$ or $v(t)$. To determine the irreversibility of this realization, the reversed trajectories, $\bar{x}(t)$ or $\bar{v}(t)$, related by a time reversal, need to be considered as realizations in the reverse process.

take a range of values over all possible "realizations" of the dynamics: as such it will be characterized by a probability distribution. Furthermore, there is nothing in its form that disallows negative values. Finally, the quantity vanishes if the reversed trajectory occurs with the same probability as the forward trajectory under the relevant protocols: a process is deemed reversible if the forward process can be "undone" with equal probability. We can simplify this form since we know how the time-reversed protocols and trajectories are related. Given the step sequence laid out for the approximation to a continuous trajectory, we can transform X and t according to $\bar{X}_i = \hat{T}X_{n-i}$ and $\bar{\lambda}(t_i) = \lambda(t_{n-i+1})$ giving

$$\begin{aligned}\mathcal{P}^R[\bar{X}] &= \mathcal{P}^R_{start}(\hat{T}X(\tau))\mathcal{P}^R[\hat{T}X(0)|\hat{T}X(\tau)] \\ &= \mathcal{P}^R_{start}(\hat{T}X_n)\mathcal{P}(\hat{T}X_{n-1}|\hat{T}X_n, \lambda(t_n)) \times \cdots \times \mathcal{P}(\hat{T}X_0|\hat{T}X_1, \lambda(t_1)).\end{aligned} \tag{1.31}$$

Pointing out that $\mathcal{P}^R_{\text{start}}(\hat{T}X_n) = \hat{T}\mathcal{P}_{\text{end}}(\hat{T}X_n) = \mathcal{P}_{\text{end}}(X_n)$, we thus have

$$\begin{aligned}\ln\left[\frac{\mathcal{P}^F[X]}{\mathcal{P}^R[\bar{X}]}\right] &= \ln\left(\frac{\mathcal{P}_{\text{start}}(X(0))}{\mathcal{P}_{\text{end}}(X(\tau))}\right) + \ln\left[\frac{\mathcal{P}^F[X(\tau)|X(0)]}{\mathcal{P}^R[\hat{T}X(0)|\hat{T}X(\tau)]}\right] \\ &= \ln\left[\frac{\mathcal{P}_{\text{start}}(X_0)}{\mathcal{P}_{\text{end}}(X_n)}\prod_{i=1}^{n}\frac{\mathcal{P}(X_i|X_{i-1},\lambda(t_i))}{\mathcal{P}(\hat{T}X_{i-1}|\hat{T}X_i,\lambda(t_i))}\right],\end{aligned} \tag{1.32}$$

noting that strictly this is for a model in discrete space and time. Let us study this quantity for a specific model to understand its meaning in physical terms. Consider the continuous stochastic process described by the Langevin equation from Section 1.3.3, where $X = v$ and we have

$$\dot{v} = -\gamma v + \left(\frac{2k_B T(t)\gamma}{m}\right)^{1/2}\xi(t), \tag{1.33}$$

where $\xi(t)$ is white noise. The equivalent Fokker–Planck equation is given by

$$\frac{\partial P(v,t)}{\partial t} = \frac{\partial(\gamma v P(v,t))}{\partial v} + \frac{k_B T(t)\gamma}{m}\frac{\partial^2 P(v,t)}{\partial v^2}. \tag{1.34}$$

where P is a probability density. By inserting probability densities and associated infinitesimal volumes into Eq. (1.32) and canceling the latter, we observe that we may use probability densities to represent the quantity $I[X]$ for this continuous behavior without a loss of generality. To introduce a distinct forward and reverse process, let us allow the temperature to vary with a protocol $\lambda(t)$. We choose for simplicity a protocol that consists only of step changes such that

$$T(\lambda(t_i)) = T_j, \quad t_i \in [(j-1)\Delta t, j\Delta t], \tag{1.35}$$

where j is an integer in the range of $1 \le j \le N$, such that $N\Delta t = \tau$. Because the process is simply the combination of different Ornstein–Uhlenbeck processes, each of which is characterized by defined solution Eq. (1.23), we can represent the path probability in a piecewise fashion. Consolidating with our notation, the continuous Langevin behavior at some fixed temperature can be considered to be the limit, $\mathrm{d}t = (t_{i+1} - t_i) \to 0$, of the discrete jump process, so that

$$\begin{aligned}&\lim_{\mathrm{d}t\to 0}\prod_{t_i=(j-1)\Delta t}^{t_i=j\Delta t}\mathcal{P}(v_i|v_{i-1},\lambda(t_i)) = P^{T_j}_{\text{OU}}[v(j\Delta t)|v((j-1)\Delta t)]\mathrm{d}v(j\Delta t) \\ &= \left(\frac{m}{2\pi k_B T_j(1-\mathrm{e}^{-2\gamma\Delta t})}\right)^{1/2}\exp\left(-\frac{m(v(j\Delta t) - v((j-1)\Delta t)\mathrm{e}^{-\gamma\Delta t})^2}{2k_B T_j(1-\mathrm{e}^{-2\gamma\Delta t})}\right)\mathrm{d}v(j\Delta t).\end{aligned} \tag{1.36}$$

The total conditional path probability density (with units equal to the inverse dimensionality of the path) over N of these step changes in temperature is then by application of the Markov property

$$P^F[v(\tau)|v(0)] = \prod_{j=1}^{N}\left(\frac{m}{2\pi k_B T_j(1-\mathrm{e}^{-2\gamma\Delta t})}\right)^{1/2}\exp\left(-\frac{m(v(j\Delta t) - v((j-1)\Delta t)\mathrm{e}^{-\gamma\Delta t})^2}{2k_B T_j(1-\mathrm{e}^{-2\gamma\Delta t})}\right) \tag{1.37}$$

and since $\hat{T}v = -v$,

$$P^R[-v(0)| - v(\tau)] = \prod_{j=1}^{N} \left(\frac{m}{2\pi k_B T_j (1 - e^{-2\gamma\Delta t})} \right)^{1/2} \exp\left(-\frac{m(-v((j-1)\Delta t) + v(j\Delta t) e^{-\gamma\Delta t})^2}{2k_B T_j (1 - e^{-2\gamma\Delta t})} \right). \tag{1.38}$$

Taking the logarithm of their ratio explicitly and abbreviating $v(j\Delta t) = v_j$ yields

$$\ln\left[\frac{P^F[v(\tau)|v(0)]}{P^R[-v(0)| - v(\tau)]} \right] = -\frac{1}{k_B} \sum_{j=1}^{N} \frac{m}{2T_j} \left(v_j^2 - v_{j-1}^2 \right), \tag{1.39}$$

which is quite manifestly equal to the sum of negative changes of the kinetic energy of the particle scaled by k_B and the environmental temperature to which the particle is exposed. Our model consists only of the particle and the environment and so each negative kinetic energy change of the particle, $-\Delta Q$, must be associated with a positive flow of heat ΔQ_{med} into the environment such that we define $\Delta Q_{med} = -\Delta Q$. For the Langevin equation, the effect of the environment is idealized as a dissipative friction term and a fluctuating white noise characterized by a defined temperature that is entirely independent of the behavior of the particle. This is the idealization of a large equilibrium heat bath for which the exchanged heat is directly related to the entropy change of the bath through the relation $\Delta Q_{med} = T\Delta S$. It may be argued that changing between N temperatures under such an idealization is equivalent to exposing the particle to N separate equilibrium baths each experiencing an entropy change according to $\Delta Q_{med,j} = T_j \Delta S_j$. We consequently assert, for this particular model at least, that

$$k_B \, ln\left[\frac{P^F[v(\tau)|v(0)]}{P^R[-v(0)| - v(\tau)]} \right] = \sum_j \frac{\Delta Q_{med,j}}{T_j} = \sum_j \Delta S_j = \Delta S_{med}, \tag{1.40}$$

where the entropy production in all N baths can be denoted as a total entropy production ΔS_{med} that occurs in a generalized medium.

Let us now examine the remaining part of our quantification of irreversibility, which here is given in Eq. (1.32) by the logarithm of the ratio of $P_{start}(v(0))$ and $P_{end}(v(\tau))$. Given an arbitrary initial distribution, one can write this as the change in the logarithm of the dynamical solution to P as given by the Fokker–Planck equation (Eq. (1.34)). Consequently, we can write

$$\ln\left(\frac{P_{start}(v(0))}{P_{end}(v(\tau))} \right) = \ln \frac{P(v, 0)}{P(v, \tau)} = -(\ln P(v, \tau) - \ln P(v, 0)). \tag{1.41}$$

If we now characterize the *mean* entropy of our Langevin particle or "system" using a Gibbs entropy that we allow to be time dependent such that

$$\langle S_{sys} \rangle = S_{Gibbs} = -k_B \int dv \, P(v, t) \ln P(v, t), \tag{1.42}$$

one can make the conceptual leap that it is an individual value for the entropy of the system for a given v and time t that is being averaged in the above integral[1)] [14], $S_{\rm sys} = -k_{\rm B} \ln P(v,t)$. If we accept these assertions, we find that our measure of irreversibility for any one individual trajectory is formed as

$$k_{\rm B} I[X] = \Delta S_{\rm sys} + \Delta S_{\rm med}. \tag{1.43}$$

Since our model consists only of the Langevin particle (the system) and a heat bath (the medium), we therefore regard this sum as the total entropy production associated with such a trajectory and make the assertion that our measure of irreversibility is identically the increase in the total entropy of the universe, in this model at least:

$$\Delta S_{\rm tot}[X] = \Delta S_{\rm sys} + \Delta S_{\rm med} = k_{\rm B} \ln \left[\frac{P^F[X]}{P^R[\bar{X}]}\right]. \tag{1.44}$$

However, we have already stated that nothing prevents this quantity from taking negative values. If this is to be the total entropy production, how is this permitted given our knowledge of the second law of thermodynamics? In essence, describing the way in which a quantity that looks like the total entropy production can take both positive and negative values, but obeys well-defined statistical requirements such that, for example, it is compatible with the second law, is the subject matter of the so-called fluctuation theorems or fluctuation relations. These relations are disarmingly simple, but allow us to make predictions far beyond those possible in classical thermodynamics. For this class of system in fact, they are so simple that we can derive in a couple of lines a most fundamental relation and immediately reconcile the second law in terms of our irreversibility functional. Let us consider the average, with respect to all possible forward realizations, of the quantity $\exp\left(-\Delta S_{\rm tot}[X]/k_{\rm B}\right)$, which we write $\langle \exp\left(-\Delta S_{\rm tot}[X]/k_{\rm B}\right)\rangle$ and where the angle brackets denote a weighted path integration. Performing the average yields,

$$\begin{aligned} \langle e^{-\Delta S_{\rm tot}[X]/k_{\rm B}} \rangle &= \int dX P^F[X] e^{-\Delta S_{\rm tot}[X]/k_{\rm B}} \\ &= \int dX P^F[X] \frac{P^R[\bar{X}]}{P^F[X]} \\ &= \int d\bar{X} P^R[\bar{X}], \end{aligned} \tag{1.45}$$

where we assume the path integral measures are equivalent, $dX = d\bar{X}$, such that the Jacobian associated with the transformation between the paths is unity (this is guaranteed for any involutive transformation). Or perhaps more transparently, in the discrete approximation, multiple summations over $X_0, \ldots, X_n$ yield the same

1) Strictly, $P(v,t)$ is a probability density and so for Eq. (1.42) to be consistent with the entropy arising from the combinatorial arguments of statistical mechanics and dimensionally correct, it may be argued that we should be considering $\ln\left(P(v,t)dv\right)$. However, for relative changes, this issue is irrelevant.

result as summation over $X_n, \ldots, X_0$. The expression above now trivially integrates to unity that allows us to write the so-called [14]:

Integral Fluctuation Theorem

$$\langle e^{-\Delta S_{\mathrm{tot}}[X]/k_B} \rangle = 1. \tag{1.46}$$

This remarkably simple relation holds for all times, protocols, and initial conditions[2)] and implies that the possibility of negative total entropy change is obligatory. Furthermore, if we make use of Jensen's inequality

$$\langle \exp (z) \rangle \geq \exp \langle z \rangle, \tag{1.47}$$

we can directly infer

$$\langle \Delta S_{\mathrm{tot}} \rangle \geq 0. \tag{1.48}$$

Since this holds for any initial condition, we may also state that the mean total entropy monotonically increases for any process. This statement, under the stochastic dynamics we consider, is the second law. It is a replacement or reinterpretation of Eq. (1.4). The expected entropy production rate is always positive, but this is not necessarily found in detail for individual realizations. The second law, when correctly understood, is statistical in nature and we have now obtained an expression that places a fundamental bound on these statistics.

1.5 Entropy Production in the Overdamped Limit

We have formulated a quantity that we assert to be the total entropy production, though it is for a very specific system and importantly has no ability to describe the application of work. To broaden the scope of application, it is instructive to obtain a general expression like that obtained in Eq. (1.39), but for a class of stochastic behavior where we can formulate and verify the total entropy production without the need for an exact analytical result. This is straightforward for systems with detailed balance [15]; however, we can generalize further. The class of stochastic behavior we shall consider will be the simple overdamped Langevin equation that we discussed in Section 1.3.3 involving a position variable described by

$$\dot{x} = \frac{\mathcal{F}(x)}{m\gamma} + \left(\frac{2k_B T}{m\gamma}\right)^{1/2} \xi(t), \tag{1.49}$$

2) We do though assume that nowhere in the initial available configuration space we have $P_{\mathrm{start}}(X) = 0$. This is a paraphrasing of the so-called ergodic consistency requirement found in deterministic systems [9] and insists that there must be a trajectory for every possible reversed trajectory and vice versa, so that all possible paths, $\bar{X}(t)$, are included in the integral in the final line of Eq. (1.45).

along with an equivalent Fokker–Planck equation:

$$\frac{\partial P(x,t)}{\partial t} = -\frac{1}{m\gamma}\frac{\partial(\mathcal{F}(x)P(x,t))}{\partial x} + \frac{k_B T}{m\gamma}\frac{\partial^2 P(x,t)}{\partial x^2}. \tag{1.50}$$

The description includes a force term $\mathcal{F}(x)$ that allows us to model most simple thermodynamic processes including the application of work. We describe the force as a sum of two contributions that arise respectively from a potential $\phi(x)$ and an external force $f(x)$ that is applied directly to the particle, both of which we allow to vary in time through application of a protocol such that

$$\mathcal{F}(x,\lambda_0(t),\lambda_1(t)) = -\frac{\partial\phi(x,\lambda_0(t))}{\partial x} + f(x,\lambda_1(t)). \tag{1.51}$$

The first step in characterizing the entropy produced in the medium according to this description is to identify the main thermodynamic quantities, including the heat exchanged with the bath. To do this, we paraphrase Sekimoto and Seifert [14, 16, 17] and start from basic thermodynamics and the first law:

$$\Delta E = \Delta Q + \Delta W, \tag{1.52}$$

which must hold rigorously despite the stochastic nature of our model. To proceed, let us consider the change in each of these quantities in response to evolving our system by a small time $\mathrm{d}t$ and corresponding displacement $\mathrm{d}x$. We can readily identify that the system energy for overdamped conditions is equal to the value of the conservative potential such that

$$\mathrm{d}E = \mathrm{d}Q + \mathrm{d}W = \mathrm{d}(\phi(x,\lambda_0(t))). \tag{1.53}$$

However, at this point, we reach a subtlety in the mathematics originating in the stochastic nature of x. Where normally we could describe the small change in ϕ using the usual chain rule of calculus, when ϕ is a function of the stochastic variable x, we must be more careful. The peculiarity is manifest in an ambiguity of expressing the multiplication of a continuous stochastic function by a stochastic increment. The product, which strictly should be regarded as a stochastic integral, is not uniquely defined because both function and increment cannot be assumed to behave smoothly on any timescale. The mathematical details [12] are not of our concern for this chapter and so we shall not rigorously discuss stochastic calculus or go beyond the following steps of reasoning and assumption. First, we assume that in order to work with thermodynamic quantities in the traditional sense, as in undergraduate physics, we require a small change to resemble that of normal calculus, and this requires, in all instances, multiplication to follow so-called Stratonovich rules. These rules, denoted in this chapter by the symbol $\circ$, are taken to mean evaluation of the preceding stochastic function at the midpoint of the following increment. Employing this procedure, we may write

$$\begin{aligned}\mathrm{d}E &= \mathrm{d}(\phi(x(t),\lambda_0(t))) \\ &= \frac{\partial\phi(x(t),\lambda_0(t))}{\partial\lambda_0}\frac{\mathrm{d}\lambda_0(t)}{\mathrm{d}t}\mathrm{d}t + \frac{\partial\phi(x(t),\lambda_0(t))}{\partial x}\circ\mathrm{d}x.\end{aligned} \tag{1.54}$$

Next, we can explicitly write down the work from basic mechanics as contributions from the change in potential and the operation of an external force:

$$dW = \frac{\partial\phi(x(t),\lambda_0(t))}{\partial\lambda_0}\frac{d\lambda_0(t)}{dt}dt + f(x(t),\lambda_1(t)) \circ dx. \tag{1.55}$$

Accordingly, we directly have an expression for the heat transfer to the system in response to a small change dx:

$$\begin{aligned} dQ &= \frac{\partial\phi(x(t),\lambda_0(t))}{\partial x} \circ dx - f(x(t),\lambda_1(t)) \circ dx \\ &= -\mathcal{F}(x(t),\lambda_0(t),\lambda_1(t)) \circ dx. \end{aligned} \tag{1.56}$$

We may then integrate these small increments over a trajectory of duration τ to find

$$\Delta E = \int_0^\tau dE = \int_0^\tau d(\phi(x(t),\lambda_0(t))) = \phi(x(\tau),\lambda_0(\tau)) - \phi(x(0),\lambda_0(0)) = \Delta\phi, \tag{1.57}$$

$$\Delta W = \int_0^\tau dW = \int_0^\tau \frac{\partial\phi(x(t),\lambda_0(t))}{\partial\lambda_0}\frac{d\lambda_0(t)}{dt}dt + \int_0^\tau f(x(t),\lambda_1(t)) \circ dx, \tag{1.58}$$

and

$$\Delta Q = \int_0^\tau dQ = \int_0^\tau \frac{\partial\phi(x(t),\lambda_0(t))}{\partial x} \circ dx - \int_0^\tau f(x(t),\lambda_1(t)) \circ dx. \tag{1.59}$$

Let us now verify what we expect; the ratio of conditional path probability densities that we use in Eq. (1.32) will be equal to the negative heat transferred to the system divided by the temperature of the environment. We no longer have a means for representing the transition probabilities in general and so we proceed using the so-called "short time propagator" [11, 12, 18], which to first order in the time between transitions, dt, describes the probability of making a transition from x_i to x_{i+1}. We may then consider the analysis valid in the limit d$t \to 0$. The short time propagator can also be thought of as a short time Green's function; it is a solution to the Fokker–Planck equation subject to a delta function initial condition, valid as the propagation time is taken to zero.

The basic form of the short time propagator is helpfully rather intuitive and most simply adopts a general Gaussian form that reflects the fluctuating component of the force about the mean due to the Gaussian white noise. Abbreviating $\mathcal{F}(x,\lambda_0(t),\lambda_1(t))$ as $\mathcal{F}(x,t)$, we may write the propagator as

$$P[x_{i+1}, t_i + dt | x_i, t_i] = \sqrt{\frac{m\gamma}{4\pi k_B T dt}} \exp\left[-\frac{m\gamma}{4k_B T dt}\left(x_{i+1} - x_i - \frac{\mathcal{F}(x_i,t_i)}{m\gamma}dt\right)^2\right]. \tag{1.60}$$

However, one must be very careful. For reasons similar to those discussed above, a propagator of this type is not uniquely defined, with a family of forms being available depending on the spatial position at which one chooses to evaluate the force $\mathcal{F}$ of which Eq. (1.60) is but one example [18]. In the same way we had to choose certain multiplication rules; it is not enough to write $\mathcal{F}(x(t),t)$

without further comment since $x(t)$ has not been fully specified. This leaves a certain mathematical freedom in how to write the propagator and we must consider which is most appropriate. Of crucial importance is that all are correct in the limit $dt \to 0$ (all lead to the correct solution of the Fokker–Planck equation), meaning our choice must rest solely on ensuring the correct representation of the entropy production. We can proceed heuristically: as we take time $dt \to 0$, we steadily approach a representation of transitions as jump processes, from which we can proceed with confidence since jump processes are the more general description of stochastic phenomena. In this limit, therefore, we are obliged to faithfully represent the ratio that appears in Eq. (1.32). In this description, the forward and reverse jump probabilities have the same functional form and to emulate this we must evaluate the short time propagators at the same position x for both the forward and reverse transitions.[3] Mathematically, the most convenient way of doing this is to evaluate all functions in the propagator midway between initial and final points. Evaluating the functions at the midpoint x' such that $2x' = x_{i+1} + x_i$ and $dx = x_{i+1} - x_i$ introduces a propagator of the form

$$P[x_{i+1}, t_i + dt | x_i, t_i] = \sqrt{\frac{m\gamma}{4\pi k_B T dt}} \exp\left[-\frac{m\gamma}{4k_B T dt}\left(dx - \frac{\mathcal{F}(x', t_i)}{m\gamma} dt\right)^2 - \frac{1}{2}\frac{\partial}{\partial x'}\left(\frac{\mathcal{F}(x', t_i)}{m\gamma}\right) dt\right] \tag{1.61}$$

and similarly

$$P[x_i, t_i + dt | x_{i+1}, t_i] = \sqrt{\frac{m\gamma}{4\pi k_B T dt}} \exp\left[-\frac{m\gamma}{4k_B T dt}\left(-dx - \frac{\mathcal{F}(x', t_i)}{m\gamma} dt\right)^2 - \frac{1}{2}\frac{\partial}{\partial x'}\left(\frac{\mathcal{F}(x', t_i)}{m\gamma}\right) dt\right]. \tag{1.62}$$

The logarithm of their ratio, in the limit $dt \to 0$, simply reduces to

$$\begin{aligned}
\lim_{dt\to 0} \ln\left[\frac{P[x_{i+1}, t_i + dt | x_i, t_i]}{P[x_i, t_i + dt | x_{i+1}, t_i]}\right] &= \ln\left(\frac{P(x_{i+1} | x_i, \lambda(t_i))}{P(x_i | x_{i+1}, \lambda(t_i))}\right) \\
&= \frac{\mathcal{F}(x', \lambda_0(t_i), \lambda_1(t_i))}{k_B T} dx \\
&= \frac{\mathcal{F}(x(t_i), \lambda_0(t_i), \lambda_1(t_i))}{k_B T} \circ dx \\
&= -\frac{dQ}{k_B T},
\end{aligned} \tag{1.63}$$

where we get to the result by recognizing that line 2 obeys our definition of Stratonovich multiplication rules since x' is the midpoint of dx and that line 3 contains the definition of an increment in the heat transfer from Eq. (1.56). We can

3) For the reader aware of the subtleties of stochastic calculus, we mention that for additive noise as considered here, this point is made largely for completeness: if one constructs the result using the relevant stochastic calculus, the ratio is independent of the choice. However, to be a well-defined quantity for cases involving multiplicative noise, this issue becomes important.

then obtain the entropy production of the entire path by constructing the integral limit of the summation over contributions for each t_i such that

$$k_B \ln \left[\frac{P^F[X(\tau)|X(0)]}{P^R[\bar{X}(\tau)|\bar{X}(0)]} \right] = -\frac{1}{T}\int_0^\tau \mathrm{d}Q = -\frac{\Delta Q}{T} = \frac{\Delta Q_{\mathrm{med}}}{T} = \Delta S_{\mathrm{med}}, \tag{1.64}$$

giving us the expected result noting that the identification of such a term from the ratio of path probabilities can also readily be achieved in full phase space [19].

1.6 Entropy, Stationarity, and Detailed Balance

Let us consider the functional for the total entropy production once more, specifically with a view to understanding when we expect an entropy change. Specifically, we aim to identify two conceptually different situations where entropy production occurs. If we consider a system evolving without external driving, it will typically, for well-defined system parameters, approach some stationary state. That stationary state is characterized by a time-independent probability density P^{st} such that

$$\frac{\partial P^{\mathrm{st}}(x,t)}{\partial t} = 0. \tag{1.65}$$

Let us write down the entropy production for such a situation. Since the system is stationary, we have $P_{\mathrm{start}} = P_{\mathrm{end}}$, but we also have a time-independent protocol, meaning we need not consider distinct forward and reverse processes such that we write path probability densities $P^R = P^F = P$. In this situation, the total entropy production for overdamped motion is given as

$$\Delta S_{\mathrm{tot}}[x] = k_B \ln \left[\frac{P^{\mathrm{st}}(x(0))P[x(\tau)|x(0)]}{P^{\mathrm{st}}(x(\tau))P[x(0)|x(\tau)]} \right]. \tag{1.66}$$

We can then ask what in general are the properties required for entropy production, or indeed no entropy production in such a situation. Clearly, there is no entropy production when the forward and reverse trajectories are equally likely and so we can write the condition for zero entropy production in the stationary state as

$$P^{\mathrm{st}}(x(0))P[x(\tau)|x(0)] = P^{\mathrm{st}}(x(\tau))P[x(0)|x(\tau)], \quad \forall\, x(0), x(\tau). \tag{1.67}$$

Written in this form, we emphasize that this is equivalent to the statement of *detailed balance*. Transitions are said to balance because the average number of all transitions to and from any given configuration $x(0)$ exactly cancel; this leads to a constant probability distribution and is the condition required for a stationary state. However, to have no entropy production in the stationary state, we require all transitions to balance in detail: we require the total number of transitions between every possible combination of two configurations $x(0)$ and $x(\tau)$ to cancel. This is also the condition required for zero probability current and for the system to be at thermal equilibrium where we understand the entropy of the universe to be maximized.

We may then quite generally place any dynamical scheme into one of two broad categories. The first is where detailed balance (Eq. (1.67)) holds and the stationary state is the thermal equilibrium.[4)] Under such dynamics, systems left unperturbed will relax toward equilibrium where there is no observed preferential forward or reverse behavior, no observed thermodynamic arrow of time or irreversibility, and therefore no entropy production. Thus, all entropy production for these dynamics is the result of driving and subsequent relaxation to equilibrium or more generally a consequence of systems being out of their stationary states.

The other category therefore is where detailed balance does not hold. In these situations, we expect entropy production even in the stationary state, which by extension must have origins beyond that of driving out of and relaxation back to stationarity. So, when can we expect detailed balance to be broken? We can first identify the situations where it does hold and for overdamped motion, the requirements are well defined. To have all transitions balancing in detail is to have zero probability current, $J^{\mathrm{st}}(x,t) = 0$, in the stationary state, where the current is related to the probability density according to

$$\frac{\partial P^{\mathrm{st}}(x,t)}{\partial t} = -\frac{\partial J^{\mathrm{st}}(x,t)}{\partial x} = 0. \tag{1.68}$$

Utilizing the form of the Fokker–Planck equation that corresponds to the dynamics, we would thus require

$$J^{\mathrm{st}}(x,t) = \frac{1}{m\gamma}\left(-\frac{\partial\phi(x,\lambda_0(t))}{\partial x} + f(x,\lambda_1(t))\right)P^{\mathrm{st}}(x,t) - \frac{k_{\mathrm{B}}T}{m\gamma}\frac{\partial P^{\mathrm{st}}(x,t)}{\partial x} = 0. \tag{1.69}$$

We can verify the consistency of such a condition by inserting the appropriate stationary distribution:

$$P^{\mathrm{st}}(x,t) \propto \exp\left[\int^{x} \mathrm{d}x' \frac{m\gamma}{k_{\mathrm{B}}T}\left(-\frac{\partial\phi(x',\lambda_0(t))}{\partial x'} + f(x',\lambda_1(t))\right)\right], \tag{1.70}$$

which is clearly of a canonical form. How can one break this condition? We would require a nonvanishing current and this can be achieved when the contents of the exponential in Eq. (1.70) are not integrable. In general, this can be achieved by using an external force that is nonconservative. However, in one dimension with natural, that is reflecting, boundary conditions, any force acts conservatively since the total distance between initial and final positions and thus work done are always path independent. To enable such a nonconservative force, one can implement periodic boundary conditions. This can be realized physically by considering motion on a ring since when a constant force acts on the particle, the work done will depend on the number of times the particle traverses the ring. If the system relaxes to its stationary state, there will be a nonzero, but constant current that

4) One can build models that have stationary states with zero entropy production where equilibrium is only local, but there is no value in distinguishing between the two situations or highlighting such cases here.

arises due to the nonconservative force driving the motion in one direction. In such a system with steady flow, it is quite easy to understand that the transitions in each direction between two configurations will not cancel and thus detailed balance is not achieved. Allowing these dynamics to relax the system to its stationary state creates a simple example of a *nonequilibrium steady state*. Generally, such states can be created by placing some constraint upon the system that stops it from reaching a thermal equilibrium. This results in a system that is perpetually attempting and failing to maximize the total entropy by equilibrating. By remaining out of equilibrium, it constantly dissipates heat to the environment and is thus associated with a constant entropy generation. As such, a system with these dynamics gives rise to irreversibility beyond that arising from driving and relaxation and possesses an underlying breakage of time reversal symmetry, leading to an associated entropy production, manifest in the lack of detailed balance.

Detailed balance may be broken in many ways and the nonequilibrium constraint that causes it may be, as we have seen, a nonconservative force or it might be an exposure to particle reservoirs with unequal chemical potentials or heat baths with unequal temperatures. The steady states of such systems in particular are of great interest in statistical physics, not only because of their qualitatively different behavior but also because they provide cases where analytical solution is feasible out of equilibrium. As we shall see later, the distribution of entropy production in these states also obeys a particular powerful symmetry requirement.

1.7
A General Fluctuation Theorem

So far, we have examined a particular functional of a path and argued from a number of perspectives that it represents the total entropy production of the universe. We have also seen that it obeys a remarkably simple and powerful relation that guarantees its positivity on average. However, we can exploit the form of the entropy production further and derive a number of fluctuation theorems that explicitly relate distributions of general entropy-like quantities. They are numerous and the differences can appear rather subtle; however, it is quite simple to derive a very general equality that we can rigorously and systematically adapt to different situations and arrive at these different relations. To do so, let us once again consider the functional that represents the total entropy production:

$$\Delta S_{\mathrm{tot}}[X] = k_{\mathrm{B}} \ln \left[\frac{P_{\mathrm{start}}(X(0))P^F[X(\tau)|X(0)]}{P^R_{\mathrm{start}}(\bar{X}(0))P^R[\bar{X}(\tau)|\bar{X}(0)]} \right]. \tag{1.71}$$

We are able to construct the probability distribution of this quantity for a particular process. Mathematically, the distribution of entropy production over the forward process can be written as

$$P^F(\Delta S_{\mathrm{tot}}[X] = A) = \int \mathrm{d}X P_{\mathrm{start}}(X(0))P^F[X(\tau)|X(0)]\delta(A - \Delta S_{\mathrm{tot}}[X]). \tag{1.72}$$

To proceed, we follow Harris and Schütz [20] and consider a new functional, but one that is very similar to the total entropy production. We shall generally refer to it as R and it can be written as

$$R[X] = k_{\mathrm{B}} \ln \left[\frac{P^R_{\mathrm{start}}(X(0)) P^R[X(\tau)|X(0)]}{P_{\mathrm{start}}(\bar{X}(0)) P^F[\bar{X}(\tau)|\bar{X}(0)]} \right]. \tag{1.73}$$

Imagine that we evaluate this new quantity over the reverse trajectory, that is, we consider $R[\bar{X}]$. It will be given by

$$R[\bar{X}] = k_{\mathrm{B}} \ln \left[\frac{P^R_{\mathrm{start}}(\bar{X}(0)) P^R[\bar{X}(\tau)|\bar{X}(0)]}{P_{\mathrm{start}}(X(0)) P^F[X(\tau)|X(0)]} \right] = -\Delta S_{\mathrm{tot}}[X], \tag{1.74}$$

which is explicitly the negative value of the functional that represents the total entropy production in the forward process. We can similarly construct a distribution for $R[\bar{X}]$ over the reverse process. This in turn would be given as

$$P^R(R[\bar{X}] = A) = \int \mathrm{d}\bar{X} P^R_{\mathrm{start}}(\bar{X}(0)) P^R[\bar{X}(\tau)|\bar{X}(0)] \delta(A - R[\bar{X}]). \tag{1.75}$$

We now seek to relate this distribution to that of the total entropy production over the forward process. To do so, we consider the value the probability distribution takes for $R[\bar{X}] = -A$. By the symmetry of the delta function, we may write

$$P^R(R[\bar{X}] = -A) = \int \mathrm{d}\bar{X} P^R_{\mathrm{start}}(\bar{X}(0)) P^R[\bar{X}(\tau)|\bar{X}(0)] \delta(A + R[\bar{X}]). \tag{1.76}$$

We now utilize three substitutions. First, $\mathrm{d}X = \mathrm{d}\bar{X}$ denoting the equivalence of the path integrals owing to the Jacobian of unity. Next, we use the definition of the entropy production functional to substitute

$$P^R_{\mathrm{start}}(\bar{X}(0)) P^R[\bar{X}(\tau)|\bar{X}(0)] = P_{\mathrm{start}}(X(0)) P^F[X(\tau)|X(0)] \mathrm{e}^{-\Delta S_{\mathrm{tot}}[X]/k_{\mathrm{B}}} \tag{1.77}$$

and finally the definition that $R[\bar{X}] = -\Delta S_{\mathrm{tot}}[X]$. Performing the above substitutions, we find

$$\begin{aligned} P^R(R[\bar{X}] = -A) &= \int \mathrm{d}X P_{\mathrm{start}}(X(0)) P^F[X(\tau)|X(0)] \mathrm{e}^{-\Delta S_{\mathrm{tot}}[X]/k_{\mathrm{B}}} \delta(A - \Delta S_{\mathrm{tot}}[X]) \\ &= \mathrm{e}^{-(A/k_{\mathrm{B}})} \int \mathrm{d}X P_{\mathrm{start}}(X(0)) P^F[X(\tau)|X(0)] \delta(A - \Delta S_{\mathrm{tot}}[X]) \\ &= \mathrm{e}^{-(A/k_{\mathrm{B}})} P^F(\Delta S_{\mathrm{tot}}[X] = A) \end{aligned} \tag{1.78}$$

and yields the following theorem [20]:

Transient Fluctuation Theorem

$$P^R(R[\bar{X}] = -A) = \mathrm{e}^{-(A/k_{\mathrm{B}})} P^F(\Delta S_{\mathrm{tot}}[X] = A). \tag{1.79}$$

This is a fundamental relation and holds for all protocols and initial conditions and is of a form referred to in the literature as a finite time, transient, or detailed fluctuation theorem depending on where you look. In addition, if we integrate over all values of A on both sides, we obtain the integral fluctuation theorem:

$$1 = \langle \mathrm{e}^{-\Delta S_{\mathrm{tot}}/k_{\mathrm{B}}} \rangle, \tag{1.80}$$

with its name now being self-explanatory. These two relations shall now form the basis of all relations we consider. However, upon returning to the transient fluctuation theorem, a valid question is what does the functional $R[\bar{X}]$ represent? In terms of traditional thermodynamic quantities, there is scant physical interpretation. It is more helpful to consider it as a related functional of the path and to understand that in general it is *not* the entropy production of the reverse path in the reverse process. It is important now to look at why. To construct the entropy production under the reverse process, we need to consider a new functional that we shall call $\Delta S^R_{\text{tot}}[\bar{X}]$, which is defined in exactly the same way as for the forward process. We consider an initial distribution, this time P^R_{start} that evolves to P^R_{end}, and compare the probability density for a trajectory starting from the initial distribution, this time under the reverse protocol $\bar{\lambda}(t)$, with the probability density of a trajectory starting from the time-reversed final distribution, $\hat{T}P^R_{\text{end}}$, so that

$$\Delta S^R_{\text{tot}}[\bar{X}] = k_{\text{B}} \ln \left[\frac{P^R_{\text{start}}(\bar{X}(0)) P^R[\bar{X}(\tau)|\bar{X}(0)]}{\hat{T}P^R_{\text{end}}(X(0)) P^F[X(\tau)|X(0)]} \right] \neq -\Delta S_{\text{tot}}[X]. \tag{1.81}$$

Crucially there is an inequality in Eq. (1.81) in general because

$$\hat{T}P_{\text{start}}(X(0))) \neq P^R_{\text{end}}(\bar{X}(\tau)) = \int \text{d}\bar{X} P^R_{\text{start}}(\bar{X}(0)) P^R[\bar{X}(\tau)|\bar{X}(0)]. \tag{1.82}$$

This is manifest in the irreversibility of the dynamics of the systems we are looking at, as is illustrated in Figure 1.1. If the dynamics were reversible, as for Hamilton's equations and Liouville's theorem, then Eq. (1.82) would hold in equality. So, examining Eqs. (1.79) and (1.81), if we wish to compare the distribution of entropy production in the reverse process with that for the forward process, we need to have $R[\bar{X}] = \Delta S^R_{\text{tot}}[\bar{X}]$ such that $\Delta S_{\text{tot}}[X] = -\Delta S^R_{\text{tot}}[\bar{X}]$. This is achieved by having $P_{\text{start}}(X(0)) = \hat{T}P^R_{\text{end}}(\bar{X}(\tau))$. When this condition is met, we may write

$$P^R(\Delta S^R_{\text{tot}}[\bar{X}] = -A) = \text{e}^{-(A/k_{\text{B}})} P^F(\Delta S_{\text{tot}}[X] = A), \tag{1.83}$$

which now relates distributions of the same physical quantity, entropy change. If we assume that arguments of a probability distribution for the reverse protocol P^R implicitly describe the quantity over the reverse process, we may write it in its more common form:

$$P^R(-\Delta S_{\text{tot}}) = \text{e}^{-\Delta S_{\text{tot}}/k_{\text{B}}} P^F(\Delta S_{\text{tot}}). \tag{1.84}$$

This will hold when the protocol and initial distributions are chosen such that evolution under the forward process followed by the reverse process together with the appropriate time reversals brings the system back into the same initial statistical distribution. This sounds somewhat challenging and indeed does not occur in any generality, but there are two particularly pertinent situations where the above does hold and has particular relevance in a discussion of thermodynamic quantities.

1.7.1
Work Relations

The first and most readily applicable example that obeys the condition $P_{\text{start}}(X(0)) = \hat{T} P^{R}_{\text{end}}(\bar{X}(\tau))$ involves changes between equilibrium states where one can trivially obtain the required condition by exploiting the fact that unperturbed, the dynamics will steadily bring the system into a stationary state that is invariant under time reversal. We start by defining the equilibrium distribution that represents the canonical ensemble where, as before, we consider the system energy for an overdamped system to be entirely described by the potential $\phi(x, \lambda_0(t))$ such that

$$P^{\text{eq}}(x(t), \lambda_0(t)) = \frac{1}{Z(\lambda_0(t))} \exp\left[-\frac{\phi(x(t), \lambda_0(t))}{k_B T}\right] \tag{1.85}$$

for $t = 0$ and τ, where Z is the partition function, uniquely defined by $\lambda_0(t)$, which can in general be related to the Helmholtz free energy through the relation

$$F(\lambda_0(t)) = -k_B T \ln Z(\lambda_0(t)). \tag{1.86}$$

To clarify, the corollary of these statements is to say that the directly applied force $f(x(t), \lambda_1(t))$ does not feature in the system's Hamiltonian.[5)] Let us now choose the initial and final distributions to be given by the respective equilibriums defined by the protocol at the start and finish of the forward process and the same temperature:

$$\begin{aligned} P_{\text{start}}(x(0), \lambda_0(0)) &\propto \exp\left[\frac{F(\lambda_0(0)) - \phi(x(0), \lambda_0(0))}{k_B T}\right], \\ P_{\text{end}}(x(\tau), \lambda_0(\tau)) &\propto \exp\left[\frac{F(\lambda_0(\tau)) - \phi(x(\tau), \lambda_0(\tau))}{k_B T}\right]. \end{aligned} \tag{1.87}$$

We are now in a position to construct the total entropy change for a given realization of the dynamics between these two states. From the initial and final distributions, we can immediately construct the system entropy change ΔS_{sys} as

$$\begin{aligned} \Delta S_{\text{sys}} &= k_B \ln\left(\frac{P_{\text{start}}(x(0), \lambda_0(0))}{P_{\text{end}}(x(\tau), \lambda_0(\tau))}\right) = k_B \ln\left(\frac{\exp\left[(F(\lambda_0(0)) - \phi(x(0), \lambda_0(0)))/k_B T\right]}{\exp\left[(F(\lambda_0(\tau)) - \phi(x(\tau), \lambda_0(\tau)))/k_B T\right]}\right) \\ &= \frac{1}{T}(-F(\lambda_0(\tau)) + F(\lambda_0(0)) + \phi(x(\tau), \lambda_0(\tau)) - \phi(x(0), \lambda_0(0))) \\ &= \frac{\Delta\phi - \Delta F}{T}. \end{aligned} \tag{1.88}$$

The medium entropy change is as we defined previously and can be written

$$\Delta S_{\text{med}} = -\frac{\Delta Q}{T} = \frac{\Delta W - \Delta\phi}{T}, \tag{1.89}$$

5) That is not to say it may not appear in some generalized Hamiltonian. For further insight into this issue, we refer the interested reader to Refs [21, 22], noting that the approach here and elsewhere [23] best resembles the extended relation used in Ref. [21].

where ΔW is the work given earlier in Eq. (1.58), but we now emphasize that this term contains contributions due to changes in the potential and due to the external force f. We thus further define two new quantities ΔW_0 and ΔW_1 such that $\Delta W = \Delta W_0 + \Delta W_1$ with

$$\Delta W_0 = \int_0^{\tau} \frac{\partial \phi(x(t), \lambda_0(t))}{\partial \lambda_0} \frac{\mathrm{d}\lambda_0(t)}{\mathrm{d}t} \mathrm{d}t \tag{1.90}$$

and

$$\Delta W_1 = \int_0^{\tau} f(x(t), \lambda_1(t)) \circ \mathrm{d}x. \tag{1.91}$$

W_0 and W_1 are not defined in the same way with W_0 being found more often in thermodynamics and W_1 being a familiar definition from mechanics: one may therefore refer to these definitions as thermodynamic and mechanical work, respectively. The total entropy production in this case is simply given by

$$\Delta S_{\mathrm{tot}}[x] = \frac{\Delta W - \Delta F}{T}. \tag{1.92}$$

In addition, since we have established that $P^R_{\mathrm{end}}(\bar{x}(\tau)) = \hat{T} P_{\mathrm{start}}(x(0))$, we can also write

$$\Delta S^R_{\mathrm{tot}}[\bar{x}] = -\frac{\Delta W - \Delta F}{T}. \tag{1.93}$$

1.7.1.1 The Crooks Work Relation and Jarzynski Equality

The derivation of several relations follows now by imposing certain constraints on the process we consider. First let us imagine the situation where the external force $f(x, \lambda_1) = 0$ and so all work is performed conservatively through the potential such that $\Delta W = \Delta W_0$. To proceed, we should clarify the form of the protocol that would take an equilibrium system to a new equilibrium such that its reversed counterpart would return the system to the same initial distribution. This would consist of a waiting period, in principle of infinite duration, where the protocol is constant, followed by a period of driving where the protocol changes, followed by another infinitely long waiting period. Such a protocol is given and explained in Figure 1.2.

For such a process, we write the total entropy production:

$$\Delta S_{\mathrm{tot}} = \frac{\Delta W_0 - \Delta F}{T}. \tag{1.94}$$

This changes its sign for the reverse trajectory and reverse protocol and so we may construct the appropriate fluctuation relation that is now simply read off Eq. (1.79) as

$$P^F((\Delta W_0 - \Delta F)/T) = \exp\left[\frac{\Delta W_0 - \Delta F}{k_{\mathrm{B}} T}\right] P^R(-(\Delta W_0 - \Delta F)/T). \tag{1.95}$$

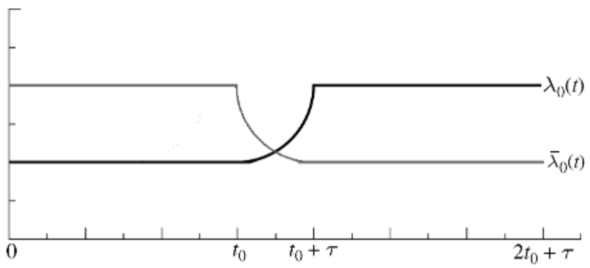

Figure 1.2 A protocol $\lambda_0(t)$, of duration $2t_0 + \tau$, which evolves a system from one equilibrium to another, defined so that the reversed protocol $\bar{\lambda}_0(t)$ returns the system to the original equilibrium. There is a period of no driving of length t_0 that corresponds to the relaxation for the reverse process, followed by a time τ of driving, followed by another relaxation period of duration t_0. As we take t_0 to infinity, we obtain a protocol that produces the condition $P^R_{\mathrm{end}}(\bar{x}(2t_0 + \tau)) = \hat{T}P_{\mathrm{start}}(x(0))$. We note here that $f(x(t), \lambda_1(t)) = 0$.

Since F and T are independent of the trajectory, we can simplify and find the following [5]:

The Crooks Work Relation

$$\frac{P^F(\Delta W_0)}{P^R(-\Delta W_0)} = \exp\left[\frac{\Delta W_0 - \Delta F}{k_B T}\right]. \tag{1.96}$$

Rearranging and integrating over all ΔW on both sides and taking the deterministic ΔF out of the path integral then yields an expression for the average over the forward process [4, 15, 24]:

The Jarzynski Equality

$$\langle \exp(-\Delta W_0/k_B T)\rangle = \exp(-\Delta F/k_B T). \tag{1.97}$$

The power of these statements is clarified in one very important conceptual point. In their formulation, the relations are *constructed* using the values of entropy change for a process that, after starting in equilibrium, is isolated for a long time, driven, and then left for a long time again to return to a stationary state. However, this does not mean that these quantities have to be *measured* over the whole of such a process. Why is this the case? It is because the entropy production for the whole process can be written in terms of the mechanical work and free energy change that are delivered exclusively during the driving phase when the protocol $\lambda_0(t)$ is changing. Since the work and free energy change are independent of the intervals where the protocol is constant and because we had no constraint on $\lambda_0(t)$ during the driving phase, we can therefore consider them to be valid for any protocol assuming the system is in equilibrium to start with. We can therefore state that the Crooks work relation and Jarzynski equality hold *for all times* for systems that start in equilibrium.[6] Historically, this has had one particularly important consequence: the

6) Although we have shown that this is the case for Langevin dynamics, it is important to note that these expressions can be obtained for other general descriptions of the dynamics. See Ref. [25].

results hold for driving, in principle, arbitrarily far from equilibrium. This is widely summed up as the ability to obtain equilibrium information from nonequilibrium averaging since, upon examining the form of the Jarzynski equality, we can compute the free energy difference by taking an average of the exponentiated work done in the course of some nonequilibrium process. Exploiting these facts, let us clarify what these two relations mean explicitly and what the implications are in the real world.

The Crooks Relation

Statement: For any time τ, the probability of observing trajectories that correspond to an application of ΔW_0 work, starting from an equilibrium state defined by $\lambda(0)$, under dynamics described by $\lambda(t)$ in $0 \leq t \leq \tau$, is exponentially more likely in $(\Delta W_0 - \Delta F)/k_B T$ than the probability of observing trajectories that correspond to an application of $-\Delta W_0$ work from an equilibrium state defined by $\bar{\lambda}(0)$, under dynamics described by $\bar{\lambda}(t)$.

Implication: Consider an isothermal gas in a piston in contact with a heat bath at equilibrium. Classically, we know from the second law that if we compress the gas, performing a given amount of work ΔW_0 on it, then after an equilibration period, we must expect the gas to perform an amount of work that is less than ΔW_0 when it is expanded (i.e., $-\Delta W_0$ work performed on the gas). To get the same amount of work back out, we need to perform the process quasi-statically such that it is reversible. The Crooks relation, however, tells us more. For the same example, we can state that if the dynamics of our system lead to some probability of performing ΔW_0 work, then the probability of extracting the same amount of work in the reverse process differs exponentially. Indeed, they only have the same probability when the work performed is equal to the free energy difference, often called the reversible work.

The Jarzynski Equality

Statement: For any time τ, the average value, as defined by the mean over many realizations of the dynamics, of the exponential of the negative work divided by the temperature arising from a defined change in protocol from $\lambda_0(0)$ to $\lambda_0(\tau)$ is identically equal to the exponential of the negative equilibrium free energy difference corresponding to the same change in protocol, divided by the temperature.

Implication: Consider once again the compression of a gas in a piston, but let us imagine that we wish to know the free energy change without knowledge of the equation of state. Classically, we may be able to measure the free energy change by attempting to perform the compression quasi-statically, which of course can never be fully realized. However, the Jarzynski equality states that we can determine this free energy change exactly by repeatedly compressing the gas *at any speed* and taking an average of the exponentiated work that we perform over all these fast compressions. One must, however, exercise caution; the average taken is patently dominated by very negative values of work. These correspond to very negative excursions in entropy and are often *extremely* rare. One may find that the estimated free energy change is significantly altered following one additional realization even if hundreds or perhaps thousands have already been averaged.

These relations very concisely extend the classical definition of irreversibility in such isothermal systems. In classical thermodynamics, we may identify the difference in free energy as the maximum amount of work we may extract from the system, or rather that to achieve a given free energy change, we must perform at least as much work as that free energy change, that is,

$$\Delta W_0 \geq \Delta F, \tag{1.98}$$

with the equality holding for a quasi-static "reversible" process. But since we saw that our entropy functional could take negative values, there is nothing in the dynamics that prevents an outcome where the work is less than the free energy change. We understand now that the second law is statistical, so more generally we must have

$$\langle \Delta W_0 \rangle \geq \Delta F. \tag{1.99}$$

The Jarzynski equality tells us more than this and replaces the inequality with an equality that it is valid for nonquasistatic processes where mechanical work is performed at a finite rate such that the system is driven away from thermal equilibrium and the process is irreversible.

1.7.2
Fluctuation Relations for Mechanical Work

Let us now consider a circumstance similar, but subtly different, to that of the Jarzynski and Crooks relations. We consider a driving process that again starts in equilibrium, but this time keeps the protocol $\lambda_0(t)$ held fixed such that all work is performed by the externally applied force $f(x(t), \lambda_1(t))$, meaning $\Delta W = \Delta W_1$. Once again we seek a fluctuation relation by constructing an equilibrium-to-equilibrium process, though this time we insist that the system relaxes back to the same initial equilibrium distribution. We note that since $f(x(t), \lambda_1(t))$ may act nonconservatively, in order to allow relaxation back to equilibrium we would require that the external force be "turned off." An example set of protocols is given in Figure 1.3 for a simple external force $f(x(t), \lambda_1(t)) = \lambda_1(t)$.

For such a process, we find

$$\Delta S_{\text{tot}} = \frac{\Delta W_1}{T}, \tag{1.100}$$

since the free energy difference between the same equilibrium states vanishes. We have constructed a process such that the distribution at the end of the reverse process is (with time reversal) the same as the initial distribution of the forward

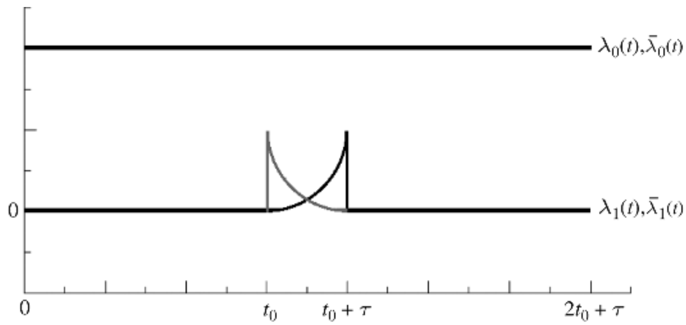

Figure 1.3 An example protocol and reversed protocol that would construct the condition $P^R_{\text{end}}(\bar{x}(2t_0+\tau)) = \hat{T}P_{\text{start}}(x(0))$ when all work is performed through the external force $f(x(t),\lambda_1(t)) = \lambda_1(t)$ and t_0 is taken to infinity.

process and so again we are permitted to read off a set of fluctuation relations [21, 22, 26, 27] that may collectively be referred to as the following:

Fluctuation relations for mechanical work

$$\frac{P^F(\Delta W_1)}{P^R(-\Delta W_1)} = \exp\left[\frac{\Delta W_1}{k_B T}\right], \tag{1.101}$$

$$\langle \exp(-\Delta W_1/k_B T)\rangle = 1. \tag{1.102}$$

For the same reasons as in the Jarzynski and Crooks relations, they are valid for all times and thus hold as a nonequilibrium result. Taking in particular the integrated relation and comparing with the Jarzynski equality in Eq. (1.97), one may think there is an inconsistency. Both are valid for all times and arbitrary driving and concern the work done under the constraint that both start in equilibrium, yet on first inspection they seem to be saying different things. But recall our distinction between the work ΔW_0 and ΔW_1 from Eqs. (1.90) and (1.91); there are two distinct ways to describe work on such a particle. If one performs work ΔW_0, one necessarily changes the form of the system energy, whereas the application of work ΔW_1 leaves the form of the system energy unchanged. The difference is manifest in the two different integrated relations because their derivations exploit the fact that the Hamiltonian, which represents the system energy, appears in initial and final distributions. To clarify, as written the Jarzynski equality explicitly concerns driving where the application of any work also changes the Hamiltonian and thus the equilibrium state. On the other hand, the relations for W_1 concern work as the path integral of an external force such that the Hamiltonian remains unchanged for the entire process.

Of course, there is nothing in the derivation of either of these relations that precludes the possibility of both types of work to be performed at the same time and so using the same arguments, we arrive at

$$\frac{P^F(\Delta W)}{P^R(-\Delta W)} = \exp\left[(\Delta W - \Delta F)/k_B T\right] \tag{1.103}$$

and

$$\left\langle \exp\left(-\frac{\Delta W - \Delta F}{k_{\mathrm{B}} T}\right)\right\rangle = 1, \tag{1.104}$$

again under the constraint that the system be initially prepared in equilibrium.

1.7.3 Fluctuation Theorems for Entropy Production

We have seen in Section 1.7.1 how relations between distributions of work can be derived from Eq. (1.84), since work can be related to the entropy production during a suitable equilibrium-to-equilibrium process. We wish now to seek situations where we can explicitly construct relations that concern the distributions of the entropy produced for forward and reverse processes that do not necessarily begin and end in equilibrium. In order to find situations where the value of the entropy production for the forward trajectory in the forward process is precisely the negative value of the entropy production for the reversed trajectory in the reverse process, we seek situations where the reverse protocol acts to return the distribution to the time-reversed initial distribution for the forward process. For the overdamped motion we have been considering, we would require

$$P_{\mathrm{start}}(x(\tau)) = \int \mathrm{d}x P_{\mathrm{end}}(x(0)) P^R[x(\tau)|x(0)]. \tag{1.105}$$

This is a slightly more general situation than that previously considered of an equilibrium-to-equilibrium process, and in such cases one can expect to see a symmetry between distributions of entropy production in such forward and reverse processes along the lines of Eq. (1.84).

However, we can specify further and find an even more direct symmetry if we insist that the evolution under the forward process is indistinguishable from that under the reverse process. Mathematically, this means $P(x_{i+1}|x_i, \lambda(t_{i+1})) = P(x_{i+1}|x_i, \bar{\lambda}(t_{i+1}))$ or $P^R[x(\tau)|x(0)] = P^F[x(\tau)|x(0)]$. Given these conditions, evolution from the initial distribution will result in the final distribution and evolution under the reverse process from the final distribution will result in the initial distribution and these distributions will be the same. If we consider in more detail the requirements for such a condition, we understand there are two main ways in which this can be achieved. The first way is to require a constant protocol $\lambda(t)$. In this way, the forward process is trivially the same as the reverse process. Alternatively, we require the protocol to be time symmetric such that $\lambda(t) = \lambda(\tau - t) = \bar{\lambda}(t)$. In both situations, the forward and reverse processes are entirely indistinguishable. As such, by careful construction, we can, in these specific circumstances, relate the probability of seeing a positive entropy production to that of a negative entropy production over the *same forward process* allowing us from Eq. (1.79) to write the following [14]:

Detailed fluctuation theorem

$$P(\Delta S_{\text{tot}}) = e^{\Delta S_{\text{tot}}/k_B} P(-\Delta S_{\text{tot}}) \qquad (1.106)$$

Physically, the two situations we have considered correspond to the following:

- $P_{\text{start}} = P_{\text{end}}, \quad \lambda(t) = \text{constant}$
  To satisfy such criteria, the system must be in a steady state, that is, all intrinsic system properties (probability distribution, mean system entropy, mean system energy etc.) must remain constant over the process. The simplest steady state is equilibrium that trivially has zero entropy production in detail for all trajectories. However, a *nonequilibrium* steady state can be achieved by breaking detailed balance through some constraint that prevents the equilibration, as we saw in Section 1.6. The mean entropy production rate of these states is constant, nonzero, and, as we have now shown, there is an explicit exponential symmetry in the probability of positive and negative fluctuations.
- $P_{\text{start}} = P_{\text{end}} = P, \quad \lambda(t) = \bar{\lambda}(t)$
  This condition can be achieved in a system that is being periodically driven characterized by a time symmetric $\lambda(t)$. If from some starting point we allow the system to undergo an arbitrarily large number of periods of driving, it will arrive at a so-called nonequilibrium oscillatory state such that $P(x,t) = P(x, t + t_p)$, where t_p is the period of oscillation. In this state, we can expect the above relation to hold for integer multiples of period t_p starting from a time such that $\lambda(t) = \bar{\lambda}(t)$.

1.8 Further Results

1.8.1 Asymptotic Fluctuation Theorems

In the class of system we have considered, the integral and detailed fluctuation theorems are guaranteed to hold. Indeed, it has not escaped some authors' attention that the reason they do is fully explained in the very definition of the functionals they concern [28]. There is, however, a class of fluctuation theorems that does not have this property. These are known as asymptotic fluctuation theorems. Their derivation for Langevin and then general Markovian stochastic systems is due to Kurchan [29] and Lebowitz and Spohn [30], respectively, and superficially bear strong similarities with results obtained by Gallavotti and Cohen for chaotic deterministic systems [3, 31]. They generally apply to systems that approach a steady state, and, for stochastic systems, strictly in their definition, concern a symmetry in the long-time limit of the generating function of a quantity known as an action functional or flux [30]. This quantity again relates to a trajectory that runs from x_0 through to x_n, described using jump probabilities $\sigma(x_i|x_{i-1})$ and is given as

$$\mathcal{W}(t) = \ln\left[\frac{\sigma(x_1|x_0)}{\sigma(x_0|x_1)}\cdots\frac{\sigma(x_n|x_{n-1})}{\sigma(x_{n-1}|x_n)}\right]. \tag{1.107}$$

An asymptotic fluctuation theorem, which we shall not prove and only briefly address here, states that there exists a long-time limit of the scaled cumulant generating function of $\mathcal{W}(t)$ such that

$$e(s) = \lim_{t\to\infty} -\frac{1}{t}\ln\langle\exp[-s\mathcal{W}(t)]\rangle \tag{1.108}$$

and that this quantity possesses the symmetry

$$e(s) = e(1-s). \tag{1.109}$$

From this somewhat technical definition, we can derive a fluctuation theorem closely related to those that we have already examined. The existence of such a limit implies that the distribution function $P(\mathcal{W}(t)/t)$ of the time-averaged action functional $\mathcal{W}(t)/t$ follows a large deviation behavior [32] such that in the long-time limit, we have

$$P(\mathcal{W}(t)/t) \simeq \mathrm{e}^{-t\hat{e}(\mathcal{W}/t)}, \tag{1.110}$$

where $\hat{e}$ is the Legendre transform of e defined as

$$\hat{e}(\mathcal{W}/t) = \max_s [e(s) - s(\mathcal{W}/t)], \tag{1.111}$$

maximizing over the conjugate variable s. Consequently, using the symmetry relation of Eq. (1.109), we may write

$$\begin{aligned}\hat{e}(\mathcal{W}/t) &= \max_s [e(s) - (1-s)(\mathcal{W}/t)]\\ &= \max_s [e(s) + s(\mathcal{W}/t)] - (\mathcal{W}/t)\\ &= \hat{e}(-\mathcal{W}/t) - (\mathcal{W}/t).\end{aligned} \tag{1.112}$$

Since we expect large deviation behavior described by Eq. (1.110), this implies

$$P(\mathcal{W}/t) \simeq P(-\mathcal{W}/t)\mathrm{e}^{\mathcal{W}} \tag{1.113}$$

or, equivalently,

$$P(\mathcal{W}) \simeq P(-\mathcal{W})\mathrm{e}^{\mathcal{W}}, \tag{1.114}$$

which is clearly analogous to the fluctuation theorems we have seen previously. Taking a closer look at the action functional $\mathcal{W}$, we see that it is, for the systems we have been considering, a representation of the entropy produced in the medium or a measure of the heat dissipated, up to a constant k_B. Unlike the fluctuation theorems considered earlier, this is not guaranteed for all systems. To get a basic understanding of this subtlety, we write the asymptotic fluctuation theorem for the medium entropy production for the continuous systems we have been considering in the form

$$\frac{P(\Delta S_{\text{med}})}{P(-\Delta S_{\text{med}})} \simeq \mathrm{e}^{\Delta S_{\text{med}}/k_B}. \tag{1.115}$$

However, we know that for stationary states, the following fluctuation theorem holds for all time:

$$\frac{P(\Delta S_{\text{tot}})}{P(-\Delta S_{\text{tot}})} = e^{\Delta S_{\text{tot}}/k_B}. \tag{1.116}$$

Since $\Delta S_{\text{tot}} = \Delta S_{\text{med}} + \Delta S_{\text{sys}}$, we may understand that the asymptotic symmetry will exist when the system entropy change is negligible compared to the medium entropy change. In nonequilibrium stationary states, we expect a continuous dissipation of heat and thus an increase of medium entropy, along with a change in system entropy that *on average* is zero. One may naively suggest that this guarantees the asymptotic symmetry since the medium entropy is unbounded and can grow indefinitely. However, if the system configuration space is unbounded, one cannot in general rule out large fluctuations to regions with *arbitrarily* low probability densities and therefore large changes in system entropy, which in principle can persist on any timescale. What is required to guarantee such a relation is the ability to neglect, in detail for all trajectories, the system entropy change compared to medium entropy change on long timescales. This can be done in general if we insist that the state space is bounded. This means that the system entropy has well-defined maximum and minimum values that can be assumed to be unimportant on long timescales and so the asymptotic symmetry necessarily follows. We note finally that systems with unbounded state space are ubiquitous and include simple harmonic oscillators [33] and so investigations of fluctuation theorems for such systems have yielded a wealth of nontrivial generalizations and extensions.

1.8.2 Generalizations and Consideration of Alternative Dynamics

What we hope the reader might appreciate following reading this chapter is the malleability of quantities that satisfy fluctuation relations. It is not difficult to identify quantities that obey relations similar to the fluctuation theorems (although it may be hard to show that they have any particular physical relevance) since the procedure simply relies on a transformation of a path integral utilizing the definition of the entropy production itself. To clarify this point, we consider some generalizations of the relations we have seen. Let us consider a new quantity that has the same form as the total entropy production:

$$G[X] = k_B \ln \left[\frac{P^F[X]}{P[Y]} \right]. \tag{1.117}$$

Here $P^F[X]$ is the same as before, yet we deliberately say very little about $P[Y]$ other than it is a probability density of observing some path Y related to X defined on the same space as X. Let us compute the average of the negative exponential of this quantity:

$$\begin{aligned}\langle \mathrm{e}^{-G/k_B} \rangle &= \int \mathrm{d}X P_{\mathrm{start}}(X(0)) P^F[X(\tau)|X(0)] \frac{P(Y(0))P[Y(\tau)|Y(0)]}{P_{\mathrm{start}}(X(0))P^F[X(\tau)|X(0)]} \\ &= \int \mathrm{d}Y P(Y(0)) P[Y(\tau)|Y(0)] \\ &= 1.\end{aligned} \tag{1.118}$$

If we have $\mathrm{d}X = \mathrm{d}Y$ such that the Jacobian of the transformation to path Y is unity and the unspecified initial distribution $P(Y(0))$ and transition probability density $P[Y(\tau)|Y(0)]$ are normalized, then *any* such quantity $G[X]$ will obey an integral fluctuation theorem averaged over the forward process. Clearly, there are as many relations as there are ways to define $P[Y(\tau)|Y(0)]$ and $P(Y(0))$ and most will mean very little at all [28]. However, there are several such relations in the literature obtained by an appropriate choice of $P(Y(0))$ and $P[Y(\tau)|Y(0)]$ that say something meaningful, including, for example, the Seifert end point relations [34]. We will very briefly allude to just two ways that this can be achieved by first noting that one may consider an alternative dynamics known as "adjoint" dynamics, leading to conditional path probabilities written $P_{\mathrm{ad}}[Y(\tau)|Y(0)]$, defined such that they generate the same stationary distribution as the normal dynamics, but with a probability current of the opposite sign [35]. For the overdamped motion that we have been considering, where $P^F[X] = P_{\mathrm{start}}(x(0))$ $P^F[x(\tau)|x(0)]$, we may derive the following results:

- **Hatano–Sasa relation:** By choosing

$$P(Y(0)) = P^R_{\mathrm{start}}(\bar{x}(0)) \tag{1.119}$$

and

$$P[Y(\tau)|Y(0)] = P^R_{\mathrm{ad}}[\bar{x}(\tau)|\bar{x}(0)], \tag{1.120}$$

we obtain the Hatano–Sasa relation [36] or integral fluctuation theorem for the "nonadiabatic entropy production" [37–39] that concerns the so-called "excess heat" transferred to the environment [40] such that

$$\langle \exp\left[-\Delta Q_{\mathrm{ex}}/k_B T - \Delta S_{\mathrm{sys}}/k_B\right] \rangle = 1. \tag{1.121}$$

The exponent here is best described as the negative of the entropy production associated with movement to stationarity, which phenomenologically includes transitions between nonequilibrium stationary states for which it was first derived. This use of the P^R_{ad} adjoint dynamics is frequently described as a reversal of both the protocol and dynamics [35] to be contrasted with reversal of just the protocol for the integral fluctuation theorem.

- **Relation for the housekeeping heat:** By choosing

$$P(Y(0)) = P_{\mathrm{start}}(x(0)) \tag{1.122}$$

and

$$P[Y(\tau)|Y(0)] = P^F_{\mathrm{ad}}[x(\tau)|x(0)], \tag{1.123}$$

we obtain the Speck–Seifert integral relation [41] or integral fluctuation theorem for the "adiabatic entropy production" [37–39], which concerns the so-called "housekeeping heat" absorbed by the environment [40] such that

$$\langle \exp\left[-\Delta Q_{\rm hk}/k_{\rm B}T\right]\rangle = 1, \tag{1.124}$$

where $\Delta Q_{\rm hk} = \Delta Q_{\rm med} - \Delta Q_{\rm ex}$ [40] and where the negative exponent is best described as the entropy production associated with the nonequilibrium steady state. Such a consideration might be called a reversal of the dynamics, but not the protocol.

Both these relations are relevant to the study of systems where detailed balance does not hold and amount to a division in the total entropy production, or irreversibility, into the two types we considered in Section 1.6, namely, the movement toward stationarity brought about by driving and relaxation, and the breaking of time reversal symmetry that arises specifically when detailed balance is absent. Consequently, if detailed balance does hold, then the exponent in Eq. (1.124) is zero and Eq. (1.121) reduces to the integral fluctuation theorem.

1.9 Fluctuation Relations for Reversible Deterministic Systems

So far we have chosen to focus on systems that obey stochastic dynamics, whereby the interaction with the environment, and the explicit breakage of time reversal symmetry, is implemented through the presence of random forces in the equations of motion. However, there exists a framework for deriving fluctuation relations that is based on deterministic dynamical equations, whereby the environmental interaction is represented through specific nonlinear terms [9], which supplement the usual forces in Newton's equations. These have the effect of constraining some aspect of the system, such as its kinetic energy, either to a chosen constant or to a particular distribution as time progresses. Most importantly, they can be reversible, such that a trajectory driven by a specified protocol and its time-reversed counterpart driven by a time-reversed protocol are both solutions to the dynamics. In practice, these so-called thermostating terms that provide the nonlinearity are taken to act solely on the boundaries (which can be made arbitrarily remote) in order for the system to be unaffected by the precise details of the input and removal of heat. This provides a parallel framework within which the dynamics of an open system, and hence fluctuation relations, can be explored. Indeed, it was through the consideration of deterministic, reversible dynamical systems that many of the seminal insights into fluctuation relations were first obtained [1].

Given that there was a choice over the framework to employ, we opted to use stochastic dynamics to develop this pedagogical overview. This has some benefit in that the concept of entropy change can be readily attached to the idea of the growth of uncertainty in system evolution, identifying it explicitly with the intrinsic irreversibility of the stochastic dynamics. Nevertheless, it is important to review the deterministic approach as well, and explore some of the additional insight that it provides.

The main outcome of seminal and ongoing studies by Evans and coworkers [1, 2, 8] is the identification of a system property that displays a tendency to grow with time under specified non-Hamiltonian reversible dynamics. The development of the H-theorem by Boltzmann was a similar attempt to identify such a quantity. However, we shall have to confront the fact that by their very nature, deterministic equations do not generate additional uncertainty as time progresses. The configuration of a system at a time t is precisely determined given the configuration at t_0. Even if the latter were specified only through a probability distribution, all future and past configurations associated with each starting configuration are fixed, and uncertainty is therefore not changed by the passage of time. Something other than the increase in uncertainty will have to emerge in a deterministic framework if it is to represent entropy change.

Within such a framework, a system is described in terms of a probability density for its dynamical variables $x, v, \ldots$, collectively denoted Γ. An initial probability density $P(\Gamma, 0)$ evolves under the dynamics into a density $P(\Gamma, t)$. Furthermore, the starting "point" of a trajectory Γ_0 (that is, $(x(0), v(0), \ldots)$ is linked uniquely to a terminating point Γ_t (that is, $x(t), v(t), \ldots$), passing through points $\Gamma_{t'}$ in between. It may then be shown [1, 2] that

$$P(\Gamma_t, t) = P(\Gamma_0, 0) \exp\left(-\int_0^t \Lambda(\Gamma_{t'})\mathrm{d}t'\right), \tag{1.125}$$

where $\Lambda(\Gamma_{t'})$ is known as the phase space contraction rate associated with configuration $\Gamma_{t'}$, which may be related specifically to the terms in the equations of motion that impose the thermal constraint. For a system without constraint, and hence thermally isolated, the phase space contraction rate is therefore zero everywhere, and the resulting $P(\Gamma_t, t) = P(\Gamma_0, 0)$ is an expression of Liouville's theorem: the conservation of probability density along any trajectory followed by the system.

For typically employed thermal constraints (denoted thermostats/ergostats, depending on their nature), it may be shown that the phase space contraction rate is related to the rate of heat transfer to the system from the implied environment. For the so-called Nose–Hoover thermostat at fixed target temperature T, we are able to write $\int_0^t \Lambda(\Gamma_{t'})\mathrm{d}t' = \Delta Q(\Gamma_0)/k_B T$, where $\Delta Q(\Gamma_0)$ is the heat transferred to the system over the course of a trajectory of duration t starting from Γ_0.

We now consider the dissipation function $\Omega(\Gamma)$ defined through

$$\int_0^t \Omega(\Gamma_{t'})\mathrm{d}t' = \ln\left(\frac{P(\Gamma_0, 0)}{P(\Gamma_t^*, 0)} \exp\left(-\int_0^t \Lambda(\Gamma_{t'})\mathrm{d}t'\right)\right), \tag{1.126}$$

where Γ_t^* is related to Γ_t by the reversal of all velocity coordinates. Assuming that the probability density at time zero is symmetric in velocities, such that

$P(\Gamma_0, 0) = P(\Gamma_0^*, 0)$ (which ensures that the right-hand side of Eq. (1.126) vanishes when $t = 0$), we can write

$$\int_0^t \Omega(\Gamma_{t'})\mathrm{d}t' = \bar{\Omega}_t(\Gamma_0)t = \ln\left(\frac{P(\Gamma_t, t)}{P(\Gamma_t, 0)}\right), \tag{1.127}$$

defining a mean dissipation function $\bar{\Omega}_t(\Gamma_0)$ for the trajectory starting from Γ_0 and of duration t. It is a quantity that will take a variety of values for a given protocol (the specified dynamics over the period in question) depending on Γ_0, and its distribution has particular properties, just as we found for the distributions of values of functionals such as ΔS_{tot} in Eq. (1.44). For example, we have

$$\begin{aligned}\left\langle \exp\left(-\bar{\Omega}_t t\right)\right\rangle &= \int \mathrm{d}\Gamma_0 P(\Gamma_0, 0)\exp\left(-\bar{\Omega}_t(\Gamma_0)t\right)\\ &= \int \mathrm{d}\Gamma_t P(\Gamma_t, t)\exp\left(-\bar{\Omega}_t t\right)\\ &= \int \mathrm{d}\Gamma_t P(\Gamma_t, 0) = 1,\end{aligned} \tag{1.128}$$

where the averaging is over the various possibilities for Γ_0, or equivalently for Γ_t, and where we have imposed a probability conservation condition $\mathrm{d}\Gamma_0 P(\Gamma_0, 0) = \mathrm{d}\Gamma_t P(\Gamma_t, t)$, implying that $\mathrm{d}\Gamma_t$ is the region of phase space around Γ_t that contains all the end points of trajectories starting within the region $\mathrm{d}\Gamma_0$ around Γ_0. This result takes the same form as the integral fluctuation theorem obtained using stochastic dynamics, but now involves the mean dissipation function. In the deterministic dynamics literature, the result Eq. (1.128) is known as a nonequilibrium partition identity. As a consequence, we can deduce that $\langle\bar{\Omega}_t\rangle \geq 0$, again a result that resembles several already encountered.

Now let us consider a protocol that is time symmetric about its midpoint, and for simplicity, consists of time variation in the form of the system's Hamiltonian. The thermal constraint, as discussed above, is imposed through reversible non-Hamiltonian terms in the dynamics (let us say the Nose–Hoover scheme) and is explicitly time independent and therefore isothermal. For such a case, it is clear that a trajectory running from Γ_0 to Γ_t over the time period $0 \to t$ can be generated in a velocity-reversed form and in reverse sequence, by evolving for the same period forward in time under the same equations of motion, but starting from the velocity-reversed configuration at time t, that is, Γ_t^*. This evolution is precisely that which would be obtained by running a movie of the normal trajectory backward. The velocity-reversed or time-reversed counterpart to each phase space point $\Gamma_{t'}$ is visited, but in the opposite order, and the final configuration is Γ_0^*. The mean dissipation function for such a trajectory would be given by

$$\bar{\Omega}_t(\Gamma_t^*)t = \ln\left(\frac{P(\Gamma_t^*, 0)}{P(\Gamma_0, 0)}\exp\left(-\Delta Q(\Gamma_t^*)/k_B T\right)\right), \tag{1.129}$$

where $\Delta Q(\Gamma_t^*)$ is the heat transfer for this time-reversed trajectory. The symmetry of the protocol, and the symmetry of the Hamiltonian under velocity reversal, allows us to conclude that the heat transfer associated with the trajectory starting from Γ_0 is equal and opposite to that associated with starting point Γ_t^*, and hence the mean dissipation functions for the two trajectories must satisfy $\bar{\Omega}_t(\Gamma_0) = -\bar{\Omega}_t(\Gamma_t^*)$. We can then proceed to derive a specific case of the *Evans–Searles fluctuation theorem* (ESFT) associated with the distribution of values $\bar{\Omega}_t$ taken by the mean dissipation function $\bar{\Omega}_t(\Gamma_0)$:

$$\begin{aligned}
P(\bar{\Omega}_t) &= \int d\Gamma_0 P(\Gamma_0, 0)\delta\big(\bar{\Omega}_t(\Gamma_0) - \bar{\Omega}_t\big) = \int d\Gamma_t P(\Gamma_t, t)\delta\big(\bar{\Omega}_t(\Gamma_0) - \bar{\Omega}_t\big) \\
&= \int d\Gamma_t P(\Gamma_t, t) \exp\big(\bar{\Omega}_t(\Gamma_0)t\big) \frac{P(\Gamma_t, 0)}{P(\Gamma_t, t)} \delta\big(\bar{\Omega}_t(\Gamma_0) - \bar{\Omega}_t\big) \\
&= \exp\big(\bar{\Omega}_t t\big) \int d\Gamma_t P(\Gamma_t, 0)\delta\big(\bar{\Omega}_t(\Gamma_0) - \bar{\Omega}_t\big) \\
&= \exp\big(\bar{\Omega}_t t\big) \int d\Gamma_t P(\Gamma_t, 0)\delta\big(-\bar{\Omega}_t(\Gamma_t^*) - \bar{\Omega}_t\big) \\
&= \exp\big(\bar{\Omega}_t t\big) \int d\Gamma_t^* P(\Gamma_t^*, 0)\delta\big(-\bar{\Omega}_t(\Gamma_t^*) - \bar{\Omega}_t\big) \\
&= \exp\big(\bar{\Omega}_t t\big) P(-\bar{\Omega}_t),
\end{aligned} \tag{1.130}$$

noting that the Jacobian for the transformation of the integration variables from Γ_t to Γ_t^* is unity. Under the assumed conditions, therefore, we have obtained a relation that resembles (but historically preceded) the transient fluctuation theorem (Eq. (1.84)) or detailed fluctuation theorem (Eq. (1.106)) derived within the framework of stochastic dynamics. It only remains to make connections between the mean dissipation function and thermodynamic quantities to complete the parallel development, though it has been argued that the mean dissipation function itself is the more general measure of nonequilibrium behavior [8].

If we assume that the initial distribution is a canonical equilibrium such that $P(\Gamma_0, 0) \propto \exp(-H(\Gamma_0, 0)/k_B T)$, where $H(\Gamma_0, 0)$ is the system's Hamiltonian at $t = 0$, then we find from Eq. (1.126) that

$$\bar{\Omega}_t(\Gamma_0)t = \frac{1}{k_B T}(H(\Gamma_t, 0) - H(\Gamma_0, 0)) - \frac{1}{k_B T}\Delta Q(\Gamma_0), \tag{1.131}$$

and if the Hamiltonian at time t takes the same functional form as it does at $t = 0$, then $H(\Gamma_t, 0) = H(\Gamma_t, t)$ and we get

$$\bar{\Omega}_t(\Gamma_0)t = \frac{1}{k_B T}(H(\Gamma_t, t) - H(\Gamma_0, 0)) - \frac{1}{k_B T}\Delta Q(\Gamma_0) \tag{1.132}$$

$$\bar{\Omega}_t(\Gamma_0)t = \frac{1}{k_B T}(\Delta E(\Gamma_0) - \Delta Q(\Gamma_0)) = \frac{1}{k_B T}\Delta W(\Gamma_0), \tag{1.133}$$

where $\Delta E(\Gamma_0)$ is the change in system energy along a trajectory starting from Γ_0. Hence, the mean dissipation function is proportional to the (here solely thermodynamic) work performed on the system as it follows the trajectory starting from Γ_0. We deduce that the expectation value of this work is positive, and that the probability distribution of work for a time symmetric protocol and starting from canonical equilibrium satisfies the ESFT.

Deterministic methods may be used to derive a variety of statistical results involving the work performed on a system, including the Jarzynski equation and the Crooks relation. Nonconservative work may be included such that relations analogous to Eq. (1.101) may be obtained. However, it seems that a parallel development of the statistics of ΔS_{tot} is not possible. The fundamental problem is revealed if we try to construct the deterministic counterpart to the total entropy production defined in Eqs. (1.44) and (1.64):

$$\Delta S_{\text{tot}}^{\text{det}} = \Delta S_{\text{sys}} + \Delta S_{\text{med}} = -k_B \ln \left(\frac{P(\Gamma_t, t)}{P(\Gamma_0, 0)} \right) - \frac{\Delta Q(\Gamma_0)}{T}. \tag{1.134}$$

According to Eq. (1.125), this is identically zero. As might have been expected, uncertainty does not change under deterministic dynamics and the total entropy, in the form that we have chosen to define it, is constant. Nevertheless, the derivation of relationships involving the statistics of work performed and heat transferred, just alluded to, corresponding to similar expressions obtained using the stochastic dynamics framework, indicate that the use of deterministic reversible dynamics is an equivalent procedure for describing the behavior. Pedagogically, it is perhaps best to focus on just one approach, but a wider appreciation of the field requires an awareness of both.

1.10 Examples of the Fluctuation Relations in Action

The development of theoretical results of the kind we have seen so far is all very well, but their meaning is perhaps best appreciated by considering examples, which we do in this section. We shall consider overdamped stochastic dynamics, such that the velocities are always in an equilibrium Maxwell–Boltzmann distribution and never enter into consideration for entropy production. In the first two cases, we shall focus on the harmonic oscillator, since we understand its properties well. The only drawback of the harmonic oscillator is that it is a rather special case and some of its properties are not general [28, 42], although we deliberately avoid situations where the distributions produced are Gaussian in these examples. In the third case, we describe the simplest of nonequilibrium steady states and illustrate a detailed fluctuation theorem for the entropy, and identify its origin in the breaking of detailed balance.

1.10.1 Harmonic Oscillator Subject to a Step Change in Spring Constant

Let us consider the most simple model of the compression–expansion-type processes that are ubiquitous within thermodynamics. We start with a 1D classical harmonic oscillator subject to a Langevin heat bath. Such a system is governed by the overdamped equation of motion:

$$\dot{x} = -\frac{\kappa x}{m\gamma} + \left(\frac{2k_{\text{B}}T}{m\gamma} \right)^{1/2} \xi(t), \tag{1.135}$$

where κ is the spring constant. The corresponding Fokker–Planck equation is

$$\frac{\partial P(x,t)}{\partial t} = \frac{1}{m\gamma} \frac{\partial(\kappa x P(x,t))}{\partial x} + \frac{k_{\text{B}}T}{m\gamma} \frac{\partial^2 P(x,t)}{\partial x^2}. \tag{1.136}$$

We consider a simple work process, whereby, starting from equilibrium at temperature T, we instigate an instantaneous step change in spring constant from κ_0 to κ_1 at $t = 0$ with the system in contact with the thermal bath at all times. This has the effect of compressing or expanding the distribution. We are then interested in the statistics of the entropy change associated with the process. Starting from Eqs. (1.44) and (1.64) for our definition of the entropy production, we may write

$$\Delta S_{\rm tot} = \frac{\Delta W - \Delta\phi}{T} + k_{\rm B} \ln\left(\frac{P_{\rm start}(x_0)}{P_{\rm end}(x_1)}\right), \tag{1.137}$$

utilizing notation $x_1 = x(t)$ and $x_0 = x(0)$ and $\phi(x) = \kappa x^2/2$. We also have

$$\Delta W = \frac{1}{2}(\kappa_1 - \kappa_0)x_0^2 \tag{1.138}$$

and

$$\Delta\phi = \frac{1}{2}\kappa_1 x_1^2 - \frac{1}{2}\kappa_0 x_0^2, \tag{1.139}$$

and so can write

$$\Delta S_{\rm tot} = -\frac{\kappa_1}{2T}\left(x_1^2 - x_0^2\right) + k_{\rm B} \ln\left(\frac{P_{\rm start}(x_0)}{P_{\rm end}(x_1)}\right). \tag{1.140}$$

Employing an initial canonical distribution

$$P_{\rm start}(x_0) = \left(\frac{\kappa_0}{2\pi k_{\rm B} T}\right)^{1/2} \exp\left(-\frac{\kappa_0 x_0^2}{2k_{\rm B} T}\right), \tag{1.141}$$

the distribution at the end of the process will be given by

$$P_{\rm end}(x_1) = \int_{-\infty}^{\infty} {\rm d}x_0 P_{\rm OU}^{\kappa_1}[x_1|x_0] P_{\rm start}(x_0). \tag{1.142}$$

This is the Ornstein–Uhlenbeck process and so has transition probability density $P_{\rm OU}^{\kappa}$ given by analogy to Eq. (1.23). Hence, we may write

$$\begin{aligned} P_{\rm end}(x_1) &= \int_{-\infty}^{\infty} {\rm d}x_0 \left(\frac{\kappa_1}{2\pi k_{\rm B} T(1 - {\rm e}^{-(2\kappa_1 t/m\gamma)})}\right)^{1/2} \exp\left(-\frac{\kappa_1\left(x_1 - x_0 {\rm e}^{-(\kappa_1 t/m\gamma)}\right)^2}{2k_{\rm B} T(1 - {\rm e}^{-(2\kappa_1 t/m\gamma)})}\right) \\ &\quad \times \left(\frac{\kappa_0}{2\pi k_{\rm B} T}\right)^{1/2} \exp\left(-\frac{\kappa_0 x_0^2}{2k_{\rm B} T}\right) \\ &= \left(\frac{\tilde{\kappa}(t)}{2\pi k_{\rm B} T}\right)^{1/2} \exp\left(-\frac{\tilde{\kappa}(t) x_1^2}{2k_{\rm B} T}\right), \end{aligned} \tag{1.143}$$

where

$$\tilde{\kappa}(t) = \frac{\kappa_0 \kappa_1}{\kappa_0 + {\rm e}^{-2\kappa_1 t/(m\gamma)}(\kappa_1 - \kappa_0)}, \tag{1.144}$$

such that P_{end} is always Gaussian. The coefficient $\tilde{\kappa}(t)$ evolves monotonically from κ_0 at $t = 0$ to κ_1 as $t \to \infty$. Substituting this into Eq. (1.140) allows us to write

$$\Delta S_{\text{tot}}(x_1, x_0, t) = -\frac{\kappa_1}{2T}\left(x_1^2 - x_0^2\right) + \frac{k_B}{2}\ln\left(\frac{\kappa_0}{\tilde{\kappa}(t)}\right) - \frac{\kappa_0 x_0^2}{2T} + \frac{\tilde{\kappa}(t)x_1^2}{2T} \tag{1.145}$$

for the entropy production associated with a trajectory that begins at x_0 and ends at x_1 at time t, and is not specified in between. We can average this over the probability distribution for such a trajectory to get

$$\begin{aligned}\langle \Delta S_{\text{tot}} \rangle &= \int dx_0 dx_1 P_{\text{OU}}^{\kappa_1}[x_1|x_0] P_{\text{start}}(x_0) \Delta S_{\text{tot}}(x_1, x_0, t) \\ &= k_B\left(-\frac{1}{2}\frac{\kappa_1}{\tilde{\kappa}(t)} + \frac{1}{2}\frac{\kappa_1}{\kappa_0} + \frac{1}{2}\ln\left(\frac{\kappa_0}{\tilde{\kappa}(t)}\right) - \frac{1}{2} + \frac{1}{2}\right) \\ &= \frac{k_B}{2}\left(\frac{\kappa_1}{\kappa_0} - \frac{\kappa_1}{\tilde{\kappa}(t)} + \ln\left(\frac{\kappa_0}{\tilde{\kappa}(t)}\right)\right),\end{aligned} \tag{1.146}$$

making full use of the separation of ΔS_{tot} into quadratic terms and the Gaussian character of the distributions. $\langle \Delta S_{\text{tot}} \rangle$ is zero at $t = 0$ and as $t \to \infty$, we find

$$\lim_{t\to\infty}\langle \Delta S_{\text{tot}} \rangle = \frac{k_B}{2}\left(\frac{\kappa_1}{\kappa_0} - 1 - \ln\left(\frac{\kappa_1}{\kappa_0}\right)\right), \tag{1.147}$$

which is positive since $\ln z \le z - 1$ for all z. Furthermore,

$$\frac{d\langle \Delta S_{\text{tot}} \rangle}{dt} = \frac{k_B}{2\tilde{\kappa}^2}\frac{d\tilde{\kappa}}{dt}(\kappa_1 - \tilde{\kappa}), \tag{1.148}$$

and it is clear that this is positive at all times during the evolution, irrespective of the values of κ_1 and κ_0. If $\kappa_1 > \kappa_0$, then $\tilde{\kappa}$ increases with time while remaining less than κ_1, and all factors on the right-hand side of Eq. (1.148) are positive. If $\kappa_1 < \kappa_0$, then $\tilde{\kappa}$ decreases but always remains greater than κ_1 and the mean total entropy production is still positive as the relaxation process proceeds.

The work done on the system is simply the input of potential energy associated with the shift in spring constant:

$$\Delta W(x_1, x_0, t) = \frac{1}{2}(\kappa_1 - \kappa_0)x_0^2, \tag{1.149}$$

and so the mean work performed up to any time $t > 0$ is

$$\langle \Delta W \rangle = \frac{k_B T}{2}\left(\frac{\kappa_1}{\kappa_0} - 1\right), \tag{1.150}$$

which is greater than $\Delta F = (k_B T/2)\ln(\kappa_1/\kappa_0)$. The mean dissipative work is

$$\langle \Delta W_d \rangle = \langle \Delta W \rangle - \Delta F = \frac{k_B T}{2}\left(\frac{\kappa_1}{\kappa_0} - 1 - \ln\left(\frac{\kappa_1}{\kappa_0}\right)\right), \tag{1.151}$$

and this equals the mean entropy generated as $t \to \infty$ derived in Eq. (1.147), which is to be expected since the system started in equilibrium. More specifically, let us verify the Jarzynski equality:

$$\begin{aligned}\langle \exp\left(-\Delta W/k_{\mathrm{B}}T\right)\rangle &= \int \mathrm{d}x_0 P_{\mathrm{start}}(x_0)\exp\left(-(\kappa_1-\kappa_0)x_0^2/2k_{\mathrm{B}}T\right)\\ &= (\kappa_0/\kappa_1)^{1/2} = \exp\left(-\Delta F/k_{\mathrm{B}}T\right).\end{aligned} \tag{1.152}$$

Now we demonstrate that the integral fluctuation relation is satisfied. We consider

$$\begin{aligned}\langle \exp\left(-\Delta S_{\mathrm{tot}}/k_{\mathrm{B}}\right)\rangle &= \left\langle \exp\left(\frac{\kappa_1}{2k_{\mathrm{B}}T}\left(x_1^2-x_0^2\right)-\frac{1}{2}\ln\left(\frac{\kappa_0}{\tilde{\kappa}}\right)+\frac{\kappa_0 x_0^2}{2k_{\mathrm{B}}T}-\frac{\tilde{\kappa}x_1^2}{2k_{\mathrm{B}}T}\right)\right\rangle\\
&= \int \mathrm{d}x_1\mathrm{d}x_0 P_{\mathrm{OU}}^{\kappa_1}[x_1|x_0]P_{\mathrm{start}}(x_0)\\
&\quad\times\exp\left(\frac{\kappa_1}{2k_{\mathrm{B}}T}\left(x_1^2-x_0^2\right)-\frac{1}{2}\ln\left(\frac{\kappa_0}{\tilde{\kappa}}\right)+\frac{\kappa_0 x_0^2}{2k_{\mathrm{B}}T}-\frac{\tilde{\kappa}x_1^2}{2k_{\mathrm{B}}T}\right)\\
&= \left(\frac{\tilde{\kappa}}{\kappa_0}\right)^{1/2}\int \mathrm{d}x_1\mathrm{d}x_0\\
&\quad\times\left(\frac{\kappa_1}{2\pi k_{\mathrm{B}}T\left(1-\mathrm{e}^{-(2\kappa_1 t/m\gamma)}\right)}\right)^{1/2}\exp\left(-\frac{\kappa_1\left(x_1-x_0\mathrm{e}^{-(\kappa_1 t/m\gamma)}\right)^2}{2k_{\mathrm{B}}T(1-\mathrm{e}^{-(2\kappa_1 t/m\gamma)})}\right)\\
&\quad\times\left(\frac{\kappa_0}{2\pi k_{\mathrm{B}}T}\right)^{1/2}\exp\left(-\frac{\kappa_0 x_0^2}{2k_{\mathrm{B}}T}\right)\exp\left(\frac{\kappa_1}{2k_{\mathrm{B}}T}\left(x_1^2-x_0^2\right)+\frac{\kappa_0 x_0^2}{2k_{\mathrm{B}}T}-\frac{\tilde{\kappa}x_1^2}{2k_{\mathrm{B}}T}\right)=1,\end{aligned} \tag{1.153}$$

which is a tedious integration, but the result is inevitable.

Furthermore, we can directly confirm the Crooks relation for this process. The work over the forward process is given by Eq. (1.149) and so, choosing $\kappa_1 > \kappa_0$, we can derive its distribution according to Eq. (1.154):

$$\begin{aligned}P^F(\Delta W) &= P_{\mathrm{eq}}(x_0)\frac{\mathrm{d}x_0}{\mathrm{d}\Delta W}\\
&= \left(\frac{\kappa_0}{2\pi k_{\mathrm{B}}T}\right)^{1/2}\exp\left(-\frac{\kappa_0 x_0^2}{2k_{\mathrm{B}}T}\right)\frac{1}{\kappa_1-\kappa_0}\left(\frac{2\Delta W}{\kappa_1-\kappa_0}\right)^{-1/2}H(\Delta W)\\
&= \frac{1}{\sqrt{4\pi k_{\mathrm{B}}T}}\left(\frac{\kappa_0}{\kappa_1-\kappa_0}\right)^{1/2}\Delta W^{-1/2}\exp\left(-\frac{\kappa_0}{\kappa_1-\kappa_0}\frac{\Delta W}{k_{\mathrm{B}}T}\right)H(\Delta W),\end{aligned} \tag{1.154}$$

where $H(\Delta W)$ is the Heaviside step function. Let us consider the appropriate reverse process. Starting in equilibrium defined by κ_1, where again to form the reversed protocol we must have $\kappa_1 > \kappa_0$, the work is

$$\Delta W = \frac{1}{2}(\kappa_0-\kappa_1)x_1^2 \tag{1.155}$$

and so we can derive its distribution according to Eq. (1.156):

$$\begin{aligned}P^R(\Delta W) &= P_{\mathrm{eq}}(x_1)\frac{\mathrm{d}x_1}{\mathrm{d}\Delta W}\\
&= \left(\frac{\kappa_1}{2\pi k_{\mathrm{B}}T}\right)^{1/2}\exp\left(-\frac{\kappa_1 x_1^2}{2k_{\mathrm{B}}T}\right)\frac{1}{\kappa_0-\kappa_1}\left(\frac{2\Delta W}{\kappa_0-\kappa_1}\right)^{-1/2}H(-\Delta W)\\
&= \frac{1}{\sqrt{4\pi k_{\mathrm{B}}T}}\left(\frac{\kappa_1}{\kappa_0-\kappa_1}\right)^{1/2}\Delta W^{-1/2}\exp\left(-\frac{\kappa_1}{\kappa_0-\kappa_1}\frac{\Delta W}{k_{\mathrm{B}}T}\right)H(-\Delta W),\end{aligned} \tag{1.156}$$

so that

$$P^R(-\Delta W) = \frac{1}{\sqrt{4\pi k_B T}}\left(\frac{\kappa_1}{\kappa_0-\kappa_1}\right)^{1/2}(-\Delta W)^{-1/2}\exp\left(\frac{\kappa_1}{\kappa_0-\kappa_1}\frac{\Delta W}{k_B T}\right)H(\Delta W)$$
$$= \frac{1}{\sqrt{4\pi k_B T}}\left(\frac{\kappa_1}{\kappa_1-\kappa_0}\right)^{1/2}(\Delta W)^{-1/2}\exp\left(-\frac{\kappa_1}{\kappa_1-\kappa_0}\frac{\Delta W}{k_B T}\right)H(\Delta W). \tag{1.157}$$

Taking the ratio of these two distributions (1.154) and (1.157) gives

$$\frac{P^F(\Delta W)}{P^R(-\Delta W)} = \sqrt{\frac{\kappa_0}{\kappa_1}}\exp\left(-\frac{\kappa_0-\kappa_1}{\kappa_1-\kappa_0}\frac{\Delta W}{k_B T}\right)$$
$$= \exp\left(\frac{\Delta W - (k_B T/2)\ln(\kappa_1/\kappa_0)}{k_B T}\right) \tag{1.158}$$
$$= \exp\left((\Delta W - \Delta F)/k_B T\right),$$

which is the Crooks work relation as required.

1.10.2 Smoothly Squeezed Harmonic Oscillator

Now let us consider a process where work is performed isothermally on a particle, but this time by a continuous variation of the spring constant. We have

$$\Delta W = \int_0^\tau \frac{\partial\phi(x(t),\lambda(t))}{\partial\lambda}\frac{d\lambda}{dt}dt, \tag{1.159}$$

where $\lambda(t) = \kappa(t)$ and $\phi(x(t),\kappa(t)) = (1/2)\kappa(t)x_t^2$, where $x_t = x(t)$, such that

$$\Delta W = \int_0^\tau \frac{1}{2}\dot{\kappa}(t)x_t^2 dt. \tag{1.160}$$

Similarly, the change in system energy will be given simply by

$$\Delta\phi = \int_0^\tau \frac{d\phi(x(t),\lambda(t))}{dt}dt = \frac{1}{2}\kappa(\tau)x_\tau^2 - \frac{1}{2}\kappa_0 x_0^2. \tag{1.161}$$

Accordingly, we can once again describe the entropy production as

$$\Delta S_{tot} = \frac{1}{2T}\int_0^\tau \dot{\kappa}x_t^2 dt - \frac{1}{2T}\kappa(\tau)x_\tau^2 + \frac{1}{2T}\kappa_0 x_0^2 + k_B\ln\left(\frac{P_{start}(x_0)}{P_{end}(x_\tau)}\right). \tag{1.162}$$

For convenience, we assume the initial state to be in canonical equilibrium. The evolving distribution P satisfies the appropriate Fokker–Planck equation:

$$\frac{\partial P}{\partial t} = \frac{\kappa(t)}{m\gamma}\frac{\partial(xP)}{\partial x} + \frac{k_B T}{m\gamma}\frac{\partial^2 P}{\partial x^2}. \tag{1.163}$$

Since P is initially canonical, it retains its Gaussian form and can be written

$$P_{end}(x_\tau) = P(x_\tau,\tau) = \left(\frac{\tilde{\kappa}(\tau)}{2\pi k_B T}\right)^{1/2}\exp\left(-\frac{\tilde{\kappa}(\tau)x_\tau^2}{2k_B T}\right), \tag{1.164}$$

where $\tilde{\kappa}(t)$ evolves according to

$$\frac{d\tilde{\kappa}}{dt} = -\frac{2}{m\gamma}\tilde{\kappa}(\tilde{\kappa} - \kappa), \tag{1.165}$$

with initial condition $\tilde{\kappa}(0) = \kappa_0$. We can solve for $\tilde{\kappa}$: write $z = \tilde{\kappa}^{-1}$ such that

$$\frac{dz}{dt} = \frac{2}{m\gamma}(1 - \kappa z). \tag{1.166}$$

This has integrating factor solution:

$$z(\tau)\exp\left(\frac{2}{m\gamma}\int_0^\tau \kappa(t)dt\right) = z(0) + \int_0^\tau \exp\left(\frac{2}{m\gamma}\int_0^t \kappa(t')dt'\right)\frac{2}{m\gamma}dt, \tag{1.167}$$

or, equivalently,

$$\frac{1}{\tilde{\kappa}(\tau)} = \frac{1}{\kappa(0)}\exp\left(-\frac{2}{m\gamma}\int_0^\tau \kappa(t)dt\right) + \frac{2}{m\gamma}\int_0^\tau \exp\left(-\frac{2}{m\gamma}\int_t^\tau \kappa(t')dt'\right)dt. \tag{1.168}$$

Returning to the entropy production, we now write

$$\Delta S_{\text{tot}} = \frac{1}{T}\int_0^\tau \frac{1}{2}\dot{\kappa}x_t^2 dt - \frac{1}{2T}\kappa(\tau)x_\tau^2 + \frac{1}{2T}\kappa_0 x_0^2 + \frac{k_B}{2}\ln\left(\frac{\kappa_0}{\tilde{\kappa}(\tau)}\right) - \frac{\kappa_0 x_0^2}{2T} + \frac{\tilde{\kappa}(\tau)x_\tau^2}{2T}, \tag{1.169}$$

and we also have

$$\Delta W = \int_0^\tau \frac{1}{2}\dot{\kappa}x_t^2 dt. \tag{1.170}$$

We can investigate the statistics of these quantities:

$$\langle \Delta W \rangle = \int_0^\tau \frac{1}{2}\dot{\kappa}\langle x_t^2 \rangle dt = \int_0^\tau \frac{1}{2}\dot{\kappa}\frac{k_B T}{\tilde{\kappa}}dt, \tag{1.171}$$

and from $\Delta W_d = \Delta W - \Delta F$, the rate of performance of dissipative work is

$$\frac{d\langle \Delta W_d \rangle}{dt} = \frac{\dot{\kappa}k_B T}{2\tilde{\kappa}} - \frac{dF(\kappa(t))}{dt} = \frac{\dot{\kappa}k_B T}{2\tilde{\kappa}} - \frac{\dot{\kappa}k_B T}{2\kappa} = \frac{k_B T}{2}\dot{\kappa}\left(\frac{1}{\tilde{\kappa}} - \frac{1}{\kappa}\right). \tag{1.172}$$

While the positivity of $\langle \Delta W_d \rangle$ is ensured for this process, as a consequence of the Jarzynski equation and the initial equilibrium condition, the rate of change can be both positive and negative, according to this result.

The expectation value for total entropy production in Eq. (1.169), on the other hand, is

$$\langle \Delta S_{\text{tot}} \rangle = \frac{1}{T}\int_0^\tau \frac{1}{2}\dot{\kappa}\frac{k_B T}{\tilde{\kappa}}dt - \frac{1}{2T}\kappa(\tau)\frac{k_B T}{\tilde{\kappa}} + \frac{1}{2T}k_B T + \frac{k_B}{2}\ln\left(\frac{\kappa_0}{\tilde{\kappa}(\tau)}\right) - \frac{k_B T}{2T} + \frac{k_B T}{2T} \tag{1.173}$$

and the rate of change of this quantity is

$$\begin{aligned}\frac{d\langle \Delta S_{tot}\rangle}{dt} &= \frac{\dot{\kappa} k_B}{2\tilde{\kappa}} - \frac{k_B}{2}\left(\frac{\dot{\kappa}}{\tilde{\kappa}} - \frac{\kappa}{\tilde{\kappa}^2}\dot{\tilde{\kappa}}\right) - \frac{k_B}{2}\frac{\dot{\tilde{\kappa}}}{\tilde{\kappa}} \\ &= \frac{k_B}{2}\dot{\tilde{\kappa}}\frac{(\kappa - \tilde{\kappa})}{\tilde{\kappa}^2} \\ &= \frac{k_B}{m\gamma}\frac{(\kappa - \tilde{\kappa})^2}{\tilde{\kappa}}.\end{aligned} \tag{1.174}$$

The monotonic increase in entropy with time is explicit. The mean dissipative work and the entropy production for a process of this kind starting in equilibrium are illustrated in Figure 1.4, where the protocol changes over a driving period followed by a subsequent period of equilibration. Note particularly that the mean entropy production never exceeds the mean dissipative work, which is delivered instantaneously, and that both take the same value as $t \to \infty$ giving insight into the operation of the Jarzynski equality, as discussed in Section 1.7.1.

It is of more interest, however, to verify that detailed fluctuation relations hold. Analytical demonstration based upon Eq. (1.169) and the probability density for a particular trajectory throughout the entire period are challenging, but a numerical

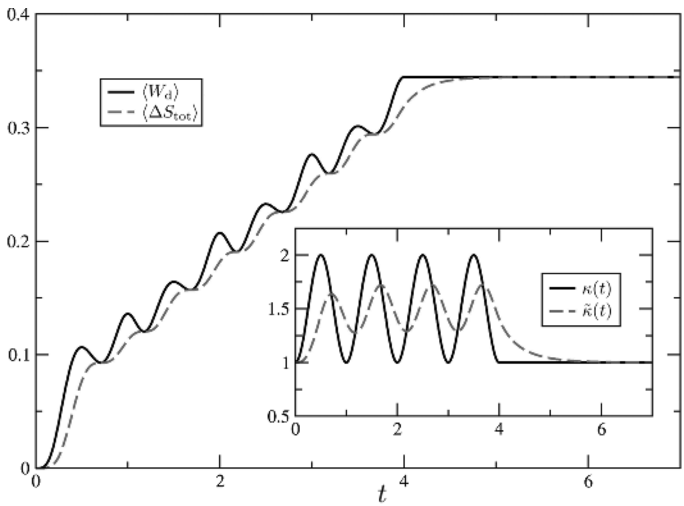

Figure 1.4 An illustration of the mean behavior of the dissipative work and entropy production for an oscillatory compression and expansion process starting in equilibrium. The mean dissipative work increases, but not monotonically, and is delivered instantly such that there is no further contribution when the protocol stops changing. The mean entropy production, however, continues to increase monotonically until it reaches the mean dissipative work after the protocol has stopped changing. The evolution of the protocol, $\kappa(t) = \sin^2(\pi t) + 1$, and the characterization of the distribution, $\tilde{\kappa}(t)$, are shown in the inset. Units are $k_B = T = m = \gamma = 1$.

approach based upon generating sample trajectories is feasible particularly since the entire distribution can always be characterized with the known quantity $\tilde{\kappa}(t)$. As such we consider the same protocol, $\kappa(t) = \sin^2(\pi t) + 1$, wait until the system has reached a nonequilibrium oscillatory steady state, as described in Section 1.7.3, characterized here by an oscillatory $\tilde{\kappa}(t)$, as seen in Figure 1.4, and measure the entropy production over a time period across which $\kappa(t)$ is symmetric. The distribution in total entropy production over such a period and the symmetry it possesses are illustrated in Figure 1.5.

1.10.3
A Simple Nonequilibrium Steady State

Let us construct a very simple nonequilibrium steady state. We consider an overdamped Brownian motion on a ring driven in one direction by a nonconservative force. We assume a constant potential $\phi(x) = c$ such that the equation of motion is simply

$$\dot{x} = \frac{f}{m\gamma} + \left(\frac{2k_B T}{m\gamma}\right)^{1/2} \xi(t). \tag{1.175}$$

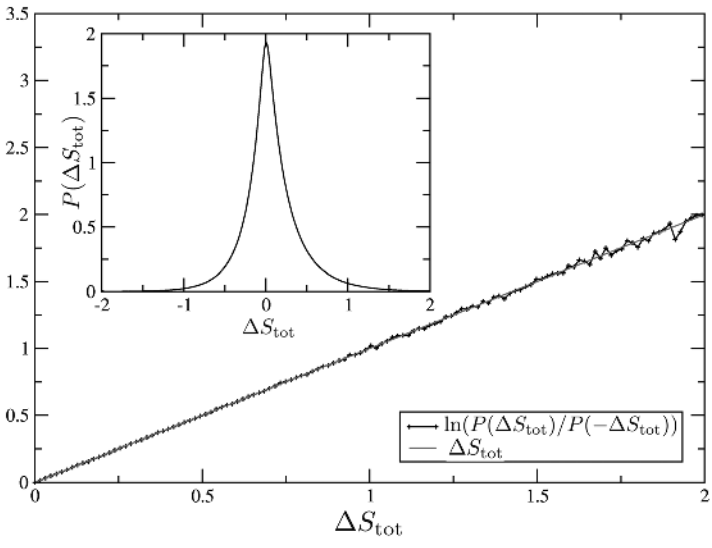

Figure 1.5 An illustration of a detailed fluctuation theorem arising for an oscillatory nonequilibrium steady state, as described in Section 1.7.3, created by compressing and expanding a particle in a harmonic potential by using protocol $\lambda(t) = \kappa(t) = \sin^2(\pi t) + 1$. The total entropy production must be measured over a time period during which the protocol is symmetric and the distribution is deemed to be oscillatory. Such a time period exists between $t = 3$ and $t = 4$, as shown in inset in Figure 1.4. Units are $k_B = T = m = \gamma = 1$.

This is just the Wiener process from Eq. (1.18) centered on a mean proportional to the external force multiplied by the time, so the probability density of a given displacement is defined by

$$P[x(\tau)|x(0)] = \sqrt{\frac{m\gamma}{4\pi k_B T\tau}} \exp\left[-\frac{m\gamma(\Delta x - (f/m\gamma)\tau)^2}{4k_B T\tau}\right], \tag{1.176}$$

where the lack of a superscript on P recognizes the constancy of the protocol, and noting that $\Delta x = x(\tau) - x(0)$ may extend an arbitrary number of times around the ring. In addition, by utilizing the symmetry of the system, we can trivially state that the stationary distribution is given by

$$P^{\mathrm{st}}(x) = L^{-1}, \tag{1.177}$$

where L is the circumference of the ring. Considering that we are in the steady state, we know that the transitions must balance in total; however, let us consider the transitions between individual configurations: comparing the probabilities of transitions, we immediately see that

$$P^{\mathrm{st}}(x(0))P[x(\tau)|x(0)] \neq P^{\mathrm{st}}(x(\tau))P[x(0)|x(\tau)]. \tag{1.178}$$

Detailed balance explicitly does not hold. For this system, not only can there be entropy production due to driving, such as is the case with expansion and compression processes, but there is also a continuous probability current in the steady state, in the direction of the force, which dissipates heat into the thermal bath. We have previously stated in Section 1.7.3 that the distribution of the entropy production in such steady states obeys a detailed fluctuation theorem for all times. We can verify that this is the case. The entropy production is rather simple and is given by

$$\Delta S_{\mathrm{tot}} = k_B \ln \frac{P^{\mathrm{st}}(x(0))P[x(\tau)|x(0)]}{P^{\mathrm{st}}(x(\tau))P[x(0)|x(\tau)]} = k_B \ln \frac{L\exp\left[-\frac{m\gamma(\Delta x - (f/m\gamma)\tau)^2}{4k_B T\tau}\right]}{L\exp\left[-\frac{m\gamma(-\Delta x - (f/m\gamma)\tau)^2}{4k_B T\tau}\right]} = \frac{f\Delta x}{T}. \tag{1.179}$$

This provides an example where the entropy production is highly intuitive. Taking $f > 0$, if the particle moves with the probability current, $\Delta x > 0$, it is carrying out the expected behavior and thus is following an entropy generating trajectory. However, if the particle moves against the current, $\Delta x < 0$, it is behaving unexpectedly and as such is performing a trajectory that destroys entropy. It follows that since an observation of the particle flowing with a current is more likely than an observation of the opposite, then on average the entropy production is positive.

Since the system is in a steady state, we expect a detailed fluctuation theorem. The transformation of the probability distribution is trivial and we have simply

$$P(\Delta S_{\rm tot}) = \sqrt{\frac{m\gamma T}{4\pi k_{\rm B} f^2 \tau}} \exp\left[-\frac{m\gamma T\left(\Delta S_{\rm tot} - (f^2/m\gamma T)\tau\right)^2}{4k_{\rm B} f^2 \tau}\right], \qquad (1.180)$$

and we can verify a detailed fluctuation theorem that holds for all time. We can however probe further. While we may conceive of some fluctuations against a steady flow for a small particle, we would be quite surprised to see such deviations if we were considering a macroscopically sized object. Despite the model's limitations, let us consider an approach to macroscopic behavior while maintaining constant the ratio f/m such that the mean particle velocity is unchanged. Both the mean and variance of the distribution of entropy production increase in proportion. On the scale of the mean, the distribution of entropy change increasingly looks like a narrower and narrower Gaussian until it inevitably, for a macroscopic object, approaches a delta function where we recover the classical thermodynamic limit and are unable to perceive the fluctuations any longer.

1.11 Final Remarks

The aim of this chapter was to explore the origin, application, and limitations of fluctuation relations. We have done this within a framework of stochastic dynamics with white noise and often employing the overdamped limit in example cases where the derivations are easier: it is in the analysis of explicit examples where understanding is often to be found. Nevertheless, the results can be extended to other more complicated stochastic systems, though the details will need to be sought elsewhere. The fluctuation relations can also be derived within a framework of deterministic, reversible dynamics, which we have discussed briefly in Section 1.9. It is interesting to note that within that framework, irreversibility finds its origins in nonlinear terms that provide a contraction of phase space, in contrast to the more direct irreversibility of the equations of motion found in stochastic descriptions. Both approaches, however, are attempts to represent a dissipative environment that imposes a thermal constraint.

The fluctuation relations concern the statistics of quantities associated with thermodynamic processes, in particular the mechanical work done upon, or the heat transferred to a system in contact with a heat bath. In the thermodynamic limit, the statistics are simple: there are negligible deviations from the mean, and work and heat transfers appear to be deterministic and the second law requires entropy change to be nonnegative. But for finite size systems, there are fluctuations, and the statistics of these will satisfy one or more fluctuation relations. These can be very specific requirements, for example relating the probability of a fluctuation with positive dissipative work to the probability of a fluctuation with negative dissipative work in the reversed process. Or, the outcome can take the form of an inequality that demonstrates that the mean dissipative work over all possible realizations of the process is positive.

The core concept in the analysis, within the framework of stochastic dynamics at least, is entropy production. This no longer need to be a mysterious concept: it is a natural measure of the departure from dynamical reversibility, the loosening of the hold of Loschmidt's expectation of reversibility, when system interactions with a coarse-grained environment are taken into account. Entropy production emerges in stochastic models where there is uncertainty in initial specification. Intuitively, uncertainty in configuration in such a situation will grow with time, and the mean entropy production is this concept commodified. And it turns out that entropy production can also be related, in certain circumstances, to heat and work transfers, allowing the growth of uncertainty to be monitored in terms of thermodynamic process variables. Moreover, although it is *expected* to be positive, entropy change can be negative; and the probability of such an excursion, possibly observed by a measurement of work done or heat transferred, can be described by a fluctuation relation. In the thermodynamic limit, the entropy production appears to behave deterministically and to violate time reversal symmetry, and only then does the second law acquire its unbreakability. But for small systems interacting with a much larger environment, this status is very much diminished, and the second law is revealed to be merely a statement about what is likely to happen to the system, according to rules governing the evolution of probability that explicitly break time reversal symmetry.

References

1 Evans, D.J. Cohen, E.G.D., and Morriss, G.P. (1993) Probability of second law violations in shearing steady states. *Phys. Rev. Lett.*, **71**, 2401.

2 Evans, D.J. and Searles, D.J. (1994) Equilibrium microstates which generate second law violating steady states. *Phys. Rev. E*, **50**, 1645–1648.

3 Gallavotti, G. and Cohen, E.G.D. (1995) Dynamical ensembles in nonequilibrium statistical mechanics. *Phys. Rev. Lett.*, **74**, 2694.

4 Jarzynski, C. (1997) Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.*, **78**, 2690.

5 Crooks, G.E. (1999) Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Phys. Rev. E*, **60**, 2721–2726.

6 Cercignani, C. (1998) *Ludwig Boltzmann: The Man Who Trusted Atoms*, Oxford University Press.

7 Kondepudi, D. (2008) *Introduction to Modern Thermodynamics*, John Wiley & Sons, Inc., New York.

8 Evans, D.J. Williams, S.R., and Searles, D.J. (2011) A proof of Clausius' theorem for time reversible deterministic microscopic dynamics. *J. Chem. Phys.*, **134**, 204113.

9 Evans, D.J. and Searles, D.J. (2002) The fluctuation theorem. *Adv. Phys.*, **51**, 1529–1585.

10 Jaynes, E. (2003) *Probability Theory: The Logic of Science*, Cambridge University Press.

11 Risken, H. (1989) *The Fokker–Planck Equation: Methods of Solution and Applications*, 2nd edn, Springer.

12 Gardiner, C. (2009) *Stochastic Methods: A Handbook for the Natural and Social Sciences*, Springer.

13 van Kampen, N. (2007) *Stochastic Processes in Physics and Chemistry*, North-Holland.

14 Seifert, U. (2005) Entropy production along a stochastic trajectory and an

integral fluctuation theorem. *Phys. Rev. Lett.*, **95**, 040602.
15 Crooks, G.E. (1998) Nonequilibrium measurements of free energy differences for microscopically reversible Markovian systems. *J. Stat. Phys.*, **90**, 1481–1487.
16 Sekimoto, K. (1998) Langevin equation and thermodynamics. *Prog. Theor. Phys. Suppl.*, **130**, 17.
17 Sekimoto, K. (2010) *Stochastic Energetics, Lecture Notes in Physics*, vol. **799**, Springer, Berlin.
18 Wissel, C. (1979) Manifolds of equivalent path integral solutions of the Fokker–Planck equation. *Z. Phys. B*, **35**, 185–191.
19 Imparato, A. and Peliti, L. (2006) Fluctuation relations for a driven Brownian particle. *Phys. Rev. E*, **74**, 026106.
20 Harris, R.J. and Schütz, G.M. (2007) Fluctuation theorems for stochastic dynamics. *J. Stat. Mech.*, P07020.
21 Horowitz, J. and Jarzynski, C. (2007) Comparison of work fluctuation relations. *J. Stat. Mech.*, P11002.
22 Jarzynski, C. (2007) Comparison of far-from-equilibrium work relations. *C. R. Phys.*, **8** (5–6), 495–506.
23 Seifert, U. (2008) Stochastic thermodynamics: principles and perspectives. *Eur. Phys. J. B*, **64** (3–4), 423–431.
24 Jarzynski, C. (1997) Equilibrium free-energy differences from nonequilibrium measurements: a master-equation approach. *Phys. Rev. E*, **56**, 5018–5035.
25 Jarzynski, C. (2004) Nonequilibrium work theorem for a system strongly coupled to a thermal environment. *J. Stat. Mech.*, P09005.
26 Bochkov, G. and Kuzovlev, Y. (1981) Nonlinear fluctuation-dissipation relations and stochastic models in nonequilibrium thermodynamics: I. Generalized fluctuation-dissipation theorem. *Physica A*, **106**, 443–479.
27 Bochkov, G. and Kuzovlev, Y. (1981) Nonlinear fluctuation-dissipation relations and stochastic models in nonequilibrium thermodynamics: II. Kinetic potential and variational principles for nonlinear irreversible processes. *Physica A*, **106**, 480–520.
28 Shargel, B.H. (2010) The measure-theoretic identity underlying transient fluctuation theorems. *J. Phys. A*, **43**, 135002.
29 Kurchan, J. (1998) Fluctuation theorem for stochastic dynamics. *J. Phys. A*, **31**, 3719.
30 Lebowitz, J.L. and Spohn, H. (1999) A Gallavotti–Cohen type symmetry in the large deviation functional for stochastic dynamics. *J. Stat. Phys.*, **95**, 333–365.
31 Gallavotti, G. and Cohen, E. (1995) Dynamical ensembles in stationary states. *J. Stat. Phys.*, **80**, 931–970.
32 Touchette, H. (2009) The large deviation approach to statistical mechanics. *Phys. Rep.*, **478** (1–3), 1–69.
33 van Zon, R. and Cohen, E.G.D. (2003) Extension of the fluctuation theorem. *Phys. Rev. Lett.*, **91**, 110601.
34 Schmiedl, T. Speck, T., and Seifert, U. (2007) Entropy production for mechanically or chemically driven biomolecules. *J. Stat. Phys.*, **128**, 77–93.
35 Chernyak, V.Y. Chertkov, M., and Jarzynski, C. (2006) Path-integral analysis of fluctuation theorems for general Langevin processes. *J. Stat. Mech*, P08001.
36 Hatano, T. and Sasa, S. (2001) Steady-state thermodynamics of Langevin systems. *Phys. Rev. Lett.*, **86**, 3463–3466.
37 Esposito, M. and Van den Broeck, C. (2010) Three detailed fluctuation theorems. *Phys. Rev. Lett.*, **104**, 090601.
38 Esposito, M. and Van den Broeck, C. (2010) Three faces of the second law: I. Master equation formulation. *Phys. Rev. E*, **82**, 011143.
39 Van den Broeck, C. and Esposito, M. (2010) Three faces of the second law: II. Fokker–Planck formulation. *Phys. Rev. E*, **82**, 011144.
40 Oono, Y. and Paniconi, M. (1998) Steady state thermodynamics. *Prog. Theor. Phys. Suppl.*, **130**, 29–44.
41 Speck, T. and Seifert, U. (2005) Integral fluctuation theorem for the housekeeping heat. *J. Phys. A*, **38**, L581.
42 Saha, A. Lahiri, S., and Jayannavar, A.M. (2009) Entropy production theorems and some consequences. *Phys. Rev. E*, **80**, 011117.