Excursions in Statistical Dynamics

Gavin E. Crooks

GEC

Excursions in Statistical Dynamics

by

Gavin Earl Crooks

B.Sc. (University of East Anglia) 1992
M.Sc. (University of East Anglia) 1993

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Chemistry

in the

GRADUATE DIVISION
of the
UNIVERSITY of CALIFORNIA at BERKELEY

Committee in charge:

Professor David Chandler, Chair
Professor Robert A. Harris
Professor Daniel S. Rokhsar

1999

Typeset April 19, 2002

http://threeplusone.com/pubs/GECthesis.html

Abstract

There are only a very few known relations in statistical dynamics that are valid for systems driven arbitrarily far away from equilibrium by an external perturbation. One of these is the fluctuation theorem, which places conditions on the entropy production probability distribution of nonequilibrium systems. Another recently discovered far-from-equilibrium expression relates nonequilibrium measurements of the work done on a system to equilibrium free energy differences. In contrast to linear response theory, these expressions are exact no matter the strength of the perturbation, or how far the system has been driven from equilibrium. In this work I show that these relations (and several other closely related results) can all be considered special cases of a single theorem. This expression is explicitly derived for discrete time and space Markovian dynamics, with the additional assumptions that the unperturbed dynamics preserve the appropriate equilibrium ensemble, and that the energy of the system remains finite.

These theoretical results indicate that the most interesting nonequilibrium phenomena will be observed during rare excursions of the system away from the stable states. However, direct simulation of the dynamics is inherently inefficient, since the majority of the computation time is taken watching small, uninteresting fluctuations about the stable states. Transition path sampling has been developed as a Monte Carlo algorithm to efficiently sample rare transitions between stable or metastable states in equilibrium systems. Here, the transition path sampling methodology is adapted to the efficient sampling of large fluctuations in nonequilibrium systems evolving according to Langevin's equations of motion. Simulations are then used to study the behavior of the Maier-Stein system, an important model for a large class of nonequilibrium systems. Path sampling is also implemented for the kinetic Ising model, which is then employed to study surface induced evaporation.

Contents

List of Figures

Acknowledgements

A number of professional mentors are responsible for getting me to where I am right now. Brian Robinson and Gareth Rees successfully launched my research career, whose trajectory was significantly perturbed by David Coker, who taught, via example, that Theory Is Fun™. Yaneer Bar-Yam taught me many, many things; but I think the most important was that it is *all* interesting. There is never any need to restrict your attention to the one little subfield in which you happen to specialize. There are interesting advances in all branches of science, all the time, and they're all interconnected. Finally, and by no means least, David Chandler has been the most consummate of advisors. Always ready to cajole, threaten, advise and suggest when needed, but also willing to let me wander off on my own, down obscure back roads and interesting byways.

Life would have been dull (or at least less caffeinated) without the company of my fellow travelers on this perilous road to knowledge: Zoran Kurtovic, Ka Lum, Jordi Marti, Xueyu Song, Raymond Yee, Hyung-June Woo, Félix Csajka, Phillip Geissler, Peter Bolhuis, Christoph Dellago, David Huang and Tom McCormick. In particular, I would like to thank Phill for his proofreading of this thesis, and subsequent lessons in the art of English.

I am also indebted to our administrative assistants: Mary Hammond, Jan Jackett, Sarah Emmer, Isabella Mendoza, and many others far to numerous to mention here, without whom the group would have rapidly fallen to pieces. (no, wait ...)

A chance meeting with Christopher Jarzinski lead to a long and extremely profitable collaboration. Many of the ideas and results in this work had their genesis in our emailed conversations.

Substantial emotional support was provided by Stefanie Breedy. Thank you for the attention. I'm sorry for the lost weekends. However, first and foremost and beyond all others, my parents are most responsible for my being here. They instilled and encouraged a love of science from a very young age. My only regret is that this dissertation was not completed a few short years earlier. My mother would really have appreciated it.

GEC

For Juliet H. Crooks

0. Introduction

This fundamental law $[\langle A\rangle = \frac{1}{Z}\sum_i A_i e^{-\beta E_i}]$ *is the summit of statistical mechanics, and the entire subject is either the slide-down from this summit, as the principle is applied to various cases, or the climb-up to where the fundamental law is derived...*
— R. P. Feynman [32]

Equilibrium statistical mechanics has a simple and elegant structure. For any appropriate system (typically one with a moderate to very large number of degrees of freedom in thermal contact with a heat bath), the phase space probability distribution of the system at equilibrium is given by the canonical distribution of Gibbs [33]. This expression is very general, since the details of a particular system enter through the density of energy states, and the dynamics of the system are almost irrelevant. The explicit calculation of the properties of the system may be hard or impossible in practice, but at least we know where to start.

The development of statistical dynamics has not (yet?) reached this level of elegance. There are simple, useful and wide ranging nonequilibrium theories, such as nonequilibrium thermodynamics and linear response. However, these theories are all approximations that are limited in their range of applicability, typically to systems very near equilibrium. This is a distressing characteristic for a subject whose ambitions should encompass general nonequilibrium dynamics, irrespective of the proximity of the ensemble to equilibrium. Ideally, the fundamental relations of statistical dynamics would be applicable to many different dynamics without modification, and would be exact no matter how far the system had been driven from equilibrium. Near equilibrium theorems would be simple approximations of these exact relations. The explicit calculation of the statistical properties of the dynamics might be hard or impossible in practice, but at least we would know where to start.

The situation has improved in recent years. Several relations have been discovered that are valid for thermostatted systems that have been driven arbitrarily far from equilibrium by an external perturbation. In 1993 Evans, Cohen, and Morriss [1] noted that fluctuations in the nonequilibrium entropy production obey a simple relation, which has subsequently been formulated as the entropy fluctuation theorem(s) (see Chapter 3). A flurry of activity has extended these relations to many different dynamics (both deterministic and stochastic), and has shown that some near equilibrium

theorems can be derived from them. In an entirely independent development, C. Jarzynski [34] derived a relation between ensemble averages of the nonequilibrium work and equilibrium free energy differences. This was also extended to cover a variety of different dynamics and ensembles.

The first part of this thesis (Chapters 1-5) is based on the observation that these disparate far-from-equilibrium relations are actually closely related, and can be derived from the same, simple assumptions. Indeed, both these relations (and several others) can be considered special cases of a single theorem. The principle that unifies these theorems and their applicable dynamics is that the probability of trajectory in a thermostatted system is related to the probability of the corresponding time reversed trajectory. This principle, referred to hereafter as microscopic reversibility, is applicable to many of the dynamics that are routinely used to model real systems. The genesis of these ideas, and their chronological development, can be traced in Refs. [35, 13, 18].

In Chapters 6-7 I turn to the related problem of efficiently simulating nonequilibrium systems. For equilibrium properties it is often advantageous to use the Monte Carlo algorithm to sample state space with the correct probabilities. Similarly, it can be advantageous to use Monte Carlo to sample the trajectory space of a nonequilibrium system. The necessary computational power is now readily available, and efficient methods for generating a Markov chain of trajectories have recently been developed. I use these techniques to study the nonequilibrium behavior of several stochastic systems.

The subject of far-from-equilibrium statistical dynamics is still in its infancy, and many limitations and obstacles remain. Perhaps the most serious of these is that it is extremely difficult to observe the direct consequences of the far-from-equilibrium theorems in real systems. However, it should be possible to test these relations for realistic models using path ensemble Monte Carlo, and it seems probable that practical approximations await discovery. The explicit calculation of dynamic properties, and the direct application of exact far-from-equilibrium relations may be hard or impossible in practice, but at least we have made a start.

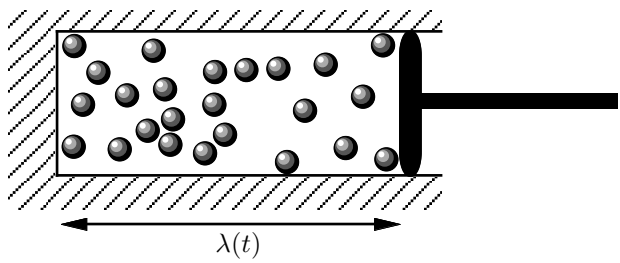
1. Microscopic Reversibility

We will begin by embarking on the climb.

— R. P. Feynman [32]

1.1 Introduction

Consider a classical system in contact with a constant temperature heat bath where some degree of freedom of the system can be controlled. Manipulation of this degree of freedom will result in an expenditure of some amount of work, a change in energy and free energy of the system, and an exchange of heat with the bath. A concrete example is illustrated below; a classical gas confined in a cylinder by a movable piston. The walls of this cylinder are diathermal so that the gas is in thermal contact with the surroundings, which therefore act as the bath. Work can be applied to the system by moving the piston in or out. The value of the controllable, work inducing, parameter (in this case the position of the piston) is traditionally denoted by the variable λ.



The results in the following chapters are based on the following principle: If a system is in contact with a heat bath then the probability of a particular trajectory (given the initial state) is related to the probability of the corresponding time reversed trajectory by a simple function of the temperature and the heat. This relation remains valid for systems that have been driven far from equilibrium* by an external perturbation.

*The phrases nonequilibrium and far-from-equilibrium have more than one connotation. In this work, far-from-equilibrium refers to a configurational ensemble that is very different from the equilibrium ensemble of the same system because the system has been

This important property, which I refer to as microscopic reversibility, was originally shown to be applicable to detailed balanced stochastic dynamics as an intermediate step in the proof of the Jarzynski nonequilibrium work relation [35] (See Chapter 4.) This has since been generalized to include many of the dynamics used to model thermostatted system [17, 18]. Here, I will show that systems are microscopically reversible if the energy is alway finite, and the dynamics are Markovian and preserve the equilibrium distribution of an unperturbed system.

1.2 Discrete time Markov chains

For simplicity, consider a system with stochastic dynamics, a finite set of states, $x \in \{1, 2, 3, \cdots, N\}$, and a discrete time scale $t \in \{0, 1, 2, 3, \cdots, \tau\}$. In other words, a discrete time Markov chain. The energies of the states of the system are given by the vector $\mathbf{E}$. If these state energies do not vary with time then the stationary probability distribution, π_x, is given by the canonical ensemble of equilibrium statistical mechanics;

$$\pi_x = \rho(x|\beta, \mathbf{E}) = \frac{e^{-\beta E_x}}{\sum_x e^{-\beta E_x}} = \exp\left\{\beta F - \beta E_x\right\}. \tag{1.1}$$

In this expression, $\beta = 1/k_{\rm B}T$, T is the temperature of the heat bath, $k_{\rm B}$ is Boltzmann's constant[†], $F(\beta, \mathbf{E}) = -\beta^{-1} \ln \sum_x \exp\{-\beta E_x\} = -\beta^{-1} \ln Z$, is the Helmholtz free energy of the system, Z is the canonical partition function, and the sums are over all states of the system. This distribution, $\rho(x|\beta, \mathbf{E})$, is the probability of state x given the temperature of the heat bath and the energies of all the states. We adopt the convention that when there is insufficient information to uniquely determine the probability distribution, then the maximum entropy distribution should be assumed [36, 37, 38].

In contrast to an equilibrium ensemble, the probability distribution of a nonequilibrium ensemble is not determined solely by the external constraints, but explicitly depends on the dynamics and history of the system. The state of the system at time t is $x(t)$, and the path, or trajectory that the system takes through this state space can be represented by the vector $\mathbf{x} = \big(x(0), x(1), x(2), \cdots, x(\tau)\big)$. For Markovian [39] dynamics the probability of making a transition between states in a particular time step depends only on the current state of the system, and not on the previous history.

perturbed away from equilibrium by an external disturbance. Elsewhere, nonequilibrium is used to refer to an ensemble that cannot reach equilibrium because the dynamics are very slow or glassy.

[†] In theoretical work such as this, it is convenient to measure entropy in nats ($= 1/\ln 2$ bits) [20]. Boltzmann's constant can then be taken to be unity.

The single time step dynamics are determined by the transition matrix[‡] $M(t)$ whose elements are the transition probabilities,

$$M(t)_{x(t+1)x(t)} \equiv \mathrm{P}\big(x(t) \to x(t{+}1)\big)\,.$$

A transition matrix M, has the properties that all elements must be non-negative, and that all columns sum to one due to the normalization of probabilities.

$$\begin{aligned} M_{ij} &\geqslant 0 \qquad \text{for all } i \text{ and } j, \\ \sum_i M_{ij} &= 1 \qquad \text{for all } j. \end{aligned}$$

Let $\rho(t)$ be a (column) vector whose elements are the probability of being in state i at time t. Then the single time step dynamics can be written as

$$\rho(t{+}1) = M(t)\,\rho(t), \tag{1.2}$$

or equivalently as

$$\rho(t{+}1)_i = \sum_j M(t)_{ij}\,\rho(t)_j\,.$$

The state energies $\mathbf{E}(t)$ and the transition matrices $M(t)$ are functions of time due to the external perturbation of the system, and the resulting Markov chain is non-homogeneous in time [43]. The vector of transition matrices $\mathbf{M} = \big(M(0), M(2), \cdots, M(\tau{-}1)\big)$ completely determine the dynamics of the system. We place the following additional constraints on the dynamics: the state energies are always finite (this avoids the possibility of an infinite amount of energy being transferred from or to the system), and the single time step transition matrices must preserve the corresponding canonical distribution. This canonical distribution, Eq. (1.1), is determined by the temperature of the heat bath and the state energies at that time step. We say that such a transition matrix is balanced, or that the equilibrium distribution $\pi(t)$ is an invariant distribution of $M(t)$.

$$\pi(t) = M(t)\,\pi(t)$$

Essentially this condition says that if the system is already in equilibrium (given $\mathbf{E}(t)$ and β), and the system is unperturbed, then it must remain in equilibrium.

It is often convenient to impose the much more restrictive condition of detailed balance,

$$M(t)_{ij}\,\pi(t)_j = M(t)_{ji}\,\pi(t)_i\,. \tag{1.3}$$

[‡]In much of the Markov chain literature (e.g., [40, 41, 39]) a different convention is used such that the transition matrix is the transpose of the matrix defined here, and the probability density is represented by a row vector rather than a column vector. The convention used in this work is common in the physics literature (e.g., [42])

In particular many Monte Carlo simulations are detailed balanced. However it is not a necessity in such simulations [42], and it is not necessary here. It is sufficient that the transition matrices are balanced.

Microscopic reversibility is a relation between probability ratios of trajectories and the heat. The heat $\mathcal{Q}$, is flow of energy into into the system from the attached heat bath, and the work $\mathcal{W}$ is the change in energy due to mechanical perturbations applied to the system. These definitions are perfectly adequate for macroscopic thermodynamics [44], but it may not be immediately clear how they should be expressed in terms of the microscopic degrees of freedom common in statistical dynamics.

Consider this sketch of the microscopic dynamics of a three state system. The horizontal axis is time, the vertical axis is energy, and the dot represents the current state of the system. In the first step the system makes a transition between states. This transition could occur even if there is absolutely no external perturbation of the system, and therefore no work done on the system. Therefore, this energy change can be associated with an exchange of energy with the heat bath. In the second step the system remains in the same state, but the energy of the system changes because the external perturbation changed the energies of all the states. We associate this change in energy with the work. Thus, on a microscopic level the heat is a change in energy due to a change in the state of the system, and the work is a change in energy due to a change in the energy of the state that the system currently resides in.

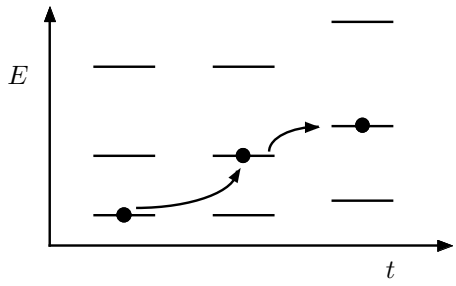


Figure 1.1: Microscopic dynamics of a three state system.

These definitions of work and heat can readily be applied to the Markovian dynamics defined above. Each time step of this dynamics can be separated into two distinct substeps. At time $t = 0$ the system is in state $x(0)$ with energy $E(0)_{x(0)}$. In the first substep the system makes a stochastic transition to a state $x(1)$ which has energy $E(0)_{x(1)}$. This causes an amount of energy, $E(0)_{x(1)} - E(0)_{x(0)}$, to enter the system in the form of heat. In the second substep the state energies change from $\mathbf{E}(0)$ to $\mathbf{E}(1)$ due to the systematic perturbation acting on the system. This requires an amount of work, $E(1)_{x(1)} - E(0)_{x(1)}$. This sequence of substeps repeats for a total of τ time steps[‡]. The total heat exchanged with the reservoir, $\mathcal{Q}$, the total work performed on the system, $\mathcal{W}$, and the total change in energy,

ΔE, are therefore

$$\mathcal{Q}[\mathbf{x}] = \sum_{t=0}^{\tau-1} \Big[E(t)_{x(t+1)} - E(t)_{x(t)} \Big], \tag{1.4}$$

$$\mathcal{W}[\mathbf{x}] = \sum_{t=0}^{\tau-1} \Big[E(t{+}1)_{x(t+1)} - E(t)_{x(t+1)} \Big], \tag{1.5}$$

$$\Delta E = E(\tau)_{x(\tau)} - E(0)_{x(0)} = \mathcal{W} + \mathcal{Q}. \tag{1.6}$$

The reversible work, $W_{\mathrm{r}} = \Delta F = F\big(\beta, \mathbf{E}(\tau)\big) - F\big(\beta, \mathbf{E}(0)\big)$, is the free energy difference between two equilibrium ensembles. It is the minimum average amount of work required to change one ensemble into another. The dissipative work, $\mathcal{W}_{\mathrm{d}}[\mathbf{x}] = \mathcal{W}[\mathbf{x}] - W_{\mathrm{r}}$, is defined as the difference between the actual work and the reversible work. Note that the total work, the dissipative work and the heat are all path functions. In this thesis they are written with script letters, square brackets and/or as functions of the path, $\mathbf{x}$, to emphasize this fact. In contrast ΔE is a state function; it depends only on the initial and final state.

Now we will consider the effects of a time reversal on this Markov chain. In many contexts a time reversal is implemented by permuting the states of the system. For example, in a Hamiltonian system a time reversal is typically implemented by inverting the momenta of all the particles. However, it is equivalent, and in the current context much more convenient, to apply the effects of the time reversal to the dynamics rather than the state space. Thus, the time reversed trajectory $\widehat{\mathbf{x}}$ is a simple reordering of the forward trajectory; $\widehat{x}(t) = x(\tau\text{-}t)$ and $\widehat{\mathbf{E}}(t) = \mathbf{E}(\tau\text{-}t)$.

We can derive the effect of a time reversal on a transition matrix by considering a time homogeneous Markov chain, a chain with a time-independent transition matrix. Let π be the invariant distribution of the transition matrix. If the system is in equilibrium then a time reversal should have no effect on that distribution. Therefore, the probability of observing the transition $i \to j$ in the forward chain should be the same as the probability of observing the transition $j \to i$ in the time reversed chain. Because the equilibrium probability distribution is the same for both chains it follows that

$$\widehat{M}_{ji}\, \pi_i = M_{ij}\, \pi_j \qquad \text{for all} \quad i, j\,. \tag{1.7}$$

In matrix notation this may conveniently be written as

$$\widehat{M} = \mathrm{diag}(\pi) M^T \mathrm{diag}(\pi)^{-1}.$$

[‡]An astute reader will note that the order of the heat and work substeps has been changed when compared with earlier work [35]. This minor change (it amounts to no more than relabeling forward and reverse chains) was made so that the time indexing of the forward chain transition matrix would be compatible with that used in the non-homogeneous Markov chain literature [43].

Here, diag(π) indicates a matrix whose diagonal elements are given by the vector π. $\widehat{M}$ is referred to as the reversal of M [39], or as the π-dual of M [41]. If the transition matrix obeys detailed balance, Eq. (1.3), then $\widehat{M} = M$.

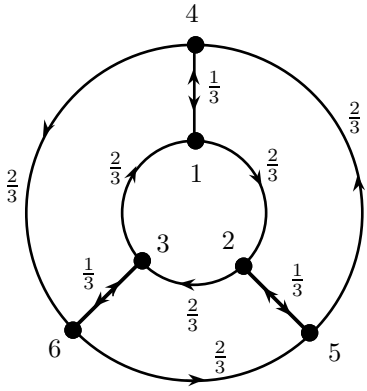
It is easy to confirm that $\widehat{M}$ is a transition matrix; all entries are nonnegative because all equilibrium and transition probabilities are nonnegative, and all rows sum to 1,

$$\sum_j \widehat{M}_{ji} = \frac{1}{\pi_i} \sum_j M_{ij}\, \pi_j = \frac{\pi_i}{\pi_i} = 1 \qquad \text{for all} \quad i\,.$$

Furthermore, we can demonstrate that $\widehat{M}$ and M have the same invariant distribution,

$$\sum_i \widehat{M}_{ji}\, \pi_i = \sum_i M_{ij}\, \pi_j = \pi_j\,.$$

As an illustration of this time reversal operation consider the Markov chain described by the diagram below, and the corresponding stochastic matrix. States are numbered 1 through 6, and transitions are indicated by arrows labeled with the transition probability. The system tends to rotate clockwise through the states (1,2,3) and anticlockwise through the states (4,5,6). The equilibrium probability distribution is $\pi = (\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6})$ and $\widehat{M} = M^T$. Accordingly, the diagram of the time reversed Markov chain is the same except that all arrows reverse direction. Thus, the order in which states are visited reverses, and the system will tend to rotate anticlockwise through the states (1,2,3) and clockwise through the states (4,5,6).



$$M = \begin{pmatrix} 0 & 0 & \frac{2}{3} & \frac{1}{3} & 0 & 0 \\ \frac{2}{3} & 0 & 0 & 0 & \frac{1}{3} & 0 \\ 0 & \frac{2}{3} & 0 & 0 & 0 & \frac{1}{3} \\ \frac{1}{3} & 0 & 0 & 0 & \frac{2}{3} & 0 \\ 0 & \frac{1}{3} & 0 & 0 & 0 & \frac{2}{3} \\ 0 & 0 & \frac{1}{3} & \frac{2}{3} & 0 & 0 \end{pmatrix}$$

For the non-homogeneous chain, the time reversal of the vector of transition matrices, $\mathbf{M}$, is defined as

$$\widehat{M}(t) = \text{diag}(\pi(\tau\text{-}t))M(\tau\text{-}t)^T\text{diag}(\pi(\tau\text{-}t))^{-1}. \tag{1.8}$$

The time reversal operation is applied to each transition matrix, and their time order is reversed. Note that for the transition matrices of the reverse chain the time index runs from 1 to τ, rather than 0 to τ–1. Therefore, $M(t)$ is the transition matrix from time t to time t+1 (see Eq. (1.2)), but $\widehat{M}(t)$ is the transition matrix from time t–1 to time t.

$$\widehat{\rho}(t) = \widehat{M}(t)\,\widehat{\rho}(t\text{–}1)\,. \tag{1.9}$$

This convention is chosen so that the time indexes of the various entities remain consistent. Thus, for the reverse chain at time t the state is $\widehat{x}(t)$, the states energies are $\widehat{\mathbf{E}}(t)$ and the corresponding equilibrium distribution is $\widehat{\pi}(t)$, which is an invariant distribution of $\widehat{M}(t)$.

Another consequence of the time reversal is that the work and heat substeps are interchanged in the reverse chain. The heat, total work, and dissipative work are all odd under a time reversal: $\mathcal{Q}[\mathbf{x}] = -\mathcal{Q}[\widehat{\mathbf{x}}]$, $\mathcal{W}[\mathbf{x}] = -\mathcal{W}[\widehat{\mathbf{x}}]$ and $\mathcal{W}_{\rm d}[\mathbf{x}] = -\mathcal{W}_{\rm d}[\widehat{\mathbf{x}}]$. The total change in energy and the free energy change are also odd under a time reversal, but to avoid ambiguity a 'Δ' always refers to a change measured along the forward process.

As an illustration, consider the following diagram of a short Markov chain.

$$\begin{aligned} x(0) &\xrightarrow{M(0)} \bullet\, x(1) \xrightarrow{M(1)} \bullet\, x(2) \xrightarrow{M(2)} \bullet\, x(3) \\ \widehat{x}(3) &\xleftarrow{\hat{M}(3)} \bullet\, \widehat{x}(2) \xleftarrow{\hat{M}(2)} \bullet\, \widehat{x}(1) \xleftarrow{\hat{M}(1)} \bullet\, \widehat{x}(0) \end{aligned}$$

The stochastic transitions are indicated by arrows which are labeled with the appropriate transition matrix. The bullets ($\bullet$) indicate the points in time when the perturbation acts on the system.

We are now in a position to prove an important symmetry for the driven system under consideration. Let $\mathcal{P}[\mathbf{x}|x(0),\mathbf{M}]$ be the probability of observing the trajectory $\mathbf{x}$, given that the system started in state $x(0)$. The probability of the corresponding reversed path is $\widehat{\mathcal{P}}[\widehat{\mathbf{x}}|\widehat{x}(0),\widehat{\mathbf{M}}]$. The ratio of these path probabilities is a simple function of the heat exchanged with the bath

$$\frac{\mathcal{P}[\mathbf{x}|x(0),\mathbf{M}]}{\widehat{\mathcal{P}}[\widehat{\mathbf{x}}|\widehat{x}(0),\widehat{\mathbf{M}}]} = \exp\Big\{-\beta\mathcal{Q}[\mathbf{x}]\Big\}. \tag{1.10}$$

At the risk of ambiguity, a system with this property will be described as microscopically reversible [35,13,16,17]. In current usage the terms "microscopically reversible" and "detailed balance" are often used interchangeably [46]. However, the original meaning of microscopic reversibility [47,48] is similar to Eq. (1.10). It relates the probability of a particular path to its reverse. This is distinct from the principle of detailed balance [45,46,49], Eq. (1.3), which refers to the probabilities of changing states without reference to a particular path.

We proceed by expanding the path probability as a product of single time step transition probabilities. This follows from the condition that the dynamics are Markovian.

$$\frac{\mathcal{P}[\mathbf{x}|x(0),\mathbf{M}]}{\widehat{\mathcal{P}}[\widehat{\mathbf{x}}|\widehat{x}(0),\widehat{\mathbf{M}}]} = \frac{\prod_{t=0}^{\tau-1} \mathrm{P}\big(x(t)\to x(t{+}1)\big)}{\prod_{t'=0}^{\tau-1} \widehat{\mathrm{P}}\big(\widehat{x}(t')\to \widehat{x}(t'{+}1)\big)}$$

For every transition in the forward chain there is a transition in the reverse chain related by the time reversal symmetry, Eq. (1.8). The path probability ratio can therefore be converted into a product of equilibrium probabilities.

$$\begin{aligned}\frac{\mathcal{P}[\mathbf{x}|x(0),\mathbf{M}]}{\widehat{\mathcal{P}}[\widehat{\mathbf{x}}|\widehat{x}(0),\widehat{\mathbf{M}}]} &= \prod_{t=0}^{\tau-1}\frac{\pi(t)_{x(t+1)}}{\pi(t)_{x(t)}} = \prod_{t=0}^{\tau-1}\frac{\rho(x(t{+}1)|\beta,\mathbf{E}(t))}{\rho(x(t)|\beta,\mathbf{E}(t))}\\ &= \exp\Big\{-\beta\sum_{t=0}^{\tau-1}\big[E(t)_{x(t+1)} - E(t)_{x(t)}\big]\Big\}\\ &= \exp\big\{-\beta\mathcal{Q}[\mathbf{x}]\big\}\end{aligned}$$

The second line follows from the definition of the canonical ensemble, Eq. (1.1), and the final line from the definition of the heat, Eq. (1.4).

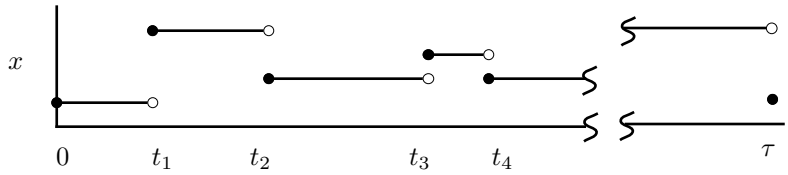
The essential assumptions leading to microscopic reversibility are that accessible state energies are always finite, and that the dynamics are Markovian, and if unperturbed preserve the equilibrium distribution. These conditions are valid independently of the strength of the perturbation, or the distance of the ensemble from equilibrium. Given that a system is microscopically reversible the derivation of a variety of far-from-equilibrium expressions is a relatively simple matter. The impatient reader may skip ahead to the next chapter without significant loss, as the remainder of this chapter will be devoted to extensions and generalizations.

1.3 Continuous-time Markov chains

Consider a Markov process with finite set of states, $x \in \{1, 2, 3, \cdots, N\}$, and a continuous time scale, $t \in [0, \tau)$ [39]. We assume that the process is non-explosive (there are only a finite number of state transitions in a finite time) and that the process is right-continuous, which in this context means that for all $0 \leqslant t < \tau$ there exists $\epsilon > 0$ such that

$$x(s) = x(t) \text{ for } t \leqslant s \leqslant t + \epsilon.$$

Therefore, during the time interval $[0, \tau)$ the system makes a finite number of transitions between states. These transitions occur instantaneously, after which the system resides in the same state for a finite amount of time. These transitions occur at the jump times, $t_1, t_2, \cdots, t_J$. We set $t_0 = 0$ and $t_J = \tau$, and require that the jump times are ordered; $t_{j-1} < t_j < t_{j+1}$. The following diagram should clarify these definitions.



The dynamics of this process can conveniently be described by a transition rate matrix¶, or Q-matrix, $Q(t)$. The off-diagonal elements of a Q-matrix are nonnegative, and all columns sum to zero.

$$\begin{aligned} Q_{ij} &\geqslant 0 \qquad \text{for all } i \neq j \\ \sum_i Q_{ij} &= 0 \qquad \text{for all } i \end{aligned}$$

The negatives of the diagonal elements $-Q_{ii}$ are the rate of leaving state i, and the off-diagonal elements Q_{ij} give the rate of going from state j to state i. For a homogeneous Markov process the finite time transition matrix, M, can be obtained from a matrix exponential of the corresponding Q matrix.

$$M = e^{\tau Q} = \sum_{k=0}^{\infty} \frac{(\tau Q)^k}{k!}$$

Because M is a polynomial in Q it follows that Q and M always have the same invariant distributions. The elements M_{ij} are the probabilities of the system being in state i at time τ given that it was in state j at time 0. This representation therefore loses information about the path that was taken between state j and state i.

An alternative approach, one that is more useful in the current context, is to consider explicitly the probability of a trajectory such as the one illustrated above. The total probability can be broken into a product of single step transition probabilities. These in turn can be split into a holding time probability and a jump probability.

$$\mathcal{P}[\mathbf{x}|x(0), \mathbf{Q}] = \prod_{j=0}^{J-1} \mathrm{P_H}\big(x(t_j)\big)\ \mathrm{P_J}\big(\, x(t_j) \to x(t_{j+1})\,\big) \tag{1.11}$$

¶As with the transition matrix M, this Q-matrix is defined as the transpose of the Q-matrix normally used in the Markov chain literature.

The holding time probability, $\mathrm{P_H}\big(x(t_j)\big)$, is the probability of holding in state $x(t_j)$ for a time t_{j+1}-t_j, before making a jump to the next state. For a non-homogeneous chain we can write the holding time probabilities in terms of the diagonal elements of $Q(t)$,

$$\mathrm{P_H}\big(x(t_j)\big) = \exp\left\{\int_{t_j}^{t_{j+1}} Q(t')_{x(t_j)x(t_j)}\,dt'\right\}.$$

The jump probabilities, $\mathrm{P_J}\big(\,x(t_j) \to x(t_{j+1})\,\big)$, are the probabilities of making the specified transition given that some jump has occurred;

$$\mathrm{P_J}\big(\,x(t_j) \to x(t_{j+1})\,\big) = -Q(t_{j+1})_{x(t_{j+1})x(t_j)}/Q_{x(t_j)x(t_j)}\,.$$

Note that the jump form state $x(t_j)$ to $x(t_{j+1})$ occurs at time t_{j+1}, and that the jump probabilities depend only on the transition rate matrix at that time.

The time reversal of a continuous time chain [39, p. 123] is very similar to the time reversal of the discrete time chain. First the states, state energies and jump times are reordered: $\widehat{x}(t) = x(\tau\text{-}t)$, $\widehat{\mathbf{E}}(t) = \mathbf{E}(\tau\text{-}t)$ and $\hat{t}_j = \tau\text{-}t_{J\text{-}j}$ for all j. These transformations have the effect of turning the right-continuous forward process into a left continuous reverse process.

Once more we insist that an equilibrium ensemble should be unaffected by a time reversal, and that the probability of a transition in equilibrium should be the same as the probability of the reverse transition in the reverse dynamics. The time reversal of the transition rate matrix is then identical to the time reversal of the discrete time transition matrix.

$$\widehat{Q}(t) = \mathrm{diag}(\pi(\tau\text{-}t))Q(\tau\text{-}t)^T\mathrm{diag}(\pi(\tau\text{-}t))^{-1}$$

Recall that $\pi(t)$ is the invariant distribution of $Q(t)$, which is identical to the probability distribution defined by equilibrium statistical mechanics, $\pi(t) = \rho(x|\beta, \mathbf{E}(t))$.

This transformation does not alter the diagonal elements of Q. It follows that the holding time probabilities of the reverse path are the same as those in the forward path. The jump probabilities of the forward path can be simply related to the jump probabilities of the reverse path via the time reversal operation. With these observations and definitions it is now possible to show that continuous time Markovian dynamics is also microscopically reversible.

$$\begin{aligned}
\widehat{\mathcal{P}}[\widehat{\mathbf{x}}|\widehat{x}(0), \widehat{\mathbf{Q}}] &= \prod_{j=0}^{J\text{-}1} \mathrm{P_H}\big(\widehat{x}(\hat{t}_{j+1})\big)\ \mathrm{P_J}\big(\,\widehat{x}(\hat{t}_j) \to \widehat{x}(\hat{t}_{j+1})\,\big) \\
&= \prod_{j=0}^{J\text{-}1} \mathrm{P_H}\big(x(t_j)\big)\ \mathrm{P_J}\big(\,x(t_j) \to x(t_{j+1})\,\big)\frac{\rho(x(t_j)|\beta, \mathbf{E}(t_j))}{\rho(x(t_{j+1})|\beta, \mathbf{E}(t_j))}
\end{aligned}$$

$$= \mathcal{P}[\mathbf{x}|x(0),\mathbf{Q}]\exp\left\{+\beta\mathcal{Q}[\mathbf{x}]\right\}$$

1.4 Continuous-time and space Markov processes

There are many advantages to working with a finite, discrete state space. It allows exploitation of the large Markov chain literature, rigorous derivations of results can be relatively easy, and it corresponds directly to simulations on classical computers. However, most physical systems studied in statistical mechanics have a continuous phase space. Typically, such systems consist of a large number of interacting particles and the state of the system can be described by an element of $\mathbf{R}^n$, where n is the number of degrees of freedom of the system. In the following section we will discuss the necessary conditions for microscopic reversibility in these systems. However, the reader should be aware that some measure theoretic subtleties are being ignored. For example, the state of the system should, strictly speaking, be represented by an element of a Borel set, rather than an arbitrary element of $\mathbf{R}^n$.

Let $x(t)$ be the state of the system at time t. The phase space probability density is $\rho(x,t)$. The time evolution of the probability density can be described by an operator , U [40, 50, 51].

$$\rho(x,t_2) = U(t_2,t_1)\rho(x,t_1)$$

The family of evolution operators have the following properties:

$$\begin{aligned}
\lim_{t_2\to t_1} U(t_2,t_1) &= I & t_2 &\geqslant t_1\\
U(t_2,t_1) &= U(t_3,t_2)U(t_2,t_1) & t_3 &\geqslant t_2 \geqslant t_1\\
\lim_{t_3\to t_2} U(t_3,t_1) &= U(t_2,t_1) & t_2 &\geqslant t_1
\end{aligned}$$

Here, I is the identity operator. The operators U can be written as integral operators of the form

$$\begin{aligned}
\rho(x,t_2) &= U(t_2,t_1)\rho(x,t_1)\\
&= \int \mathrm{P}(x,t_2|x',t_1)\,\rho(x',t_1)\,dx' \qquad t_2 \geqslant t_1\,.
\end{aligned}$$

For a Markovian process the propagators (or integral kernels) $\mathrm{P}(x,t_2|x',t_1)$ are the transition probability distribution densities. They

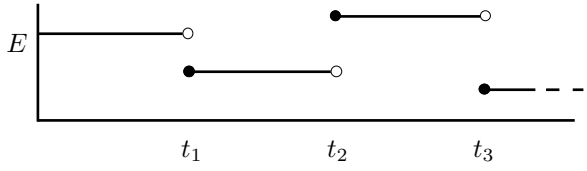
have the following properties:

$$\begin{aligned}\lim_{t_2\to t_1} \mathrm{P}(x,t_2|x',t_1) &= \delta(x-x')\\ \mathrm{P}(x,t_3|x'',t_1) &= \int \mathrm{P}(x,t_3|x',t_2)\mathrm{P}(x'',t_2|x',t_1)dx'\end{aligned}$$

We assume that statistical mechanics correctly describes the equilibrium behavior of the system. It follows that for an unperturbed system the propagator must preserve the appropriate equilibrium distribution.

$$\rho(x|\beta,\mathbf{E}) = \int \mathrm{P}(x,t|x',t')\rho(x'|\beta,\mathbf{E})dx'$$

The energy of the system is a function of both time and space, $E(x,t)$. The perturbation of the system is assumed to occur at a discrete set of times. Thus the energy of each state is constant for a finite length of time before making an instantaneous jump to a new value. This is in contrast to a continuous time, discrete space Markov chain, where the energies could change continuously but the state transitions were discrete. The jump times are labeled $t_0, t_1, \cdots, t_J$ such that $t_j < t_{j+1}$. $t_0 = 0$ and $t_J = \tau$.



Note that the energies are right continuous; $E(x',s) = E(x',t_j)$ for $t_j \leqslant s < t_{j+1}$.

The probability density of the path $x(t)$, given the set of evolution operators $\mathbf{U}$ is

$$\mathcal{P}[x(t)|\mathbf{U}] = \prod_{j=0}^{J\text{-}1} \mathrm{P}(x,t_{j+1}|x,t_j)$$

The time reversal of this dynamics can be defined in an analogues way to that used for a Markov chain. An equilibrium ensemble should be unaffected by a time reversal, and the probability of a transition should be the same as the probability of the reverse transition in the reverse dynamics.

$$\rho(\widehat{x},\tau\text{-}t|\beta,\mathbf{E})\,\widehat{\mathrm{P}}(\widehat{x},\tau\text{-}t'|\widehat{x},\tau\text{-}t) = \rho(x,t|\beta,\mathbf{E})\mathrm{P}(x,t|x,t')$$

For the forward process the equilibrium probabilities are determined by the state energy at the beginning of an interval between transitions since the energies are right continuous. In the reverse process the order of the energy transitions is reversed, $\widehat{E}(x,\tau\text{-}t) = E(x,t)$, and the state energies

become left continuous. Thus the stationary distributions of the reverse process are determined by the state energies at the end of an interval.

With these subtleties in mind the derivation of microscopic reversibility follows as it did for Markov chains. We start with the probability of the forward path and apply the time reversal operation to the constant state energy trajectory segments.

$$\begin{aligned}\mathcal{P}[x(t)|x(0),\mathbf{U}] &= \prod_{j=0}^{J-1} \mathrm{P}(x,t_{j+1}|x,t_j) \\ &= \prod_{j=0}^{J-1} \mathrm{P}(\widehat{x},t_{j+1}|\widehat{x},t_j)\frac{\rho(x(t_j)|\beta,\mathbf{E}(t_j))}{\rho(x(t_{j+1})|\beta,\mathbf{E}(t_j))} \\ &= \widehat{\mathcal{P}}[\widehat{x}(t)|\widehat{x}(0),\widehat{\mathbf{U}}]\exp\Big\{-\beta\mathcal{Q}[\mathbf{x}]\Big\}\end{aligned}$$

In this derivation, it was assumed that the state energies only changed in discrete steps. The generalization to continuously varying energies can be accomplished by letting the time between steps go to zero. However, we need the additional assumption that the energies are piecewise continuous.

1.5 Multiple baths with variable temperatures

So far we have only considered systems coupled to a single isothermal heat bath. However, it is often useful to consider nonequilibrium systems coupled to multiple heat baths of different temperatures (e.g., heat engines [52]) or to a bath whose temperature changes with time (e.g., simulated annealing [53]). The ensembles of equilibrium statistical mechanics suggest coupling systems to more general types of reservoirs with different intensive variables, either constant pressure baths (isothermal-isobaric ensemble) or constant chemical potential baths (grand canonical ensemble). Other, more exotic baths are possible, at least in principle.

The microscopic reversibility symmetry that has been derived for a single isothermal bath can be extended to these more general situations in a very straightforward manner. First note that if an amount of energy $\mathcal{Q}$ flows out of a bath of inverse temperature β then the entropy change of the bath is $\Delta S_{\mathrm{bath}} = -\beta\mathcal{Q}$. This follows because the bath is a large thermodynamic system in equilibrium. Accordingly, we can rewrite microscopic reversibility as

$$\frac{\mathcal{P}[\mathbf{x}|x(0),\mathbf{U}]}{\widehat{\mathcal{P}}[\widehat{\mathbf{x}}|\widehat{x}(0),\widehat{\mathbf{U}}]} = \exp\Big\{\Delta S_{\mathrm{baths}}\Big\}\,. \tag{1.12}$$

This expression suffices to cover all the situations discussed above if we interpret $\Delta S_{\mathrm{baths}}$ as the total change in entropy of all the baths attached

to the system. First, consider the situation when the system is coupled to an extra bath, which acts as a reservoir of an extensive quantity other than energy. As an example, consider a classical gas confined in a diathermal cylinder by a movable piston. We suppose that the piston is attached to a large spring which exerts a constant force that tries to compress the gas. The far side of the piston therefore acts as an isobaric volume reservoir.

As before, we can consider that this perturbation occurs in a number of discrete steps. Once again, we can derive the relation between the probability of a path segment between steps, and the probability of the corresponding time reversed path segment from the equilibrium distribution, $\rho(x) \propto \exp\{-\beta E_x + \beta p V\}$;

$$\frac{\mathrm{P}(x,t|x,t')}{\widehat{\mathrm{P}}(\widehat{x},\tau\text{-}t'|\widehat{x},\tau\text{-}t)} = \frac{\rho(x,t\ |\beta,\beta p,\mathbf{E})}{\rho(\widehat{x},t'|\beta,\beta p,\mathbf{E})} = \exp\bigl\{-\beta\mathcal{Q} - \beta p\Delta V\bigr\}. \tag{1.13}$$

This probability ratio continues to hold when many path segments are strung together to form the complete path. In this expression ΔV is the change in volume of the system measured along the forward path, and p is the pressure of the external isobaric volume reservoir. This change in volume can be considered a baric equivalent of the thermal heat [36]. Therefore, $-\beta p\Delta V$ is the change in entropy of the isobaric bath and microscopic reversibility can indeed be written as the total change in entropy of all the attached baths. Other isothermal ensembles of statistical mechanics can be treated in a similar manner.

This expression is not directly applicable to the isobaric ensemble as normally implemented in a computer simulation [54]. In the example above, the volume change was accomplished by a piston. In a computer simulation a volume change is normally implemented by uniformly shrinking or swelling the entire system. (This is consistent with periodic boundary conditions.) However, microscopic reversibility can be recovered by adding an additional term, $-N\ln\Delta V$, to the change in entropy of the baths. This term can be interpreted as the change in configurational entropy of an N particle system that is uniformly expanded or shrunk by ΔV.

Systems that are coupled to multiple heat baths can be treated in a similar manner. However, these systems are inherently out of equilibrium, and therefore there is no simple expression for the stationary state probability distribution. This problem can be circumvented by insisting that the connections of the heat baths to the system are clearly segregated, so that the system only exchanges energy with one bath during any time step. One possibility for implementing this restriction is allow the coupling to change at each time step, so that only one bath is coupled at a time. (This is equivalent to rapidly changing the temperature of the bath. See below.) Another possibility is to couple separate parts of the system to different heat baths. For example in an Ising model evolving according to single

spin flip dynamics each spin could be coupled to its own heat bath [55].

With this restriction to unambiguous coupling we can treat microscopic reversibility as before. The time reversed dynamics can be found by considering the equilibrium ratio of the two states that the system transitioned between at the temperature of the heat bath with which the system exchanged energy during that transition.

Finally, consider the situation where the temperature of the heat bath changes with time. We require that this temperature change occurs in a number of discrete steps which occur at the same time points that the perturbations of the system occur. Indeed, it is possible to view this change as an additional perturbation that exerts a sort of "entropic work" on the system. During every stochastic step the temperature is constant, and therefore the forward/reverse path probability ratios are given in terms of the temperature of the bath at that time step. Therefore, the change in entropy of the bath that is featured in the microscopic reversibility above should be interpreted as the change in entropy due to the exchange of energy with the bath, and not as the change in entropy of the bath due to the change in temperature.

It should also be noted that Eq. (1.12) is also applicable to microcanonical ensembles, systems not coupled to any baths. Then $\Delta S_{\text{baths}} = 0$ and the probabilities of the forward and corresponding reverse path are identical.

1.6 Langevin dynamics

We have considered microscopic reversibility only in generality. However, it is useful to consider specific examples. Recently Jarzynski [17] exhibited a derivation of microscopic reversibility for Hamiltonian systems coupled to multiple heat baths. In this section I would like to consider a specific stochastic dynamics that has been widely used to model nonequilibrium systems.

The state of a system evolving according to Langevin's equations of motion [56, 57] is determined by the positions, q, and momenta, p, of a collection of particles. Langevin dynamics can then be written as

$$\dot{\mathbf{q}} = \frac{\mathbf{p}}{\mathbf{m}}, \quad \dot{\mathbf{p}} = \mathbf{F}(\mathbf{q}, \lambda) - \gamma \mathbf{p} + \xi(t)\,. \tag{1.14}$$

Here, m is the mass of the particle, $\mathbf{F}$ is the position dependent force (parameterized by the potentially time dependent variable λ), γp is a frictional drag force proportional to the velocity, and $\xi(t)$ is a rapidly varying random force due to the coupling of the system to the many degrees of freedom of the environment. Normally, ξ is taken to be δ correlated Gaussian white

noise with zero mean,.

$$\langle \xi(t) \rangle = 0, \qquad \langle \xi(t)\xi(0) \rangle = \frac{2m\gamma}{\beta}\delta t$$

If the frictional force is large (the high friction limit) then the inertial parts of the dynamics become irrelevant, and Langevin's equation simplifies to

$$\dot{x} = \beta D\, F(x,\lambda) + \xi(t),$$

where x is a generic phase coordinate vector ($q \equiv x$), and D is the diffusion coefficient, which is related to γ by Einstein's' relation $D = m\gamma/\beta$. Another common parameterization is to set $\epsilon = 2D = 2/\beta$. Then $\dot{x} = F(x,\lambda) + \xi(t)$. We will use this convention in Chapter 6.

Here we will explicitly demonstrate microscopic reversibility for Langevin dynamics in the high friction limit. We require that the non-random force on the system is the gradient of a potential energy $F(x,\lambda) = -\partial E(x,\lambda)/\partial x$. (This condition is not strictly necessary, but without it the interpretation of microscopic reversibility in terms of the heat flow becomes problematic.) As the kinetic energy is zero this is the total energy of the system. The total change in internal energy, ΔE, can split into two parts.

$$\Delta E = \int dE = \underbrace{\int \left(\frac{\partial E}{\partial \lambda}\right)_x \frac{d\lambda}{dt}\, dt}_{\text{Work, } \mathcal{W}} + \underbrace{\int \left(\frac{\partial E}{\partial x}\right)_\lambda \frac{dx}{dt}\, dt}_{\text{Heat, } \mathcal{Q}} \tag{1.15}$$

The first term on the right is the work, the change in internal energy due to a change in the control parameter, λ. By the first law of thermodynamics the second term is the heat.

The probability of observing a trajectory through phase space $x(t)$, given $x(0)$, is given by [58, 59]

$$\mathcal{P}[x(+t)|\lambda(+t)] \propto \exp\left\{ -\int_{-\tau}^{+\tau} \frac{\left(\dot{x} + \beta D F(x,\lambda)\right)^2}{4D} + D\nabla\cdot F\, dt \right\}.$$

For convenience we have taken the time interval to be $t \in [-\tau, +\tau)$. By making the change of variables $t \to -t$ (the Jacobian is unity) we find the probability of the time reversed path.

$$\mathcal{P}[x(-t)|\lambda(-t)] \propto \exp\left\{ -\int_{-\tau}^{+\tau} \frac{(-\dot{x} + \beta D F(x,\lambda))^2}{4D} + D\nabla\cdot F\, dt \right\}.$$

The probability ratio for these two trajectories is

$$\frac{\mathcal{P}[x(+t)|\lambda(+t)]}{\mathcal{P}[x(-t)|\lambda(-t)]} = \exp\left\{ -\beta \int_{-\tau}^{+\tau} \dot{x} F(x,\lambda)\, dt \right\}. \tag{1.16}$$

The integral is the heat as defined by Eq. (1.15), thus demonstrating microscopic reversibility for the dynamics of a system evolving according to the high friction Langevin equation in a gradient force field. It is also possible to show that systems evolving according to the Langevin equation in a time dependent field are microscopically reversible by showing that the dynamics are detailed balanced for infinitesimal time steps [24].

The Langevin equation has another curious and suggestive property, one that may imply useful relations that would not hold true for a general stochastic dynamics. The probability of a Langevin trajectory (high or low friction) can be written as [60, 61, 24]

$$\mathcal{P}[\mathbf{x}, x(0)] = \frac{e^{-\beta U_p[\mathbf{x}]}}{\mathcal{Z}}\,, \tag{1.17}$$

where $U_p[\mathbf{x}]$ is a temperature independent "path energy", and $\mathcal{Z}$ is a path independent normalization constant. The similarity to equilibrium statistical mechanics is striking. Because we have shown that Langevin trajectories are microscopically reversible it follows that

$$U_p[\mathbf{x}] - U_p[\widehat{\mathbf{x}}] = \beta \mathcal{Q}[\mathbf{x}]\,.$$

It is possible to impose this path energy structure on a discrete time, discrete space dynamics. However, the necessary conditions appear unrealistic. Suppose that the single time step transition probabilities could be written in the form of Eq. (1.17).

$$\mathrm{P}(i \to i') = \frac{e^{-\beta U_p(i,i')}}{\mathcal{Z}} \qquad \text{where} \qquad \mathcal{Z} = \sum_{i'} e^{-\beta U_p(i,i')}\,.$$

The normalization constant $\mathcal{Z}$ must be path independent, and therefore independent of i. This can only be true for arbitrary β if the set of path energies is the same for each and every initial state. This implies that

$$P(i \to i') = P(i \to j') \text{ for all } i, j \text{ and some } i', j'\,.$$

For a discrete system this condition seems overly restrictive. It would be very difficult (if not impossible) to implement a dynamics with this property for the discrete lattice models, such as the Ising model, that are so popular in equilibrium statistical mechanics. Nonetheless, this assumption that the dynamics of a discrete lattice system can be written in this form (occasionally referred to as a Gibbs measure [62]) has been used to derive a version of the Gallavotti-Cohen fluctuation theorem. [12, 16, and see Chapter 3] As we shall see this theorem can derived from the much more general condition that the dynamics are microscopically reversible.

1.7 Summary

If a system is coupled to a collection of thermodynamic baths, and the dynamics of the system are Markovian and balanced, and the energy of the system is always finite, then the dynamics are microscopically reversible.

$$\boxed{\frac{\mathcal{P}[\mathbf{x}|x(0)]}{\widehat{\mathcal{P}}[\widehat{\mathbf{x}}|\widehat{x}(0)]} = \exp\Big\{\Delta S_{\text{baths}}\Big\}} \tag{1.18}$$

Here, $\mathcal{P}[\mathbf{x}|x(0)]$ is the probability of the trajectory $\mathbf{x}$, given that the system started in state $x(0)$, $\widehat{\mathcal{P}}[\widehat{\mathbf{x}}|\widehat{x}(0)]$ is the probability of the time reversed trajectory, with the time reversed dynamics, and ΔS_{baths} is the total change in entropy of the baths attached to the system, measured along the forward trajectory.

For the special, but common, case of a system coupled to a single isothermal bath this expression becomes

$$\boxed{\frac{\mathcal{P}[\mathbf{x}|x(0)]}{\widehat{\mathcal{P}}[\widehat{\mathbf{x}}|\widehat{x}(0)]} = \exp\Big\{+\beta\mathcal{Q}[\mathbf{x}]\Big\}} \tag{1.19}$$

where $\mathcal{Q}$ is the heat exchanged with the bath along the forward trajectory.

2. Path Ensemble Averages

One of the principle objects of theoretical research in any department of knowledge is to find the point of view from which the subject appears in its greatest simplicity.

— J. W. Gibbs

2.1 Introduction

If a system is gently driven from equilibrium by a small time-dependent perturbation, then the response of the system to the perturbation can be described by linear response theory. On the other hand, if the system is driven far from equilibrium by a large perturbation then linear response, and other near-equilibrium approximations, are generally not applicable. However, there are a few relations that describe the statistical dynamics of driven systems which are valid even if the system is driven far from equilibrium. These include the Jarzynski nonequilibrium work relation [34, 63, 64, 35], which gives equilibrium free energy differences in terms of nonequilibrium measurements of the work required to switch from one ensemble to another; the Kawasaki relation [65, 66, 67, 68, 69], which specifies the nonlinear response of a classical system to an arbitrarily large perturbation; and the Evans-Searles transient fluctuation theorem [2, 5] which provides information about the entropy production of driven systems that are initially in equilibrium. This last theorem is one of a group that can be collectively called the "entropy production fluctuation theorems" [1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17], which will be considered in greater detail in Chapter 3. (In particular, the Gallavotti-Cohen [3, 4] differs from the Evans-Searles identity is several important respects, and will not be considered in this chapter.)

The relations listed above have been derived for a wide range of deterministic and stochastic dynamics. However, the various expressions and applicable dynamics have several commonalities: the system starts in thermal equilibrium, it is driven from that equilibrium by an external perturbation, the energy of the system always remains finite, the dynamics are Markovian, and if the system is unperturbed then the dynamics preserve the equilibrium ensemble. It should be clear from the previous chapter that a system with these properties is microscopically reversible (1.10). It will

be shown that this property is a sufficient condition for the validity of all the far-from-equilibrium expressions mentioned above. Indeed, they can all be considered special cases of a single theorem:

$$\langle \mathcal{F} \rangle_{\mathrm{F}} = \langle \widehat{\mathcal{F}}\, e^{-\beta \mathcal{W}_{\mathrm{d}}} \rangle_{\mathrm{R}}. \tag{2.1}$$

Here, $\langle \mathcal{F} \rangle_{\mathrm{F}}$ indicates the average of the path function $\mathcal{F}$. Path functions (such as the heat and work) are functionals of the trajectory that the system takes through phase-space. An average of a path function is implicitly an average over a suitably defined ensemble of paths. In this chapter, the path ensemble is defined by the initial thermal equilibrium and the process by which the system is subsequently perturbed from that equilibrium. The left side of the above relation is simply $\mathcal{F}$ averaged over the ensemble of paths generated by this process. We arbitrarily label this the forward process (subscript 'F').

For every such process that perturbs the system from equilibrium we can imagine a corresponding reverse perturbation (subscript 'R'). We shall construct this process by insisting that it too starts from equilibrium, and by considering a formal time reversal of the dynamics. The right side of Eq. (2.1) is $\widehat{\mathcal{F}}$, the time reversal of $\mathcal{F}$, averaged over the reverse process, and weighted by the exponential of $\beta \mathcal{W}_{\mathrm{d}}$. Here $\beta = 1/k_{\mathrm{B}}T$, T is the temperature of the heat bath, k_{B} is Boltzmann's constant, and $\mathcal{W}_{\mathrm{d}}$ is the dissipative work. The dissipative work is a path function and is defined as $\mathcal{W}_{\mathrm{d}} = \mathcal{W} - W_{\mathrm{r}}$, where $\mathcal{W}$ is the total work done on the system by the external perturbation and W_{r} is the reversible work, the minimum average amount of work required to perturb the system from its initial to its final ensemble.

In summary, Eq. (2.1) states that an average taken over an ensemble of paths, which is generated by perturbing a system that is initially in equilibrium, can be equated with the average of another, closely related quantity, averaged over a path ensemble generated by the reverse process. This relation is valid for systems driven arbitrarily far from equilibrium, and several other far-from-equilibrium relations can be derived from it. In the next section I show that the relation between path ensemble averages (2.1) is an almost trivial identity given that the dynamics are microscopically reversible. A variety of special cases are introduced, many of which will be dealt with in much greater detail in subsequent chapters.

2.2 Path ensemble averages

Consider the following situation; A system that is initially in thermal equilibrium is driven away from that equilibrium by an external perturbation, and the path function $\mathcal{F}[\mathbf{x}]$ is averaged over the resulting nonequilibrium ensemble of paths. The probability of a trajectory is determined by

the equilibrium probability of the initial state, and by the vector of transition matrices that determine the dynamics. Therefore, the average of $\mathcal{F}$ over the ensemble of trajectories can be explicitly written as

$$\langle \mathcal{F} \rangle_{\mathrm{F}} = \sum_{\mathbf{x}} \rho\big(x(0)\big|\beta, \mathbf{E}(0)\big)\, \mathcal{P}[\,\mathbf{x}\,|\,x(0), \mathbf{M}]\, \mathcal{F}[\mathbf{x}]\,.$$

The sum is over the set of all paths connecting all possible initial and final states.

$$\sum_{\mathbf{x}} = \sum_{x(0)} \sum_{x(1)} \sum_{x(2)} \cdots \sum_{x(\tau)} \tag{2.2}$$

If the system has a continuous phase space then the above sum can be written as the path integral

$$\langle \mathcal{F} \rangle_{\mathrm{F}} = \iiint_{x(0)}^{x(\tau)} \rho\big(x(0)\big|\beta, \mathbf{E}(0)\big)\, \mathcal{P}[\,\mathbf{x}\,|\,x(0), \mathbf{U}]\, \mathcal{F}[\mathbf{x}]\, \mathcal{D}[x] dx(0) dx(\tau)\,. \tag{2.3}$$

The additional integrals are over all initial and final states, since the standard path integral has fixed boundary points. For simplicity, only the discrete system will be explicitly considered. However, the derivation immediately generalizes to continuous systems.

Given that the system is microscopically reversible it is a simple matter to convert the above expression to an average over the reverse process. We first note that

$$\begin{aligned} \frac{\rho\big(x(0)\big|\beta, \mathbf{E}(0)\big)\, \mathcal{P}[\mathbf{x}|x(0), \mathbf{M}]}{\rho\big(\widehat{x}(0)\big|\beta, \widehat{\mathbf{E}}(0)\big)\, \widehat{\mathcal{P}}[\widehat{\mathbf{x}}|\widehat{x}(0), \widehat{\mathbf{M}}]} &= e^{+\beta\Delta E - \beta\Delta F - \beta\mathcal{Q}[\mathbf{x}]}, \\ &= e^{+\beta\mathcal{W}[\mathbf{x}] - \beta\Delta F}, \\ &= e^{+\beta\mathcal{W}_{\mathrm{d}}[\mathbf{x}]}. \end{aligned} \tag{2.4}$$

The first line follows from the condition of microscopic reversibility (1.10), and the definition of the canonical ensemble (1.1). Recall that $\Delta E = E(\tau)_{x(\tau)} - E(0)_{x(0)}$ (1.6) is the change in energy of the system measured along the forward trajectory, that ΔF is the reversible work of the forward process, and that $\mathcal{W}_{\mathrm{d}}[\mathbf{x}]$ is the dissipative work. The set of reverse trajectories is the same as the set of forward trajectories, and the time reversal of the path function is $\mathcal{F}[\mathbf{x}] = \widehat{\mathcal{F}}[\widehat{\mathbf{x}}]$.

Therefore,

$$\begin{aligned} \langle \mathcal{F} \rangle_{\mathrm{F}} &= \sum_{\widehat{\mathbf{x}}} \rho\big(\widehat{x}(0)\big|\beta, \widehat{\mathbf{E}}(0)\big)\, \widehat{\mathcal{P}}[\,\widehat{\mathbf{x}}\,|\,\widehat{\mathrm{M}}\,]\, \widehat{\mathcal{F}}[\widehat{\mathbf{x}}]\; e^{-\beta\mathcal{W}_{\mathrm{d}}[\widehat{\mathbf{x}}]} \\ &= \big\langle \widehat{\mathcal{F}}\, e^{-\beta\mathcal{W}_{\mathrm{d}}} \big\rangle_{\mathrm{R}}\,. \end{aligned}$$

It is frequently convenient to rewrite Eq. (2.1) as

$$\left\langle \mathcal{F}\, e^{-\beta \mathcal{W}_{\rm d}} \right\rangle_{\rm F} = \left\langle \widehat{\mathcal{F}} \right\rangle_{\rm R}, \tag{2.5}$$

where $\mathcal{F}$ has been replaced with $\mathcal{F}\, e^{-\beta \mathcal{W}_{\rm d}}$, and $\widehat{\mathcal{F}}$ with $\widehat{\mathcal{F}}\, e^{+\beta \mathcal{W}_{\rm d}}$.

2.2.1 Jarzynski nonequilibrium work relations

A variety of previously known relations can be considered special cases of, or approximations to this nonequilibrium path ensemble average. In the simplest case we start with Eq. (2.5), and then set $\mathcal{F} = \widehat{\mathcal{F}} = 1$ (or any other constant of the dynamics). Then

$$\left\langle e^{-\beta \mathcal{W}_{\rm d}} \right\rangle_{\rm F} = \left\langle 1 \right\rangle_{\rm R} = 1. \tag{2.6}$$

The right side is unity due to normalization of probability distributions. This relation contains only one path ensemble average. Therefore, it is no longer necessary to distinguish between forward and reverse perturbations and the remaining subscript, "F", is superfluous. The dissipative work, $\mathcal{W}_{\rm d}$ can replaced by $\mathcal{W} - \Delta F$, and the change in free energy can be moved outside the average since it is path independent. The result is the Jarzynski nonequilibrium work relation [34, 63, 70, 64, 35, 13].

$$\langle e^{-\beta \mathcal{W}} \rangle = e^{-\beta \Delta F} \tag{2.7}$$

This relation states that if we convert one system into another by changing the energies of all the states from an initial set of values to a final set of values over some finite length of time, then the change in the free energies of the corresponding equilibrium ensembles can be calculated by repeating the switching process many times, each time starting with an initial state drawn from an equilibrium ensemble, and taking the above average of the amount of work required to effect the change. In the limit of instantaneous switching between ensembles, (we change the energies of all the states in a single instantaneous jump) this relation is equivalent to the standard thermodynamic perturbation method that is frequently used to calculate free energy differences by computer simulation [71].

It is possible to extend Eq. (2.7) to a more general class of relations between the work and the free energy change [72]. Suppose that $\mathcal{F} = f(\mathcal{W})$ where $f(\mathcal{W})$ is any finite function of the work. Then $\widehat{\mathcal{F}} = f(\text{-}\mathcal{W})$, because the work is odd under a time reversal. Inserting these definitions into Eq. (2.1) and rearranging gives

$$e^{-\beta \Delta F} = \frac{\left\langle f(+\mathcal{W}) \right\rangle_{\rm F}}{\left\langle f(\text{-}\mathcal{W})\, e^{-\beta \mathcal{W}} \right\rangle_{\rm R}} \tag{2.8}$$

Recall that ΔF is defined in terms of the forward process. Further applications and approximations of these nonequilibrium work relations can be found in Chapter 4.

2.2.2 Transient fluctuation theorem

Another interesting application of the path ensemble average is to consider $\mathcal{F} = \delta(\beta\mathcal{W}_{\rm d}{-}\beta\mathcal{W}_{\rm d}[\mathbf{x}])$, $\widehat{\mathcal{F}} = \delta(\beta\mathcal{W}_{\rm d}{+}\beta\mathcal{W}_{\rm d}[\widehat{\mathbf{x}}])$. Substituting into Eq. (2.1) gives

$$\begin{aligned} \left\langle \delta(\beta\mathcal{W}_{\rm d}{-}\beta\mathcal{W}_{\rm d}[\mathbf{x}])\; e^{-\beta\mathcal{W}_{\rm d}} \right\rangle_{\rm F} &= \left\langle \delta(\beta\mathcal{W}_{\rm d}{+}\beta\mathcal{W}_{\rm d}[\widehat{\mathbf{x}}]) \right\rangle_{\rm R} , \\ \text{or} \qquad \mathrm{P}_{\rm F}(+\beta\mathcal{W}_{\rm d})\, e^{-\beta\mathcal{W}_{\rm d}} &= \mathrm{P}_{\rm R}(-\beta\mathcal{W}_{\rm d}) . \end{aligned}$$

Here, $\mathrm{P}_{\rm F}(+\beta\mathcal{W}_{\rm d})$ is the probability of expending the specified amount of dissipative work in the forward process, and $\mathrm{P}_{\rm R}(-\beta\mathcal{W}_{\rm d})$ is the probability of expending the negative of that amount of work in the reverse process. If $\mathrm{P}_{\rm R}(-\beta\mathcal{W}_{\rm d}) \neq 0$ then we can rearrange this expression as

$$\frac{\mathrm{P}_{\rm F}(+\beta\mathcal{W}_{\rm d})}{\mathrm{P}_{\rm R}(-\beta\mathcal{W}_{\rm d})} = e^{+\beta\mathcal{W}_{\rm d}}. \tag{2.9}$$

Equation (2.9) can be considered as follows. The system of interest starts in equilibrium and is perturbed for a finite amount of time. If it is allowed to relax back to equilibrium then the change in entropy of the heat bath will be $-\beta\mathcal{Q}$, and the change in entropy of the system will be $\beta\Delta E{-}\beta\Delta F$. Therefore, the total change in entropy of the universe resulting from the perturbation of the system is $-\beta\mathcal{Q}{+}\beta\Delta E{-}\beta\Delta F = \beta\mathcal{W}{-}\beta\Delta F = \beta\mathcal{W}_{\rm d}$, the dissipative work. Thus Eq. (2.9) can be interpreted as an entropy production fluctuation theorem. It relates the distribution of entropy productions of a driven system that is initially in equilibrium to the entropy production of the same system driven in reverse. As such it is closely related to the transient fluctuation theorems of Evans and Searles [2, 5]. The connection between this fluctuation theorem, the Jarzynski nonequilibrium work relation and microscopic reversibility was originally presented in [13].

2.2.3 Kawasaki response and nonequilibrium distributions

All of the relations in Sec. 2.2.1 and Sec. 2.2.2 were derived from Eq. (2.1) by inserting a function of the work. Another group of relations can be derived by instead considering $\mathcal{F}$ to be a function of the state of the system at some time. In particular, if we average a function of the final

state in the forward process, $\mathcal{F} = f\big(x(\tau)\big)$, then we average a function of the initial state in the reverse process, $\widehat{\mathcal{F}} = f\big(\widehat{x}(0)\big)$:

$$\big\langle f\big(x(\tau)\big)\, e^{-\beta \mathcal{W}_{\rm d}} \big\rangle_{\rm F} = \big\langle f\big(\widehat{x}(0)\big) \big\rangle_{\rm R} \,.$$

Therefore, in the reverse process the average is over the initial equilibrium ensemble of the system, and the subsequent dynamics are irrelevant. We can once more drop reference to forward or reverse processes, and instead use labels to indicate equilibrium and nonequilibrium averages:

$$\big\langle f\big(x(\tau)\big)\, e^{-\beta \mathcal{W}_{\rm d}} \big\rangle_{\rm neq} = \big\langle f\big(x(\tau)\big) \big\rangle_{\rm eq} \,. \tag{2.10}$$

This relation (also due to Jarzynski [72]) states that the average of a state function in a nonequilibrium ensemble, weighted by the dissipative work, can be equated with an equilibrium average of the same quantity.

Another interesting relation results if we insert the same state functions into the alternative form of the path ensemble average, Eq. (2.5): (This is ultimately equivalent to interchanging $\mathcal{F}$ and $\widehat{\mathcal{F}}$.)

$$\big\langle f\big(x(\tau)\big) \big\rangle_{\rm F} = \big\langle f\big(\widehat{x}(0)\big)\, e^{-\beta \mathcal{W}_{\rm d}} \big\rangle_{\rm R} \,. \tag{2.11}$$

This is the Kawasaki nonlinear response relation [65,66,67,68,69], applied to stochastic dynamics, and generalized to arbitrary forcing. This relation can also be written in an explicitly renormalized form [68] by expanding the dissipative work as $-\Delta F + \mathcal{W}$, and rewriting the free energy change as a work average using the Jarzynski relation, Eq. (2.7).

$$\big\langle f\big(x(\tau)\big) \big\rangle_{\rm F} = \big\langle f\big(\widehat{x}(0)\big)\, e^{-\beta \mathcal{W}} \big\rangle_{\rm R} \Big/ \big\langle e^{-\beta \mathcal{W}} \big\rangle_{\rm R} \tag{2.12}$$

Simulation data indicates that averages calculated with the renormalized expression typically have lower statistical errors [68].

The probability distribution of a nonequilibrium ensemble can be derived from the Kawasaki relation, Eq. (2.12), by setting the state function to be $\mathcal{F} = f(x(\tau)) = \delta\big(x - x(\tau)\big)$, a δ function of the state of the system at time τ;

$$\rho_{\rm neq}\big(x, \tau | \mathbf{M}\big) = \rho\big(x \big| \beta, \mathbf{E}(\tau)\big) \frac{\big\langle e^{-\beta \mathcal{W}} \big\rangle_{{\rm R},x}}{\big\langle e^{-\beta \mathcal{W}} \big\rangle_{\rm R}}. \tag{2.13}$$

Here $\rho_{\rm neq}\big(x, \tau | \mathbf{M}\big)$ is the nonequilibrium probability distribution and $\rho\big(x \big| \beta, \mathbf{E}(\tau)\big)$ is the equilibrium probability of the same state. The subscript 'x' indicates that the average is over all paths that start in state x. In contrast, the lower average is over all paths starting from an equilibrium ensemble. Thus the nonequilibrium probability of a state is, to zeroth order, the equilibrium probability, and the correction factor can be related to a nonequilibrium average of the work.

2.3 Summary

All of the relations derived in this chapter are directly applicable to systems driven far-from-equilibrium. These relations follow if the dynamics are microscopically reversible in the sense of Eq. (1.10), and this relation was shown to hold if the dynamics are Markovian and balanced. Although I have concentrated on stochastic dynamics with discrete time and phase space, this should not be taken as a fundamental limitation. The extension to continuous phase space and time appears straightforward, and deterministic dynamics can be taken as a special case of stochastic dynamics.

In the following chapters many of these same relations will be considered in greater depth, and various useful approximations will be considered, some of which are only applicable near equilibrium.

3. The Fluctuation Theorem(s)

I have yet to see any problem, however complicated, which when you looked at it in the right way, did not become still more complicated.

— Paul Alderson*

3.1 Introduction

The fluctuation theorems [1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19] are a group of relations that describe the entropy production of a finite classical system coupled to a constant temperature heat bath, that is driven out of equilibrium by a time dependent work process. Although the type of system, range of applicability, and exact interpretation differ, these theorems have the same general form,

$$\frac{\mathrm{P}(+\sigma)}{\mathrm{P}(-\sigma)} \simeq e^{\tau\sigma}. \tag{3.1}$$

Here $\mathrm{P}(+\sigma)$ is the probability of observing an entropy production rate, σ, measured over a trajectory of time τ. Evans and Searles [2] gave a derivation for driven thermostatted deterministic systems that are initially in equilibrium (this result is referred to either as the transient fluctuation theorem, or as the Evans-Searles identity), Gallavotti and Cohen [3] rigorously derived their fluctuation theorem for thermostatted deterministic steady state ensembles, and Kurchan [9], Lebowitz and Spohn [11], and Maes [12] have considered systems with stochastic dynamics. The exact interrelation between these results is currently under debate [14,15].

In this chapter I will derive the following, somewhat generalized, version of this theorem for stochastic microscopically reversible dynamics:

$$\frac{\mathrm{P}_{\mathrm{F}}(+\omega)}{\mathrm{P}_{\mathrm{R}}(-\omega)} = e^{+\omega}. \tag{3.2}$$

Here ω is the entropy production of the driven system measured over some time interval, $\mathrm{P}_{\mathrm{F}}(\omega)$ is the probability distribution of this entropy production, and $\mathrm{P}_{\mathrm{R}}(\omega)$ is the probability distribution of the entropy production

**New Scientist* 25 Sept 1969, p. 638

when the system is driven in a time-reversed manner. This distinction is necessary because we will consider systems driven by a time-dependent process, rather than the steady perturbation considered elsewhere. The use of an entropy production, rather than an entropy production rate, will prove convenient.

As a concrete example of a system for which the above theorem is valid, consider a classical gas confined in the cylinder by a movable piston. The walls of this cylinder are diathermal so that the gas is in thermal contact with the surroundings, which therefore act as the constant temperature heat bath. The gas is initially in equilibrium with a fixed piston. The piston is then moved inwards at a uniform rate, compressing the gas to some new, smaller volume. In the corresponding time reversed process, the gas starts in equilibrium at the final volume of the forward process, and is then expanded back to the original volume at the same rate that it was compressed by the forward process. The microscopic dynamics of the system will differ for each repetition of this process, as will the entropy production, the heat transfer and the work performed on the system. The probability distribution of the entropy production is measured over the ensemble of repetitions.

In the following section we will derive the entropy production fluctuation theorem, Eq. (3.2). Then in Sec. 3.3 we will discuss two distinct groups of driven systems for which the theorem is valid. In the first group, systems start in equilibrium, and are then actively perturbed away from equilibrium for a finite amount of time. In the second group, systems are driven into a time symmetric nonequilibrium steady state. We conclude by discussing several approximations that are valid when the entropy production is measured over long time periods. The fluctuation theorem and its approximations are illustrated with data from a simple computer model.

3.2 The fluctuation theorem

In contrast to the previous chapter, we no longer restrict attention to systems that start in equilibrium. A particular work process is defined by the initial phase-space distribution, $\rho\big(x(0)\big)$, and by the time dependent dynamics[†], $\mathbf{U}$. Each individual realization of this process is characterized by the path that the system follows through phase space, $x(t)$. The entropy production, ω, must be a functional of this path. Clearly there is a change in entropy due to the exchange of energy with the bath. If $\mathcal{Q}$ is the amount of energy that flows out of the bath and into the system, then the entropy of the bath must change by $-\beta\mathcal{Q}$. There is also a change in entropy associated

[†]In order to be consistent with the bulk of the fluctuation theorem literature, the emphasis in this chapter is on systems with a continuous time dynamics and a continuous phase space.

with the change in the microscopic state of the system. From an information theoretic [20] perspective, the entropy of a microscopic state of a system, $s(x) = -\ln \rho(x)$, is the amount of information required to describe that state given that the state occurs with probability $\rho(x)$. The entropy of this (possibly nonequilibrium) *ensemble* of systems is $S = -\sum_x \rho(x) \ln \rho(x)$. Thus, for a single realization of a process that takes some initial probability distribution, $\rho(x(0))$, and changes it to some different final distribution, $\rho(x(\tau))$, the entropy production[‡] is

$$\omega = \ln \rho\big(x(\tau)\big) - \ln \rho\big(x(0)\big) - \beta \mathcal{Q}[x(t)]. \tag{3.3}$$

This is the change in the amount of information required to describe the microscopic state of the system plus the change in entropy of the bath.

Recall that the fluctuation theorem, Eq. (3.2), compares the entropy production probability distribution of a process with the entropy production distribution of the corresponding time-reversed process. For example, with the confined gas we compare the entropy production when the gas is compressed to the entropy production when the gas is expanded. To allow this comparison of forward and reverse processes we will require that the entropy production is odd under a time reversal (i.e., $\omega_{\rm F} = -\omega_{\rm R}$) for the process under consideration. This condition is equivalent to requiring that the final distribution of the forward process, $\rho(x(\tau))$, is the same (after a time reversal) as the initial phase-space distribution of the reverse process, $\rho(\widehat{x}(0))$, and vice versa, i.e., $\rho(x(\tau)) = \rho(\widehat{x}(0))$ and $\rho(x(0)) = \rho(\widehat{x}(\tau))$. In the next section, we will discuss two broad types of work process that fulfill this condition. Either the system begins in equilibrium or the system begins and ends in the same time symmetric nonequilibrium steady state.

This time-reversal symmetry of the entropy production allows the comparison of the probability of a particular path, $x(t)$, starting from some specific point in the initial distribution, with the corresponding time-reversed path,

$$\frac{\rho(x(0))\, \mathcal{P}[x(t)|x(0), \mathbf{U}]}{\rho(\widehat{x}(0))\, \widehat{\mathcal{P}}[\widehat{x}(t)|\widehat{x}(0), \widehat{\mathbf{U}}]} = e^{+\omega_{\rm F}}. \tag{3.4}$$

This follows from the conditions that the system is microscopically reversible (1.10) and that the entropy production is odd under a time reversal.

Now consider the probability, $\mathrm{P}_{\rm F}(\omega)$, of observing a particular value of this entropy production. It can be written as a δ function averaged over the ensemble of forward paths,

$$\mathrm{P}_{\rm F}(\omega) \quad = \quad \big\langle \delta(\omega - \omega_{\rm F}) \big\rangle_{\rm F}$$

[‡]This definition of the entropy production is equivalent to the quantityΣ used by Jarzynski [63]

$$= \iiint_{x(0)}^{x(\tau)} \rho(x(0))\, \mathcal{P}[x(t)|x(0), \mathbf{U}] \delta(\omega - \omega_{\rm F}) \mathcal{D}[x(t)] dx(0) dx(\tau).$$

Here $\iiint_{x(0)}^{x(\tau)} \cdots \mathcal{D}[x(t)] dx(0) dx(\tau)$ indicates a sum or suitable normalized integral over all paths through phase space, and all initial and final phase space points, over the appropriate time interval. We can now use Eq. (3.4) to convert this average over forward paths into an average over reverse paths.

$$\begin{aligned}
\mathrm{P}_{\rm F}(\omega) &= \iiint_{x(0)}^{x(\tau)} \rho(\widehat{x}(0))\, \mathcal{P}[\widehat{x}(t)|\widehat{x}(0), \widehat{\mathbf{U}}] \delta(\omega - \omega_{\rm F})\, e^{+\omega_{\rm F}}\, \mathcal{D}[x(t)] dx(0) dx(\tau) \\
&= e^{+\omega} \iiint_{\widehat{x}(0)}^{\widehat{x}(\tau)} \rho(\widehat{x}(0))\, \mathcal{P}[\widehat{x}(t)|\widehat{x}(0), \widehat{\mathbf{U}}]\, \delta(\omega + \omega_{\rm R}) \mathcal{D}[\widehat{x}(t)] d\widehat{x}(0) d\widehat{x}(\tau) \\
&= e^{+\omega} \left\langle \delta(\omega + \omega_{\rm R}) \right\rangle_{\rm R} \\
&= e^{+\omega}\, \mathrm{P}_{\rm R}(-\omega)
\end{aligned}$$

The remaining average is over reverse paths as the system is driven in reverse. The final result is the entropy production fluctuation theorem, Eq. (3.2).

The fluctuation theorem readily generalizes to other ensembles. As an example, consider an isothermal-isobaric system. In addition to the heat bath, the system is coupled to a volume bath, characterized by βp, where p is the pressure. Both baths are considered to be large, equilibrium, thermodynamic systems. Therefore, the change in entropy of the heat bath is $-\beta Q$ and the change in entropy of the volume bath is $-\beta p \Delta V$, where ΔV is the change in volume of the system. The entropy production should then be defined as

$$\omega = \ln \rho(x(0)) - \ln \rho(x(\tau)) - \beta \mathcal{Q} - \beta p \Delta V, \tag{3.5}$$

and microscopic reversibility should be defined as in Eq. (1.13). The fluctuation theorem, Eq. (3.2), follows as before. It is possible to extend the fluctuation theorem to any standard set of baths, so long as the definitions of microscopic reversibility and the entropy production are consistent. In the rest of this chapter we shall only explicitly deal with systems coupled to a single heat bath, but the results generalize directly.

3.3 Two groups of applicable systems

In this section we will discuss two groups of systems for which the entropy fluctuation theorem, Eq. (3.2), is valid. These systems must satisfy the condition that the entropy production, Eq. (3.3), is odd under a time reversal, and therefore that $\rho(x(0)) = \rho(\widehat{x}(\tau))$.

First consider a system that is in equilibrium from time $t = -\infty$ to $t = 0$. It is then driven from equilibrium by an external perturbation of the system. The system is actively perturbed up to a time $t = \tau$, and is then allowed to relax, so that it once again reaches equilibrium at $t = +\infty$. For the forward process the system starts in the equilibrium ensemble specified by $\mathbf{E}(-\infty)$, and ends in the ensemble specified by $\mathbf{E}(+\infty)$. In the reverse process, the initial and final ensembles are exchanged, and the entropy production is odd under this time reversal. The gas confined in the diathermal cylinder (page 3) satisfies these conditions if the piston moves only for a finite amount of time.

At first it may appear disadvantageous that the entropy production has been defined between equilibrium ensembles separated by an infinite amount of time. However, for these systems the entropy production has a simple and direct interpretation. The probability distributions of the initial and final ensembles are known from equilibrium statistical mechanics (1.1). If these probabilities are substituted into the definition of the entropy production, Eq. (3.3), then we find that

$$\omega_{\mathrm{F}} = -\beta\Delta F + \beta\mathcal{W} = \beta\mathcal{W}_{\mathrm{d}}. \tag{3.6}$$

Recall that $\mathcal{W}$ is the work, ΔF is the change in free energy, and $\mathcal{W}_{\mathrm{d}}$ is the dissipative work. (See page 7) It is therefore possible to express the fluctuation theorem in terms of the amount of work performed on a system that starts in equilibrium [73],

$$\frac{\mathrm{P}_{\mathrm{F}}(+\beta\mathcal{W})}{\mathrm{P}_{\mathrm{R}}(-\beta\mathcal{W})} = e^{-\Delta F} e^{+\beta\mathcal{W}} = e^{+\beta\mathcal{W}_{\mathrm{d}}}. \tag{3.7}$$

The work in this expression is measured over the finite time that the system is actively perturbed.

The validity of this expression can be illustrated with the simple computer model described in Fig. 3.1. Although not of immediate physical relevance, this model has the advantage that the entropy production distributions of this driven nonequilibrium system can be calculated exactly, apart from numerical roundoff error. The resulting work distributions are shown in Fig. 3.2. Because the process is time symmetric, $\Delta F = 0$ and $\mathrm{P}(+\beta\mathcal{W}) = \mathrm{P}(-\beta\mathcal{W}) \exp \beta\mathcal{W}$. This expression is exact for each of the distributions shown, even for short times when the distributions display very erratic behavior.

Systems that start and end in equilibrium are not the only ones that satisfy the fluctuation theorem. Consider again the classical gas confined in a diathermal cylinder. If the piston is driven in a time symmetric periodic manner (for example, the displacement of the piston is a sinusoidal function of time), the system will eventually settle into a nonequilibrium steady-state ensemble. We will now require that the dynamics are entirely diffusive, so

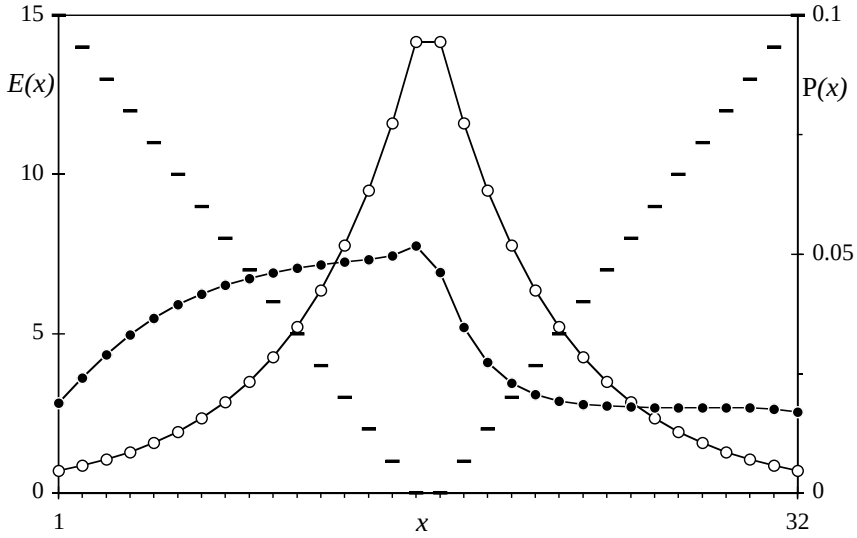


Figure 3.1: A simple Metropolis Monte Carlo simulation is used to illustrate the fluctuation theorem and some of its approximations. The master equation for this system can be solved, providing exact numerical results to compare with the theory. A single particle occupies a finite number of positions in a one-dimensional box with periodic boundaries, and is coupled to a heat bath of temperature $T = 5$. The energy surface, $E(x)$, is indicated by – in the figure. At each discrete time step the particle attempts to move left, right, or remain in the same location with equal probabilities. The move is accepted with the standard Metropolis acceptance probability [74]. Every eight time steps, the energy surface moves right one position. Thus the system is driven away from equilibrium, and eventually settles into a time symmetric nonequilibrium steady-state. The equilibrium (○) and steady state (●) probability distributions are shown in the figure above. The steady state distribution is shown in the reference frame of the surface.

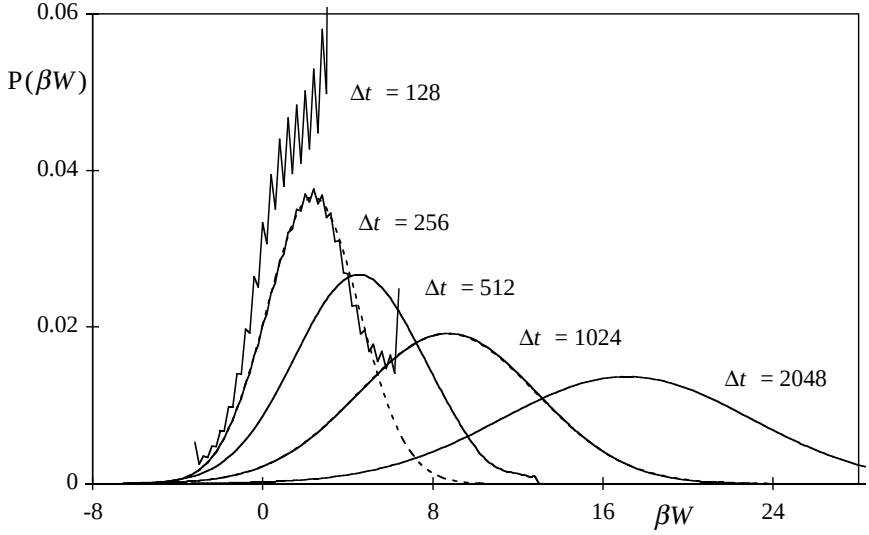


Figure 3.2: Work probability distribution (—) for the system of Fig. 3.1, starting from equilibrium. The work $\mathcal{W}$ was measured over 16, 32, 64, 128, and 256 cycles (Δt = 128, 256, 512, 1024, and 2048). For each of these distributions the work fluctuation theorem, Eq. (3.7), is exact. The dashed line (- - -) is a Gaussian fitted to the mean of the $\Delta t = 256$ distribution. (See Sec. 3.4).

that there are no momenta. Then at any time that the dynamics are time symmetric the entire system is invariant to a time-reversal. We start from the appropriate nonequilibrium steady-state, at a time symmetric point of the perturbation, and propagate forward in time a whole number of cycles. The corresponding time reversed process is then identical to the forward process, with both starting and ending in the same steady state ensemble. The entropy production for this system is odd under a time reversal and the fluctuation theorem is valid.

As a second example, consider a fluid under a constant shear [1]. The fluid is contained between parallel walls which move relative to one another. Eventually the system settles into a nonequilibrium steady state. A time reversal of this steady-state ensemble will reverse all the velocities, including the velocity of the walls. The resulting ensemble is related to the original one by a simple reflection, and is therefore effectively invariant to the time reversal. Again, the forward process is identical to the reverse process, and the entropy production is odd under a time reversal.

In general, consider a system driven by a time symmetric, periodic process. We require that the resulting nonequilibrium steady-state ensemble be

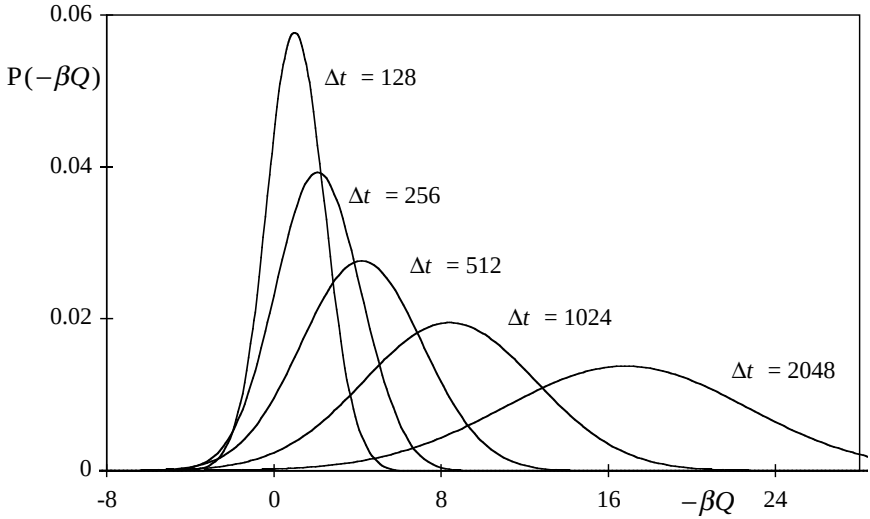


Figure 3.3: Heat probability distribution (—) for the nonequilibrium steady state. The model described in Fig. 3.1 was relaxed to the steady state, and the heat $\mathcal{Q}$ was then measured over 16, 32, 64, 128 and 256 cycles ($\Delta t =$ 128, 256, 512, 1024, and 2048). Note that for long times the system forgets its initial state and the heat distribution is almost indistinguishable from the work distribution of the system that starts in equilibrium.

invariant under time reversal. This symmetry ensures that the the forward and reverse process are essentially indistinguishable, and therefore that the entropy production is odd under a time reversal. It is no longer necessary to explicitly label forward and reverse processes. $\mathrm{P_F}(\omega) = \mathrm{P_R}(\omega) = \mathrm{P}(\omega)$. For these time symmetric steady-state ensembles the fluctuation theorem is valid for any integer number of cycles, and can be expressed as

$$\frac{\mathrm{P}(+\omega)}{\mathrm{P}(-\omega)} = e^{+\omega}. \tag{3.8}$$

For a system under a constant perturbation (such as the sheared fluid) this relation is valid for any finite time interval.

3.4 Long time approximations

The steady-state fluctuation theorem, Eq. (3.8), is formally exact for any integer number of cycles, but is of little practical use because, unlike the equilibrium case, we have no simple method for calculating the probability

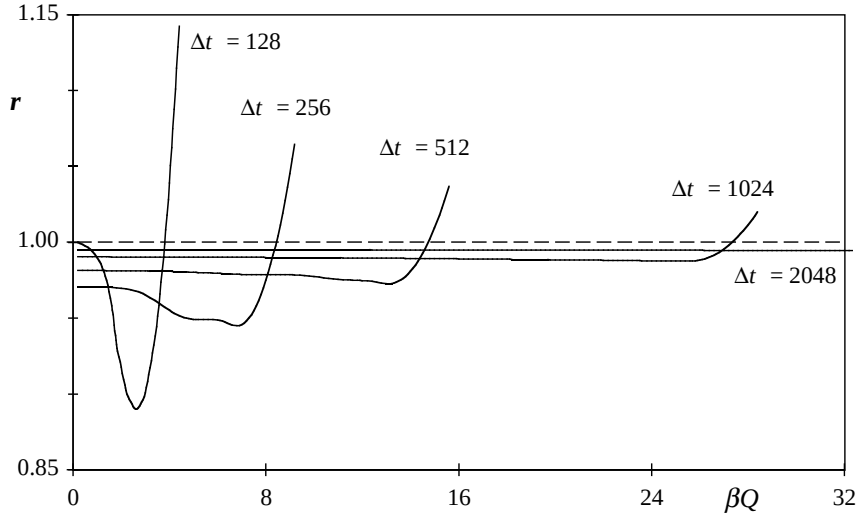


Figure 3.4: Deviations from the heat fluctuation theorem, Eq. (3.10), for the distributions of Fig. 3.3. If the heat fluctuation theorem were exact, then the ratio $r = \beta\mathcal{Q}/\ln[\mathrm{P}(+\beta\mathcal{Q})/\,\mathrm{P}(-\beta\mathcal{Q})]$ would equal 1 for all $|\beta\mathcal{Q}|$. For short times ($\Delta t \leqslant 256$) the fluctuation theorem is wholly inaccurate. For times significantly longer than the relaxation time of the system ($\Delta t \approx 100$), the fluctuation theorem is accurate except for very large values of $|\beta\mathcal{Q}|$.

of a state in a nonequilibrium ensemble. The entropy production is not an easily measured quantity. However, we can make a useful approximation for these nonequilibrium systems which is valid when the entropy production is measured over long time intervals.

We first note that whenever Eq. (3.2) is valid, the following useful relation holds [8, Eq. (16)]:

$$\begin{aligned} \langle e^{-\omega} \rangle &= \int_{-\infty}^{+\infty} \mathrm{P}_{\mathrm{F}}(+\omega) e^{-\omega} d\omega = \int_{-\infty}^{+\infty} \mathrm{P}_{\mathrm{R}}(-\omega) d\omega \\ &= 1. \end{aligned} \tag{3.9}$$

From this identity, and the inequality $\langle \exp x \rangle \geqslant \exp\langle x \rangle$ (which follows from the convexity of e^x) we can conclude that $\langle \omega \rangle \geqslant 0$. On average the entropy production is positive. Because the system begins and ends in the same probability distribution, the average entropy production depends only on the average amount of heat transferred to the bath. $-\langle \omega \rangle = \langle \beta\mathcal{Q} \rangle \leqslant 0$. On average, over each cycle, energy is transferred through the system and into the heat bath (This is the Clausius inequality). The total heat transferred

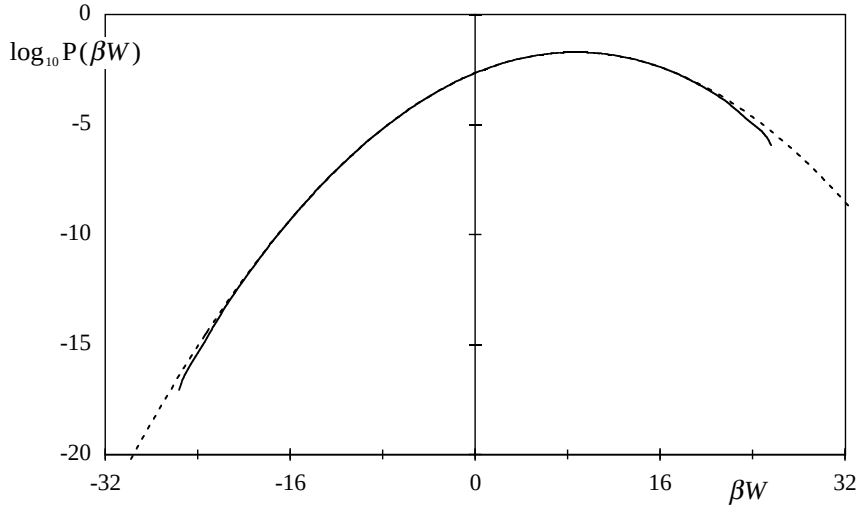


Figure 3.5: Work distribution (—) and Gaussian approximation (- - -) for Δt =2048, with the probabilities plotted on a logarithmic scale. The Gaussian is fitted to the mean of the distribution and has a variance half the mean (see Sec. 3.4). This Gaussian approximation is very accurate, even to the very wings of the distribution, for times much longer than the relaxation time ($\Delta t \approx 100$) of the system. This same distribution is shown on a linear scale in Fig. 3.2.

tends to increase with each successive cycle. When measurements are made over many cycles, the entropy production will be dominated by this heat transfer and $\omega \approx -\beta\mathcal{Q}$. Therefore, in the long time limit the steady state fluctuation theorem, Eq. (3.8), can be approximated as

$$\lim_{\Delta t\to\infty} \frac{\mathrm{P}(+\beta\mathcal{Q})}{\mathrm{P}(-\beta\mathcal{Q})} = \exp(\beta\mathcal{Q}). \tag{3.10}$$

Because $-\beta\mathcal{Q}$ is the change in entropy of the bath, this heat fluctuation theorem simply ignores the relatively small, and difficult to measure microscopic entropy of the system.

Heat distributions for the simple computer model of a nonequilibrium steady state are shown in Fig. 3.3, and the validity of the above approximation is shown in Fig. 3.4. As expected, the heat fluctuation theorem, Eq. (3.10), is accurate for times much longer than the relaxation time of the system.

Another approximation to the entropy production probability distribution can be made in this long time limit. For long times the entropy

production ω is the sum of many weakly correlated values and its distribution should be approximately Gaussian by the central limit theorem. If the driving process is time symmetric, then $\mathrm{P}_{\mathrm{F}}(\omega) = \mathrm{P}_{\mathrm{R}}(\omega) = \mathrm{P}(\omega)$, and the entropy production distribution is further constrained by the fluctuation theorem itself. The only Gaussians that satisfy the fluctuation theorem have a variance twice the mean, $2\langle\omega\rangle = \left\langle(\omega - \langle\omega\rangle)^2\right\rangle$. This is a version of the standard fluctuation-dissipation relation [75]. The mean entropy production (dissipation) is related to the fluctuations in the entropy production. If these distributions are Gaussian, then the fluctuation theorem implies the Green-Kubo relations for transport coefficients [1,68,76]. However, we have not used the standard assumption that the system is close to equilibrium. Instead, we have assumed that the system is driven by a time symmetric process, and that the entropy production is measured over a long time period.

Gaussian approximations are shown in Figs. 3.2 and 3.5 for the work distribution of the simple computer model. For times much longer than the relaxation time of the system these approximations are very accurate, even in the wings of the distributions. This is presumably due to the symmetry imposed by the fluctuation theorem. For a nonsymmetric driving process, this symmetry is broken, and the distributions will not necessarily satisfy the fluctuation-dissipation relation in the long time limit. For example, see Fig. 8 of Ref. [63] and Fig 4.2. Clearly these distributions will be poorly approximated by the Gaussian distributions considered here.

3.5 Summary

The fluctuation theorem, Eq. (3.2), appears to be very general. In this Chapter we have derived a version that is exact for finite time intervals, and which depends on the following assumptions: the system is finite and classical, and coupled to a set of baths, each characterized by a constant intensive parameter. The dynamics are required to be microscopically reversible (1.10) and the entropy production, defined by Eq. (3.3), must be odd under a time reversal. This final condition was shown to hold for two broad classes of systems, those that start in equilibrium and those in a time symmetric nonequilibrium steady state. For the latter systems the fluctuation theorem holds for entropy productions measured over any integer number of cycles. This generality suggests that other nontrivial consequences of the fluctuation theorem are awaiting study.

4. Free Energies From Nonequilibrium Work Measurements

We obtain the equation $\int \frac{dQ}{T} = S - S_0$ *which, while somewhat differently arranged, is the same as that which was formerly used to determine* S. — R. Clausius [77]

4.1 Introduction

Free energies and entropies are of primary importance both in thermodynamics and statistical mechanics. It is therefore somewhat unfortunate that these quantities cannot be directly measured in a computer simulation [79]. For a quantity such as the energy it is sufficient to take an average over a small representative sample of states, but for the entropy (and therefore free energy) it is necessary to consider all the states accessible to the system, the number of which is normally very large.

This is also a problem for experiments on physical systems. There is no such thing as an entropy meter. Fortunately, thermodynamics provides a method for calculating entropy *changes*. Indeed, entropy was originally defined in terms of this procedure [77, 44]. Consider a classical system in contact with a constant temperature heat bath where some degree of freedom of the system can be controlled. (For a concrete example refer to the confined gas discussed on page 3.) Let λ be a parameter specifying the current value of this controllable degree of freedom. We wish to know the entropy change as this parameter is switched from some initial (λ_i) to some final (λ_f) value. If this transformation is carried out infinitely slowly, and is therefore reversible (since the system is always in equilibrium), then Clausius' theorem states that

$$\Delta S = S_f - S_i = \beta \int_{\lambda_i}^{\lambda_f} d\mathcal{Q} = \beta \mathcal{Q}_r \,, \tag{4.1}$$

where ΔS is the entropy change of the system, and $\mathcal{Q}_\mathrm{r}$ is the heat flow associated with this reversible transformation. Note that β is the inverse temperature of the heat bath. For a reversible transformation the temperature of the system and bath are identical.

A similar expression for the free energy change can be obtained from Clausius' theorem by subtracting $\beta\Delta E$ from both sides of the expression above;

$$\begin{aligned} \Delta S - \beta\Delta E &= \beta\mathcal{Q}_r - \beta\Delta E\,, \\ \beta\Delta F &= \beta\mathcal{W}_{\mathrm{r}}\,. \end{aligned} \tag{4.2}$$

Since this is a thermodynamic description, E should be interpreted as the average energy of the equilibrium system. Incidentally, this relation justifies the identification of the free energy change with the reversible work (page 7).

The application of the Clausius theorem to simulations provides a simple algorithm, thermodynamic integration, for calculating free energy differences. There are two, slightly different implementations. In a slow growth simulation the controllable parameter is switched from the initial to final values very slowly. The hope is that the system remains in equilibrium (to a good approximation) throughout, so that the work done during this transformation is (approximately) the free energy difference. This method is referred to a "slow growth", since it has frequently been used to calculate chemical potentials by slowly growing an additional particle in a simulation. The chemical potential is related to the free energy change of this process [80, 81].

Alternatively (4.2) can be rewritten as

$$\Delta F = \int_{\lambda_i}^{\lambda_f} \left\langle \frac{\partial E}{\partial \lambda} \right\rangle_\lambda d\lambda\,. \tag{4.3}$$

Here, $\langle\cdots\rangle_\lambda$ indicates an average over an equilibrium ensemble with fixed λ. The partial derivative is the work done on the system due to an infinitesimal change in λ. In practice, the average is evaluated for a series of closely spaced equilibrium simulations, and the integral is approximated by a sum. This variant is referred to as staged thermodynamic integration, or simply as thermodynamic integration.

Another common algorithm for calculating free energy differences in simulations is thermodynamic perturbation [82, 71];

$$\left\langle e^{-\beta\Delta E}\right\rangle_{\lambda_{\mathrm{i}}} = e^{-\beta\Delta F}\,. \tag{4.4}$$

Again, $\langle\cdots\rangle_\lambda$ indicates an equilibrium average with the specified value of λ. The energy difference, $\Delta E = E(\lambda_{\mathrm{f}})_x - E(\lambda_{\mathrm{i}})_x$, is the change in energy of a state induced by an instantaneous change in the control parameter from λ_{i} to λ_{f}. This relation can be validated by expanding the equilibrium probability as $\rho(x) = \exp\{\beta F - \beta E_x\}$ (1.1).

$$\left\langle e^{-\beta\Delta E}\right\rangle_{\lambda_{\mathrm{i}}} = \sum_x \rho(x|\beta,\mathbf{E}) \exp\bigl\{-\beta\Delta E\bigr\}$$

$$\begin{aligned} &= \sum_x \exp\{\beta F(\lambda_i) - \beta E(\lambda_{\rm i})_x - \beta E(\lambda_{\rm f})_x + \beta E(\lambda_{\rm i})_x\} \\ &= \exp\{-\beta F(\lambda_i)\} \sum_x \exp\{-\beta E(\lambda_{\rm f})\} \\ &= e^{-\beta \Delta F} \end{aligned} \tag{4.5}$$

Although free energy perturbation is exact in principle, in practice it is only accurate (given finite computational resources) if the change in energies are, on average, small. Then the final ensemble can be considered a small perturbation of the initial ensemble. If the initial and final ensembles are very different, then the total change in the controllable parameter can be broken down into a series of smaller changes. In this staged free energy perturbation method the above average is evaluated in a series of equilibrium simulations, each separated by a small change in λ [79, 83, 71].

Entropy and free energy calculations are inherently computationally expensive. For this reason there has been a large amount of work directed at optimizing these methods [79,84]. These efforts are complicated, because these methods all suffer from both statistical and systematic error, which arise because not only because of finite sample sizes, but also because the simulations are never truly in equilibrium.

4.2 Jarzynski nonequilibrium work relation

Christopher Jarzynski [34] has recently derived an interesting nonequilibrium relation that contains both thermodynamic perturbation and thermodynamic integration as limiting cases. This relation was originally derived for a Hamiltonian system weakly coupled to a heat bath [34], and for Langevin dynamics [34], and was soon generalized to detailed balanced stochastic systems [63,35]. It is now clear that this relation is a simple and direct consequence of microscopic reversibility (1.10) [18]. We begin with a system in thermal equilibrium with fixed λ_i. This controllable parameter is then switched to its final value, λ_f, over some *finite* length of time. The work done on this system due to this switching process is averaged over the resulting nonequilibrium ensemble of paths.

$$\langle e^{-\beta \mathcal{W}} \rangle = e^{-\beta \Delta F} \tag{4.6}$$

Recall that an average of a path function indicates an average over a nonequilibrium path ensemble. Contrast this relation with the conventional methods considered above. Thermodynamic integration corresponds to an infinitely slow change in λ, and thermodynamic perturbation to an infinitely fast change in λ.

This nonequilibrium relation was briefly mentioned in Sec. (2.2.1), and shown to be a special case of the path ensemble average (2.1). It is instructive to consider the explicit derivation of this relation from microscopic reversibility (1.10).

$$\begin{aligned}\langle e^{-\beta\mathcal{W}}\rangle &= \sum_{\mathbf{x}} \rho\big(x(0)\big|\beta,\mathbf{E}(0)\big)\,\mathcal{P}[\mathbf{x}|x(0),\mathbf{M}]\,e^{-\beta\mathcal{W}[\mathbf{x}]} \\ &= \sum_{\mathbf{x}} \rho\big(\widehat{x}(0)\big|\beta,\widehat{\mathbf{E}}(0)\big)\,\widehat{\mathcal{P}}[\widehat{\mathbf{x}}|\widehat{x}(0),\widehat{\mathbf{M}}]\,e^{-\beta\Delta F+\beta\mathcal{W}[\mathbf{x}]}e^{-\beta\mathcal{W}[\mathbf{x}]} \\ &= e^{-\beta\Delta F}\end{aligned}$$

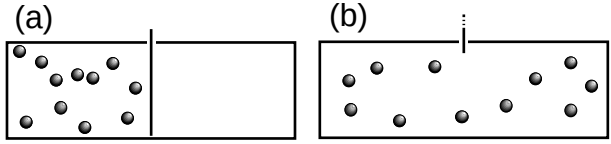
The second line follows from Eq. (2.4), which is itself a direct consequence of microscopic reversibility.

Another useful point of view is to consider the Jarzynski nonequilibrium work relation a direct consequence of the transient fluctuation theorem (2.9) [13]. The sum over paths then becomes a simple integral over work.

$$\begin{aligned}\langle e^{-\beta\mathcal{W}}\rangle &= \int_{-\infty}^{+\infty} \mathrm{P}_{\mathrm{F}}(+\mathcal{W})e^{-\mathcal{W}}d\mathcal{W} = e^{-\beta\Delta F}\int_{-\infty}^{+\infty}\mathrm{P}_{\mathrm{R}}(-\mathcal{W})d\mathcal{W} \\ &= e^{-\beta\Delta F}. \qquad (4.7)\end{aligned}$$

This relation is essentially identical to Eq. (3.9), $\langle e^{-\omega}\rangle = 1$, because for a system that starts in equilibrium the entropy production ω is $\mathcal{W}_{\mathrm{d}}$, the dissipative work (3.6).

There are a number of situations that can cause a naive application of Eq. (4.6) to fail. For example, consider the system illustrated below*. A gas is confined to one half of a box by a thin, movable barrier, and is initially in thermal equilibrium (a).



At some point in time the barrier is removed, and the gas is allowed to expand into the other half of the system. Instead of removing the barrier entirely, we could instead open a series of vents, creating holes through which the gas may pass. In either case the equilibrium free energy of the system has changed, yet removing the barrier requires no work, and this

*This particular failure was pointed out by Prof. D. Rokhsar.

free energy change cannot be calculated from (4.6). However, the Jarzynski nonequilibrium work relation is inapplicable because the system was not truly in equilibrium at the beginning of this process. Although the gas was locally in equilibrium, the system was not ergodic, and therefore not in the relevant global equilibrium. Had the system initially been in the global equilibrium state (gas evenly distributed on either side of the barrier) then the free energy change would have been zero, in agreement with the work done on the system. A similar failure can occurs whenever the perturbation of the system effectively changes the accesible phase space.

Several interesting relations can be considered near equilibrium approximations of Eq. (4.6). The average of the exponential work can be expanded as a sum of cumulants [85],

$$\left\langle e^{cz} \right\rangle = \exp\left\{ \sum_{n=1}^{\infty} \frac{c^n \kappa_n}{n!}, \right\} \tag{4.8}$$

Here, κ_n is the nth cumulant average of z. The higher order cumulants represent fluctuations in the work. Hence, for a very slow, almost reversible switching process the free energy can be well approximated by the first term of this expansion, which is $-\beta\langle \mathcal{W} \rangle$. In this manner we obtain thermodynamic integration (4.3) as the infinite switching time limit of the nonequilibrium work relation.

Alternatively, if the first two terms of the cumulant expansion are retained (which should be a good approximation if the work probability distribution is approximately Gaussian) then we obtain a version of the fluctuation-dissipation ratio [75].

$$\begin{aligned} \beta\Delta F &\approx \beta\langle \mathcal{W} \rangle - \frac{\beta^2 \sigma_{\mathcal{W}}^2}{2} \\ \text{or} \qquad \beta\mathcal{W}_{\mathrm{d}} &\approx \frac{\beta^2 \sigma_{\mathcal{W}}^2}{2} \end{aligned} \tag{4.9}$$

Here, $\langle \mathcal{W} \rangle$ is the mean work (the first cumulant) and $\sigma_{\mathcal{W}}^2$ is the variance of the work (the second moment.) Referring to the final line, we see that the dissipative work (effectively the entropy production (3.6)) is related to the fluctuations in the work. [See also Fig. (4.3)]

In the limit of infinitely fast switching the nonequilibrium work expression becomes equivalent to thermodynamic perturbation (4.5), since the work becomes simply ΔE, the energy change induced by an instantaneous change in λ. Free energy perturbation can also be used to calculate the equilibrium properties of one ensemble from an equilibrium simulation of the other [71]. A very similar relation can now be derived for nonequilibrium processes (2.10).

$$\langle f \rangle_{\lambda_{\mathrm{f}}} = \frac{\langle f(x(\tau)) e^{-\beta \mathcal{W}} \rangle}{\langle e^{-\beta \mathcal{W}} \rangle} \tag{4.10}$$

In this relation $f(x(\tau))$ is a function of the state of the system at time τ. The average on the left is over the equilibrium ensemble with the final value of the control parameter. The averages on the right are over the nonequilibrium procedure that switches λ from the initial to final values.

$$\begin{aligned}
\frac{\langle f(x(\tau))e^{-\beta\mathcal{W}}\rangle}{\langle e^{-\beta\mathcal{W}}\rangle} &= \langle f(x(\tau))e^{-\beta\mathcal{W}_{\rm d}}\rangle \\
&= \sum_{\mathbf{x}} \rho\big(x(0)\big|\beta,\mathbf{E}(0)\big)\,\mathcal{P}[\mathbf{x}|x(0),\mathbf{M}]\,f(x(\tau))e^{-\beta\mathcal{W}_{\rm d}} \\
&= \sum_{\mathbf{x}} \rho\big(\widehat{x}(0)\big|\beta,\widehat{\mathbf{E}}(0)\big)\,\widehat{\mathcal{P}}[\widehat{\mathbf{x}}|\widehat{x}(0),\widehat{\mathbf{M}}]\,f(\widehat{x}(0)) \\
&= \sum_{\mathbf{x}} \rho\big(\widehat{x}(0)\big|\beta,\widehat{\mathbf{E}}(0)\big)\,f(\widehat{x}(0)) \\
&= \langle f\rangle_{\lambda_{\rm f}}\,.
\end{aligned}$$

Thus equilibrium averages can be found by weighting the quantity of interest by a function of the work done on the system. It is also possible to extract the equilibrium probabilities from this nonequilibrium process by the same method. The function f is replaced by a delta function.

$$\rho(x'|\beta,\mathbf{E}(\tau)) = \frac{\langle \delta(x'-x(\tau))e^{-\beta\mathcal{W}}\rangle}{\langle e^{-\beta\mathcal{W}}\rangle} \tag{4.11}$$

Any equilibrium probability can be related to a function of the work performed on the system, averaged over all paths that end in the desired state.

4.3 Optimal computation time

The nonequilibrium work relation (4.6) interpolates smoothly between thermodynamic integration and thermodynamic perturbation. This raises the following interesting question: given a simulation of a switching process, and finite computational resources, which procedure provides the best estimate of the free energy difference? Is it better to use scarce computer time to take many samples of a fast switching process, or only a few samples of a longer and slower perturbation? This question can easily be answered in the limit of large switching time and large sample size.

Suppose that we have made n independent measurements of the amount of work required to switch ensembles over some time Δt. Using the nonequilibrium work relation gives the following estimate of the free energy difference;

$$\beta\Delta F_{\rm est} = -\ln\frac{1}{n}\sum_{i=1}^{n} e^{-\beta w_i}\,. \tag{4.12}$$

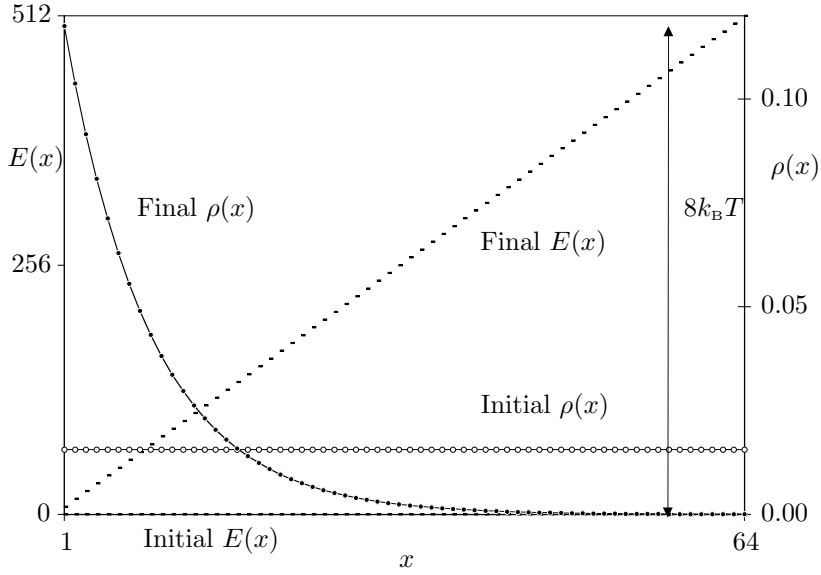


Figure 4.1: A Metropolis Monte Carlo simulation was used to study a simple switching process. The dynamics and implementation are essential the same as those used in the previous Chapter. [See Fig. (3.1) for more details.] A free energy calculation was simulated using a system with the time dependent state energies $E(x) = \lfloor 8xt/\Delta t \rfloor$. Here, x is the state of the system at time t and Δt is the total switching time. At $t = 0$ all energies are zero. During the switching process the slope of this surface gradually increases, resulting in an expenditure of some amount of work. The initial and final state energies (–) are indicated in the figure. The system is coupled to a heat bath of temperature $T = 64$, so that the difference between the lowest and highest energy levels is $8k_{\mathrm{B}}T$. The initial (○) and final (●) equilibrium probability distributions, $\rho(x)$, are also shown in this figure.

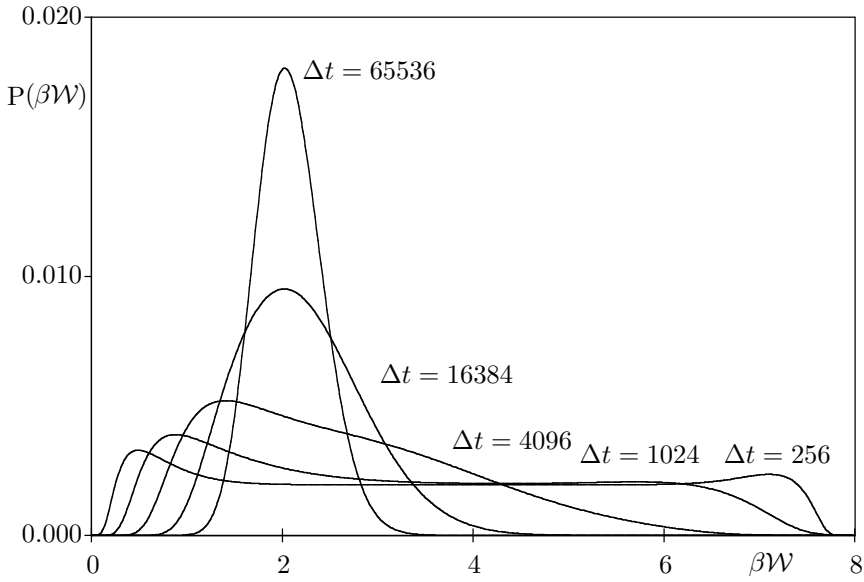


Figure 4.2: Work probability distribution (—) for the system of Fig. 4.1, starting from equilibrium. The state energies were switched from the initial to final values over 256, 1024, 4096, 16384, and 65536 Monte Carlo time steps. The free entropy difference for this process is $\beta\Delta F \approx 2.14096$. Note that in the large switching time limit the work distribution are approximately Gaussians, and are centered around $\beta\Delta F$.

Here the w_i's are the individual, independent measurements of the work. The mean bias of this estimate is

$$\langle \text{bias} \rangle = \langle \beta\Delta F_{\text{est}} - \beta\Delta F \rangle \,, \tag{4.13}$$

and the mean squared error is

$$\alpha = \langle (\beta\Delta F_{\text{est}} - \beta\Delta F)^2 \rangle \,. \tag{4.14}$$

This quantity is sometimes written as $\sigma^2_{\beta\Delta F_{\text{est}}}$, which is, at best, misleading. The mean squared error is not the variance of $\beta\Delta F_{\text{est}}$, since the mean of $\beta\Delta F_{\text{est}}$ is not $\beta\Delta F$, unless the estimate is unbiased.

The bias and mean squared error can be related using the following identity:

$$\langle e^{-\text{bias}} \rangle \quad = \quad \langle e^{-\beta\Delta F_{\text{est}} + \beta\Delta F} \rangle$$

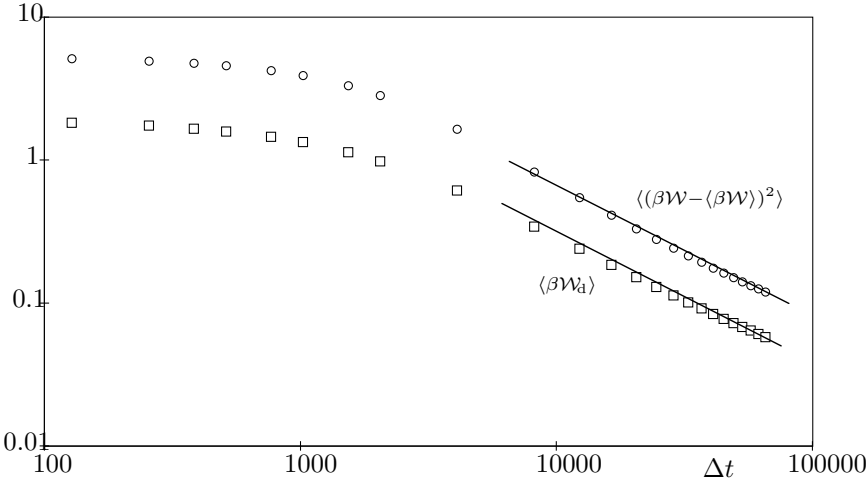


Figure 4.3: The mean dissipative work and the variance of the work versus switching time, plotted on a logarithmic scale. Note that for long switching time $\beta\mathcal{W}_{\rm d} \approx \langle(\beta\mathcal{W} - \langle\beta\mathcal{W}\rangle)^2\rangle/2$, and that both these quantities scale as $1/\Delta t$.

$$\begin{aligned} &= \langle \frac{1}{n}\sum_{i=1}^{n} e^{-\beta w_i}\rangle e^{+\beta\Delta F} \\ &= 1\,. \end{aligned} \tag{4.15}$$

It immediately follows that the bias is always positive for any finite number of samples (since $\langle\exp\{-x\}\rangle \geqslant \exp\langle x\rangle$). For long switching times the bias will be small, and the exponential in (4.15) can be expanded as a Taylor series. Retaining only the first few terms, we may conclude that the bias is approximately twice the mean squared error in this limit, i.e., $\langle\text{bias}\rangle \approx \alpha/2$.

The mean squared error can also be explicitly written as

$$\alpha = \left\langle \left(\ln \frac{\frac{1}{n}\sum_{i=1}^{n} e^{-\beta w_i}}{\langle e^{-\beta\mathcal{W}}\rangle} \right)^2 \right\rangle \tag{4.16}$$

In the limit of large sample sizes, or large switching times, the argument of the logarithm is approximately unity, and we can use the approximation that $\ln x \approx x - 1$.

$$\alpha \approx \frac{1}{n}\left\langle \left(e^{-\beta\mathcal{W}_{\rm d}} - \langle e^{-\beta\mathcal{W}_{\rm d}}\rangle\right)^2 \right\rangle \tag{4.17}$$

This can be further simplified by once more taking the long switching time

limit, and using a truncated Taylor series expansion of the exponentials.

$$\alpha \approx \frac{1}{n}\left\langle \left(\beta\mathcal{W} - \langle\beta\mathcal{W}\rangle\right)^2 \right\rangle \tag{4.18}$$

If the switching time is much longer than the correlation time of the system, then the work done on the system is the sum of many weakly correlated values. It follows that, in this limit, the probability distributions of the work are approximately Gaussian [see Fig. (4.2)], and that the variance of the work scales inversely with the switching time, Δt [see Fig. (4.3)]. The combination of these results gives the following chain of approximations.

$$\langle \text{bias} \rangle \approx \frac{\alpha}{2} \approx \frac{1}{n}\left\langle \left(\beta\mathcal{W} - \langle\beta\mathcal{W}\rangle\right)^2 \right\rangle \sim \frac{1}{n\Delta t} \sim \frac{1}{\mathcal{T}} \tag{4.19}$$

Here $\mathcal{T}$ is the total simulation time, which is approximately $n\Delta t$ for large Δt.

We are therefore led to the following conclusion; in the large sample, large switching time limit both the systematic and statistical errors scale inversely with the simulation time. It does not matter whether the computation time is used to generate many samples, or fewer samples of longer processes. However, large sample sizes allow a more accurate estimation of the error in the calculation, and are therefore preferable, all other things being equal.

4.4 Bennett acceptance ratio method

It is possible to extend the Jarzynski nonequilibrium work relation Eq. (4.6) to a more general class of equalities between the work and the free energy change [See Sec (2.2.1)]. If $f(\mathcal{W})$ is any finite function of the work then the path ensemble average (2.1) predicts that

$$e^{-\beta\Delta F} = \frac{\langle f(+\mathcal{W})\rangle_{\text{F}}}{\langle f(-\mathcal{W})e^{-\beta\mathcal{W}}\rangle_{\text{R}}} \tag{4.20}$$

Recall that ΔF is defined in terms of the forward process. This equation clearly contains the Jarzynski relation as a special case. ($f(\mathcal{W}) = \exp\{-\beta\mathcal{W}\}$)

Suppose that we have made n_{F} independent measurements of the amount of work required to switch ensembles using the forward process, and n_{R} independent measurements from the reverse process. An interesting question is what choice of $f(\mathcal{W})$ leads to the highest statistical accuracy for ΔF? For instantaneous switching this question was answered by Bennett [79, 71, 84] in his derivation of the acceptance ratio method for calculating free energy differences. Bennett's methods and results can be readily adapted to finite switching times.

Given a finite number of samples then the mean squared error is

$$\begin{aligned}\alpha &= \left\langle (\beta\Delta F_{\rm est} - \beta\Delta F)^2 \right\rangle \\ &= \left\langle \left(\ln \frac{\frac{1}{n_{\rm F}}\sum_{i=1}^{n_{\rm F}} f(+w_i)}{\langle f(+\mathcal{W})\rangle_{\rm F}} - \ln \frac{\frac{1}{n_{\rm R}}\sum_{j=1}^{n_{\rm R}} f(-w'_j)\, e^{-\beta w'_j}}{\langle f(-\mathcal{W})\, e^{-\beta\mathcal{W}}\rangle_{\rm R}} \right)^2 \right\rangle\end{aligned}$$

Here the w_i's are the individual measurements of the work from the forward process and the w'_j's are the measurements from the reverse process. This expression is difficult to evaluate. However in the limit of large sample size the arguments of the logarithms approach 1, so that we may use the approximation that $\ln x \approx x - 1$. Applying this approximation, expanding the square and evaluating the cross terms (recall that all the w_i and w'_j's are statistical independent) we obtain

$$\alpha \approx \frac{\left\langle f(+\mathcal{W})^2 \right\rangle_{\rm F} - \left\langle f(+\mathcal{W}) \right\rangle_{\rm F}^2}{n_{\rm F}\left\langle f(+\mathcal{W}) \right\rangle_{\rm F}^2} + \frac{\left\langle f(-\mathcal{W})^2\, e^{-2\beta\mathcal{W}} \right\rangle_{\rm R} - \left\langle f(-\mathcal{W})\, e^{-\beta\mathcal{W}} \right\rangle_{\rm R}^2}{n_{\rm R}\left\langle f(-\mathcal{W})\, e^{-\beta\mathcal{W}} \right\rangle_{\rm R}^2}$$

The path ensemble average (2.1) can be used to convert all averages taken over the reverse process to averages over the forward process.

$$\begin{aligned}\langle f(-\mathcal{W})\, e^{-\beta\mathcal{W}}\rangle_{\rm R} &= e^{+\beta\Delta F}\langle f(+\mathcal{W})\rangle_{\rm F} \\ \langle f(-\mathcal{W})^2\, e^{-2\beta\mathcal{W}}\rangle_{\rm R} &= \langle f(+\mathcal{W})^2\, e^{+\beta\Delta F+\beta\mathcal{W}}\rangle_{\rm F}\end{aligned}$$

In addition we may arbitrarily choose the normalization of the function f such that,

$$\langle f(+\mathcal{W})\rangle_{\rm F} = k\,, \tag{4.21}$$

where k is a constant. Making these substitutions we obtain

$$\begin{aligned}\alpha &\approx \frac{\langle f(+\mathcal{W})^2\rangle_{\rm F}}{k^2\, n_{\rm F}} + \frac{e^{-\beta\Delta F}\langle f(+\mathcal{W})^2\, e^{+\beta\mathcal{W}}\rangle_{\rm F}}{k^2\, n_{\rm R}} - \frac{1}{n_{\rm F}} - \frac{1}{n_{\rm R}} \\ &\approx \int_{-\infty}^{+\infty} {\rm P_F}(\mathcal{W})\, \frac{1}{k^2}\left[\frac{f(+\mathcal{W})^2}{n_{\rm F}} + \frac{f(+\mathcal{W})^2\, e^{-\beta\Delta F+\beta\mathcal{W}}}{n_{\rm R}} \right] d\mathcal{W} - \frac{1}{n_{\rm F}} - \frac{1}{n_{\rm R}}\end{aligned}$$

We wish to minimize the mean squared error, α, with respect to the function f, with the constraint Eq. (4.21). This can easily be accomplished by introducing a Lagrange undetermined multiplier, λ.

$$0 = {\rm P_F}(\mathcal{W})\, \frac{1}{k^2}\Big(\frac{1}{n_{\rm F}} + \frac{e^{-\beta\Delta F-\beta\mathcal{W}}}{n_{\rm R}} \Big) f(\mathcal{W})\, \delta f(\mathcal{W}) - \lambda {\rm P_F}(\mathcal{W})\, \delta f(\mathcal{W})$$

The constant k can be conveniently set to $1/\sqrt{\lambda}$. The least error then occurs if the function f is set to

$$f(\mathcal{W}) = (1 + \exp\{-\beta\mathcal{W} - C\})^{-1}$$

where C is $-\beta\Delta F + \ln n_{\rm F}/n_{\rm R}$.

Inserting this function into Eq. (4.20), we obtain the following optimal estimate for the free energy difference;

$$e^{-\beta\Delta F} = \frac{\left\langle (1+\exp\{+\beta\mathcal{W}+C\})^{-1}\right\rangle_{\rm F}}{\left\langle (1+\exp\{+\beta\mathcal{W}-C\})^{-1}\right\rangle_{\rm R}}\,\exp\{-C\}\,. \tag{4.22}$$

The optimal choice of the constant C is $-\beta\Delta F + \ln n_{\rm F}/n_{\rm R}$. Because C appears on both sides of this relation it must be solved self-consistently.

The nonequilibrium free energy calculations studied in this Chapter provide an interesting set of alternatives to the more conventional methods for calculating free energy differences. However, it has yet to be seen whether these methods are significantly more efficient in practice.

5. Response Theory

Clouds are not spheres, mountains are not cones, coastlines are not circles, and bark is not smooth, nor does lightning travel in a straight line.

— Benoit B. Mandelbrot [86]

5.1 Introduction

The response of a many particle classical system to an external time-dependent perturbation is generally a complex, often nonanalytic function of the strength of the perturbation. Fortunately, for many relevant systems the response is a linear function of the strength of the perturbation. This observation is the basis of linear response theory, which has its simplest, quantitative expression in the fluctuation-dissipation theorem*;

$$\big\langle A(\tau)\big\rangle_{\rm neq} - \big\langle A\,\big\rangle_{\rm eq} \approx \beta\lambda\Big[\big\langle A(0)B(\tau)\big\rangle_{\rm eq} - \big\langle A\big\rangle_{\rm eq}\big\langle B\big\rangle_{\rm eq}\Big] \tag{5.1}$$

Here, β is the inverse temperature of the environment, A and B are state functions of the system, and λ is a parameter that controls the strength of the perturbation. This perturbation is due to a field that couples to the variable B so that the energy of the system is changed by $-\lambda B$. (For an Ising model λ could be the strength of the external magnet field, and B the net magnetization.) The system is prepared in an equilibrium ensemble with a finite λ. At time $t = 0$ the field is discontinuously and instantaneously switched off (λ is set to zero). The equilibrium averages correspond to $\lambda = 0$, and $\langle A(\tau)\rangle_{\rm neq}$ is the nonequilibrium average of A at a time τ after the perturbation. This fluctuation-dissipation theorem relates equilibrium fluctuations (the time correlation function on the right in the expression above) to the relaxation of the system towards equilibrium from a nonequilibrium state. On this simple foundation the entire Green-Kubo linear response theory [87, 88, 89, 90, 50] can be built.

Although linear response is useful and widely applicable, it does have its limitations. For example, Skaf and Ladanyi recently studied a charge transfer reaction in a simple fluid and found significant deviations from

*Some scientists reserve the terminology "fluctuation-dissipation theorem" for a relationship that is equivalent . . . but somewhat different in appearance—D. Chandler [49, p. 255]

the fluctuation-dissipation theorem [91]. In addition, there are situations where the lowest order term is not linear (such as hydrodynamic is 2 dimension [92]) and others where higher order terms diverge [93]. For a fuller discussion of these issues see the introduction to [67].

5.2 Kawasaki response and the nonlinear fluctuation-dissipation theorem

Despite the difficulties associated with linear response, it is possible to write nontrivial expressions valid far-from-equilibrium, which reduce to linear response in the appropriate limit. Kubo [89] himself gave a relation for the nonlinear response, expressed as a sum of convolution of operators. It does not appear to have been used with any success. A more practical expression was developed by Yamada and Kawasaki [65], which gives the response of a adiabatic Hamiltonian system to an instantaneous perturbation. This relation has subsequently been generalized to deterministic, thermostatted dynamics with arbitrary time dependent perturbations [66, 67, 68, 69], using the same systems and similar reasoning used to derive the fluctuation theorem (3.2). From the relations developed earlier it is now possible to derive an extension of the Kawasaki response to thermostatted systems with stochastic dynamics. Also working from the fluctuation theorem, Kurchan derived what is effectively a nonlinear fluctuation-dissipation theorem for stochastic dynamics [9]. We will see that this can be derived as a special case of the Kawasaki response.

The average of the state function A, measured over a nonequilibrium ensemble generated by an arbitrary time dependent perturbation, can be written as

$$\big\langle A(\tau)\big\rangle_{\mathrm{F}} = \sum_{\mathbf{x}} \rho\big(x(0)\big|\beta, \mathbf{E}(0)\big)\, \mathcal{P}[\,\mathbf{x}|x(0), \mathbf{M}]\, A(\tau)\,. \tag{5.2}$$

In anticipation of the manipulations to come, I have added the subscript "F" as a reminder that this average is over the forward perturbation. Note that the system is initially in equilibrium. As usual, I will assume the microscopic reversibility (1.10) ($\mathcal{P}[\,\mathbf{x}|x(0), \mathbf{M}] = \widehat{\mathcal{P}}[\,\widehat{\mathbf{x}}|\widehat{x}(0), \widehat{\mathbf{M}}]\exp\{-\beta\mathcal{Q}[\mathbf{x}]\}$) of the underlying dynamics. Using this property the nonequilibrium average of A can be related to another average, taken over the same system driven in the "reverse" direction (2.4):

$$\begin{aligned}\big\langle A(\tau)\big\rangle_{\mathrm{F}} &= \sum_{\mathbf{x}} \rho\big(\widehat{x}(0)\big|\beta, \widehat{\mathbf{E}}(0)\big)\, \widehat{\mathcal{P}}[\widehat{\mathbf{x}}|\widehat{x}(0), \widehat{\mathbf{M}}]\, A(0)\, e^{-\beta\mathcal{W}_{\mathrm{d}}[\widehat{\mathbf{x}}]} \\ &= \big\langle A(0)\, e^{-\beta\mathcal{W}_{\mathrm{d}}}\big\rangle_{\mathrm{R}}\,. \end{aligned} \tag{5.3}$$

Recall that $\mathcal{W}_\mathrm{d}$ is the dissipative work. It will prove advantageous to expand the dissipative work as $-\Delta F + \mathcal{W}$, and rewriting the free energy change as a work average using the Jarzynski relation (2.7);

$$\big\langle A(\tau)\big\rangle_\mathrm{F} = \big\langle A(0)\, e^{-\beta\mathcal{W}}\big\rangle_\mathrm{R} \Big/ \big\langle e^{-\beta\mathcal{W}}\big\rangle_\mathrm{R} \,. \tag{5.4}$$

Equation 5.3 is referred to as the bare form, and Eq. (5.4) as the renormalized form of the Kawasaki response. Simulation data indicates that averages calculated with the renormalized expression typically have lower statistical errors [68]. The expressions above differ from the standard Kawasaki forms in three important respects: The response is explicitly defined in terms of the work, it is generalized to arbitrary forcing, and it incorporates an average over all possible trajectories, which is necessitated by the extension to stochastic dynamics. This relation can also be considered a simple application of the path ensemble average (1.10) with $\mathcal{F}[x(t)] = A(\tau)$. See Sec. 2.2.3.

Let us now consider a more restrictive situation, the kind of perturbation envisaged in the derivation of linear response. A system begins in equilibrium with a perturbation that changes the energy of the system by $-\lambda B(0)$, where B is a state function and λ a parameter. At time 0 this perturbation is removed (λ is set to zero), and the state of the system is observed at time τ. In the reverse perturbation the system is in equilibrium with $\lambda = 0$ up to time τ. The perturbation is then switched on changing the energy by $-\lambda B(\tau)$. This is the only work done on the system. Furthermore, this is a purely equilibrium average, effectively a time correlation function between A at time 0 and $e^{+\beta\lambda B}$ at time time τ. This can be made more evident by subtracting $\big\langle A\big\rangle_\mathrm{eq}$ from both sides of the Kawasaki relation. Therefore, with these replacements and changes, the Kawasaki response (5.4) for a single, instantaneous change to the system can be written as

$$\big\langle A(\tau)\big\rangle_\mathrm{neq} - \big\langle A\big\rangle_\mathrm{eq} = \frac{\big\langle A(0)\, e^{+\beta\lambda B(\tau)}\big\rangle_\mathrm{eq} - \big\langle A\big\rangle_\mathrm{eq}\big\langle e^{+\beta\lambda B}\big\rangle_\mathrm{eq}}{\big\langle e^{+\beta\lambda B}\big\rangle_\mathrm{eq}} \,. \tag{5.5}$$

This expression is a nonlinear fluctuation-dissipation theorem valid for arbitrary strong perturbations. It is effectively equivalent to that derived by Kurchan [9, Eq. (3.27)]. The standard, linear, fluctuation dissipation theorem (5.1) can be obtained as an approximation by expanding in powers of λ, and retaining only the lowest order terms.

5.3 Nonequilibrium probability distributions

The probability distribution of a nonequilibrium ensemble is not determined solely by the external constraints, but explicitly depends on the

history and dynamics of the system. A particular nonequilibrium ensemble can be generated by starting with an equilibrium system, and perturbing the system away from equilibrium in some predetermined manner. An expression for this nonequilibrium probability distribution can derived from the Kawasaki relation (5.4) by setting the state function to be $A(\tau) = \delta\big(x - x(\tau)\big)$, a δ function of the state of the system at time τ;

$$\rho_{\text{neq}}\big(x, \tau \mid \mathbf{M}\big) = \rho\big(x \mid \beta, \mathbf{E}(\tau)\big) \frac{\big\langle e^{-\beta \mathcal{W}} \big\rangle_{\text{R},x}}{\big\langle e^{-\beta \mathcal{W}} \big\rangle_{\text{R}}} \,. \tag{5.6}$$

Here, $\rho\big(x, \tau | \mathbf{M}\big)$ is the nonequilibrium probability distribution and $\rho\big(x | \beta, \mathbf{E}(\tau)\big)$ is the equilibrium probability of the same state. The subscript 'x' indicates that the average is over all paths that start in state x. In contrast the lower average is over all paths starting from an equilibrium ensemble. Thus, the nonequilibrium probability of a state is, to zeroth order, the equilibrium probability, and the correction factor can be related to a nonequilibrium average of the work.

This relation for the nonequilibrium probability should be compared with the nonequilibrium relation for equilibrium probabilities, Eq. (4.11), that was derived in the previous chapter. The two expressions are very similar. To obtain nonequilibrium probabilities it is necessary to associate the work with the initial state, whereas for equilibrium probabilities it was necessary to correlate the work with the final state.

A particular type of nonequilibrium ensemble can be generated by applying a continuous, periodic perturbation to the system. Eventually the system may relax (if the corresponding transition matrix is aperiodic and irreducible [39]) into a steady-state that is determined solely by this perturbation, and not by the initial state. The sheared fluid discussed on page 34 provides a useful example. However, the relaxation of the system into this nonequilibrium steady state requires an infinite amount of time, just as it requires infinite time to fully equilibrate a system. During this time the average work done on the system increases with each subsequent cycle (3.9), and becomes infinite in the infinite time limit that we are most interested in. Fortunately, due to the formal renormalization of the Kawasaki relation (5.4) it is possible to reexpress Eq. (5.6) in terms of finite quantities. Both the upper and lower average can be expanded as a sum of cumulants (see page 4.8);

$$\rho_{\text{neq}}\big(x, \tau \mid \mathbf{M}\big) = \rho\big(x \mid \beta, \mathbf{E}(\tau)\big) \exp\left\{ \sum_{n=1}^{\infty} \frac{(-\beta)^n}{n!} \big(\kappa_n^x(\mathcal{W}) - \kappa_n^{\text{eq}}(\mathcal{W}) \big) \right\} . \tag{5.7}$$

Here, $\kappa_n(\mathcal{W})$ is the nth cumulant average of the work done on the system, and the superscripts distinguish averages for a system starting in state x

and those starting from an equilibrium ensemble. Cumulants of all orders are additive for independent random variables. Since trajectory segments separated in time are weakly correlated, it follows that all the cumulants in the average above are either identically zero, or scale linearly with time. However, the phase space distribution tends to the steady state distribution exponentially fast [39], irrespective of the initial state. It follows that each difference of cumulants in the sum above is finite, and converges to its long time limit exponentially fast.

To proceed further it is necessary to develop useful approximations. The first cumulant is the mean of the work, and the second is the variance. If the fluctuations in the work are relatively small, then it seems plausible that the right side of Eq. (5.7) could be well approximated by the first term of the cumulant expansion.

$$\rho_{\rm neq}\big(x,\tau \bigm| \mathbf{M}\big) \approx \rho\big(x \bigm| \beta, \mathbf{E}(\tau)\big) e^{-\beta\left(\langle \mathcal{W}\rangle_x - \langle \mathcal{W}\rangle_{\rm eq}\right)} \tag{5.8}$$

Here, $\langle \mathcal{W}\rangle_x$ is the average amount of work done on the system if it starts in state x, and $\langle \mathcal{W}\rangle_{\rm eq}$ is the average amount of work if the system started from an equilibrium ensemble. Comparison with similar expressions that have been derived for thermodynamic fluctuations [94, 95, 96, 97] suggest that this expression may be accurate for two distinct types of systems; either the system is close to equilibrium or the dynamics are almost deterministic. In the latter case the system could be far-from-equilibrium, but the randomness is, at worst, a small perturbation to the deterministic dynamics. Essentially, this is the approach used by Evans and Morris in Ref. [67], and several related papers. A driven, thermostatted system is modeled with Hamiltonian dynamics incorporating a Gaussian constraint. This provides a thermostatted, yet completely deterministic dynamics. Then Eq. (5.8) would be exact. (However, there are several pitfalls and technical difficulties associated with these deterministic dynamics, including a fractal phase space measure in the steady state.) The accuracy for a simple stochastic system is shown in Fig. (5.1).

An expression for the nonequilibrium entropy can be derived from this expression by substituting the explicit canonical expression (1.1) for the equilibrium probability, and using the definition $S = -\sum_x \rho_x \ln \rho_x$.

$$S_{\rm neq} \approx -\beta F + \beta \langle E\rangle_{\rm neq} - \beta \langle \mathcal{W}^{\rm ex}\rangle \tag{5.9}$$

Here, F is the equilibrium free energy of the system, and $\langle \mathcal{W}^{\rm ex}\rangle$ is the mean excess work, the difference in the mean work done on a system that starts in the nonequilibrium steady state, and the mean work done on the system if it starts in the equilibrium ensemble. By subtracting the equilibrium entropy, $S_{\rm eq} = -\beta F + \beta \langle E\rangle_{\rm eq}$ we obtain an intriguing approximation;

$$S_{\rm neq} - S_{\rm eq} \approx \beta \langle E\rangle_{\rm neq} - \beta \langle E\rangle_{\rm eq} - \beta \langle \mathcal{W}^{\rm ex}\rangle \approx \beta \langle \mathcal{Q}^{\rm ex}\rangle \,. \tag{5.10}$$

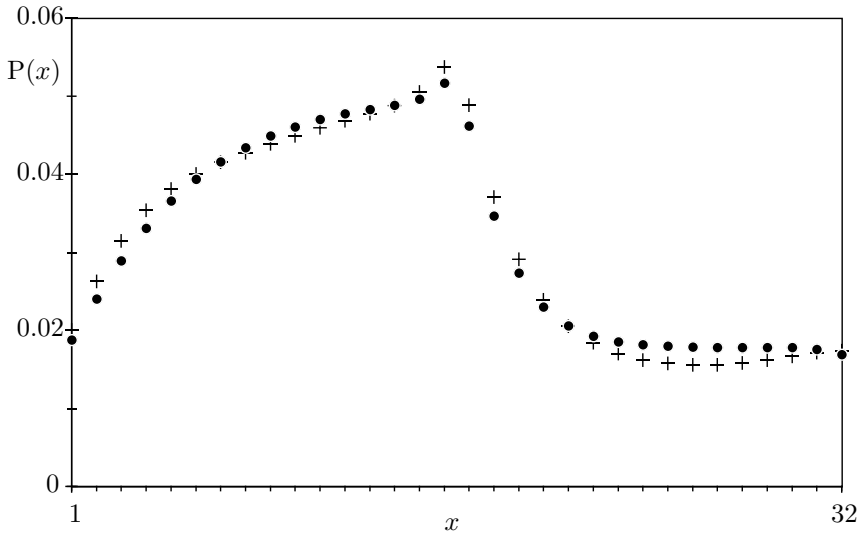


Figure 5.1: The exact relation, Eq. (5.6), and near equilibrium approximations, Eq. (5.8), for the steady state probability distribution were tested against the sawtooth system studied in Ch. (3). The nonequilibrium steady state probability distribution (•) was calculated by propagating the system forward in time until the probabilities converged. The work distributions for the system starting from each state were calculated for a time much longer than the relaxation time of the system. Probabilities were calculated from these distributions using both Eqs. (5.6) and (5.8). As expected, the first gave results almost indistinguishable from the true steady state probabilities. Small errors were largely due to the finite time that each system was perturbed, and partial due to numerical roundoff error. Unfortunately, this procedure is completely impractical for a real system. The near equilibrium approximation, Eq. (5.8), gave surprisingly accurate results (+). (Compare with the equilibrium probability distribution in Fig. (3.1)) The distribution is qualitatively correct, with about 10% error. Interestingly, utilizing the next term in the cumulant expansion, Eq. (5.7), did not significantly improve the accuracy.

The mean excess heat, $\langle \mathcal{Q}^{ex} \rangle$, is defined in an analogues manner to the excess work. This expression bears a striking resemblance to the Clausius relation (4.1) for equilibrium entropy differences, and appears to be very closely related to the work of Oono and Paniconi [98], who define an excess heat in a very similar manner, and suggest a steady state thermodynamic entropy that is compatible with this expression.

5.4 Miscellaneous comments

There are several other far-from-equilibrium relations that have been derived from, or are related to the Kawasaki response. The transient time correlation function (TTCF) [99, 100, 101, 102, 103] gives another set of relations for the nonlinear response of a system, which maybe of greater practical utility [67] than the Kawasaki response relation. Unfortunately it appears that TTCF can not be applied to the systems considered here, since a crucial step linking the two formalisms [67] makes the assumption that the dynamics are deterministic, and therefore that only an average over initial conditions is needed. Similarly Evans and Morriss have derived several interesting relations for the heat capacity of a nonequilibrium steady-state [67], but again these relations are not generally applicable because it is assumed that the probability of a trajectory is independent of the temperature of the heat bath.

The Kawasaki relation itself has not proved particular useful in practice. However, its chief utility may be to suggest approximations, other than linear response, that provide interesting perspectives, and that are (potentially) of practical utility. The steady-state entropy (5.10) may represent one such relation, but it is not presently clear how accurate this approximation generally is.

6. Path Ensemble Monte Carlo

Science is what we understand well enough to explain to a computer. — Donald E. Knuth [105]

6.1 Introduction

In modern statistical physics computer simulation and theory play complementary roles. Computer simulations are used not only to validate theories and their approximations, but also as a source of new observations which then inspire and inform further analytic work. Therefore, in this chapter I turn from the statistical theory of nonequilibrium systems, to the efficient simulation of stochastic dynamics.

Recently, approximate theories have been developed that describe large, rare fluctuations in systems with Langevin dynamics that have been driven from equilibrium by a time dependent or non-gradient (i.e., not the gradient of a potential) force field [106,107,108,109,110,111,112,113]. These theories are only good approximations in the zero noise limit; computer simulations are needed to explore the behavior of the system and the accuracy of the approximations at finite noise intensities. However, the direct simulation of the dynamics is inherently inefficient, since the majority of the computation time is taken watching small, uninteresting fluctuations about the stable states, rather than the interesting and rare excursions away from those states. Transition path sampling [23, 24, 25, 26, 27, 28, 29, 30, 31] has been developed as a Monte Carlo algorithm to efficiently sample rare transitions between stable or metastable states in equilibrium systems. Only trajectories that undergo the desired transition in a short time are sampled. In this chapter, transition path sampling is adapted to nonequilibrium, dissipative, stochastic dynamics. The principle innovation is the development of a new algorithm to generate trial trajectories.

The high friction limit of Langevin's equations of motion describe overdamped Brownian motion in a force field;

$$\dot{\mathbf{x}}_i = \mathbf{F}_i(\mathbf{x}, t) + \xi_i(t). \tag{6.1}$$

The state of the system is specified by the vector $\mathbf{x}$. The system is subjected to a systematic force $\mathbf{F}(\mathbf{x}, t)$, and a stochastic force $\xi(t)$, resulting from δ

correlated white noise with variance ϵ;

$$\langle \xi_i(t) \rangle = 0, \quad \langle \xi_i(t) \xi_j(0) \rangle = \epsilon \, \delta_{ij} \delta t. \tag{6.2}$$

In this chapter we are interested in systems that are not in equilibrium, either because the force field, $\mathbf{F}(\mathbf{x}, t)$, is time dependent, or because it is non-gradient. Dynamics of this class can model a large range of interesting problems, including thermal ratchets [114], computer networks [115] and chemical reactions [116, 117, 118]. As a particular example consider the following reaction scheme, the Selkov model of glycolysis [119, 118].

$$A \underset{k_2}{\overset{k_1}{\rightleftharpoons}} X, \qquad X + 2Y \underset{k_4}{\overset{k_3}{\rightleftharpoons}} 3Y, \qquad Y \underset{k_6}{\overset{k_5}{\rightleftharpoons}} B \tag{6.3}$$

The concentrations of these species are written as the corresponding lower case letter, i.e., $a = A/V$ where V is the volume. The reaction vessel is stirred to ensure spatially homogeneity, and the concentration of A and B are held fixed by external sources or sinks. Therefore, the state of the system can be described by just two parameters, ($\mathbf{x} = (x, y)$), and the effective force on the system is

$$\mathbf{F}(x, y) = \big(k_1 a + k_4 y^3 - k_2 x - k_3 x y^2, k_6 b + k_3 x y^2 - k_5 y - k_4 y^3\big). \tag{6.4}$$

Fixing the concentrations of A and B continuously perturbs the system away from equilibrium; this force is non-gradient and the dynamics are not detailed balanced. The system fluctuates around one of two distinct stable states, and occasionally makes a transition from one basin of attraction to another.

The Selkov model is interesting and physical relevant, but suffers from a proliferation of parameters. Instead, I will concentrate on two simpler examples, which exhibit all of the interesting behavior of these nonequilibrium dynamics. An example of a system driven by a time dependent force is the driven double well Duffing oscillator [111, 120] (see Fig. 6.1)

$$F(x, t) = x - x^3 + A \cos(\omega t) \tag{6.5}$$

An extensively studied example of a non-gradient force field is the following 2 dimensional system ($\mathbf{x} = (x, y)$) proposed by Maier and Stein [109];

$$\mathbf{F}(x, y) = \big(x - x^3 - \alpha x y^2, -\mu y (1 + x^2)\big). \tag{6.6}$$

This field is not the gradient of a potential energy unless $\alpha = \mu$. The potential energy surface for the gradient field, $\alpha = \mu = 1$, is shown in Fig. 6.2, which should serve to orient the reader. Of primary interest are the rare transitions between the stable states. For weak noise almost all transitions closely follow the optimal trajectory, the most probable exit

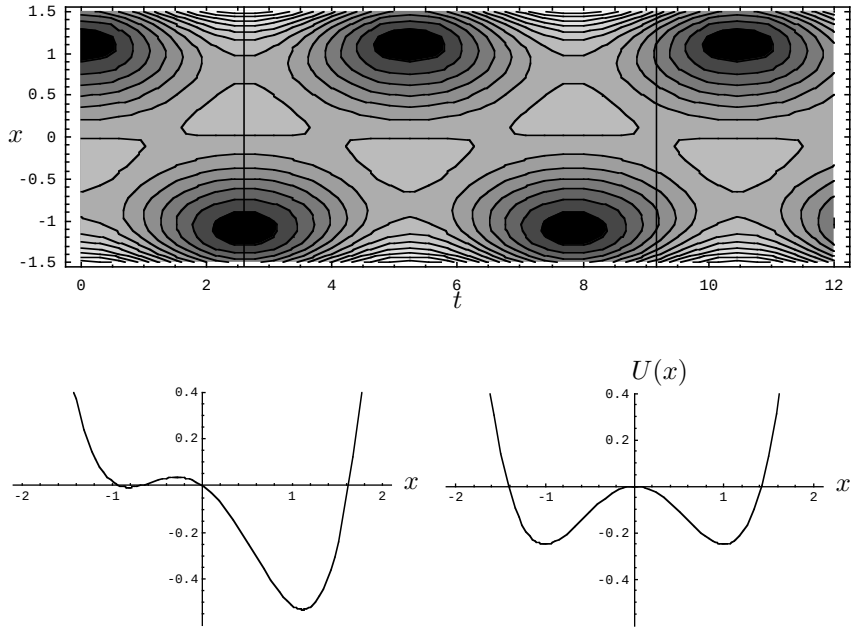


Figure 6.1: The potential energy of a driven double well Duffing oscillator, $U(x) = \frac{x^2}{2} - \frac{x^4}{4} + A\cos(\omega t)x$, with $A = 0.264$ and $\omega = 1.2$. The top figure is the potential energy as a function of time and position. Darker shading indicates lower energies. The lower graphs show the potential at two times, indicated by vertical lines in the upper figure. At the noise intensities studied here ($\epsilon = 0.012$) transitions between the metastable states are rare.

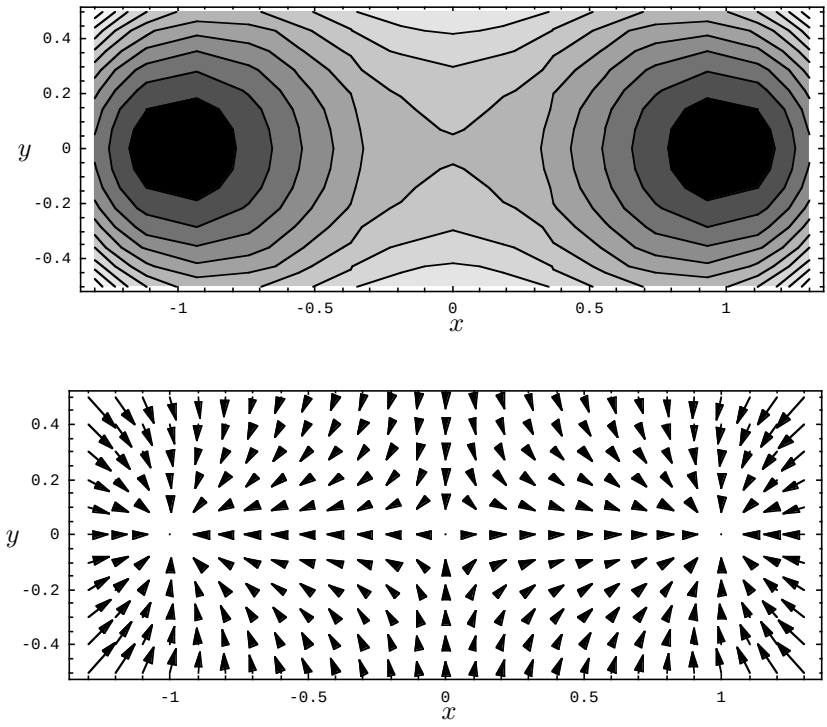


Figure 6.2: The potential energy surface (top) and the corresponding force field (bottom) for the Maier-Stein system, Eq. (6.6), with $\alpha = 1$ and $\mu = 1$. Darker shading indicates lower energies. Note the stable states at $(\pm 1, 0)$, the transition state at $(0, 0)$ and the surface dividing the stable states (the separatrix) at $x = 0$. These general features persist for the other values of the parameters used in this chapter, although the force field is no longer the gradient of a potential energy. For this equilibrium system the most probable path connecting the stable states (and therefore the path that dominates transitions in the weak noise limit) runs directly along the x axis.

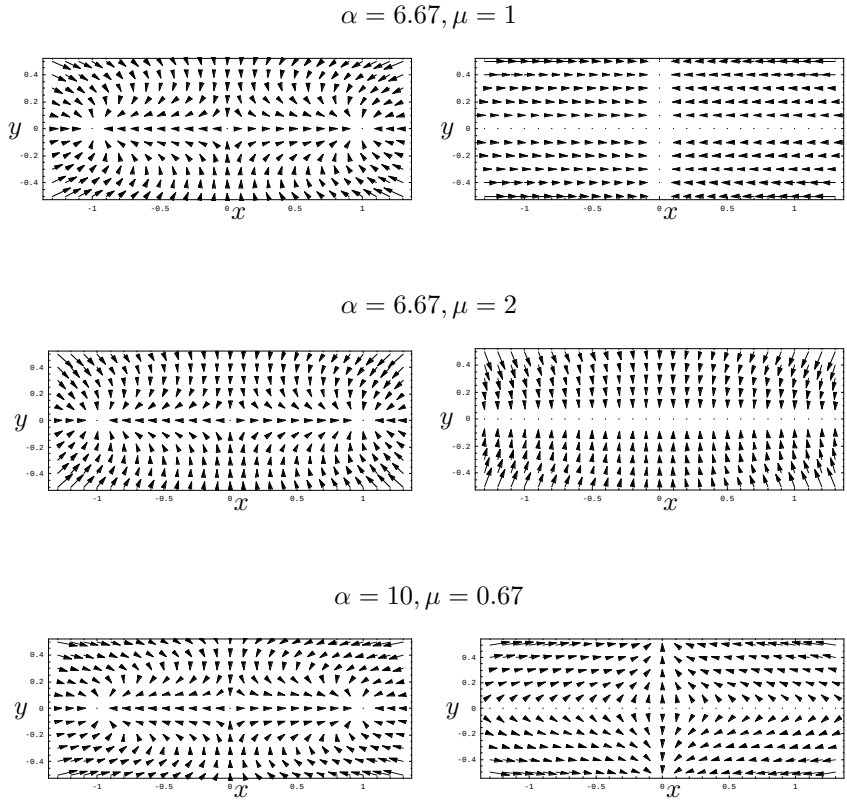


Figure 6.3: Force fields of the Maier-Stein system with various parameters (left), and the difference between that force field and the force field of the same system with $\alpha = 1$ and $\mu = 1$ (right).

path (MPEP) (See Fig. 6.1). There are extensive theoretical predictions [112, 123] and simulation results [111, 121, 120, 122, 123] for these systems against which the algorithms developed in this chapter can be tested.

Exploring the weak noise behavior of these systems has pushed conventional simulation techniques to their limits, even for the very simple, low dimensional dynamics so far considered. A single, very long trajectory is generated, and one is obliged to wait for interesting events to occur. Therefore, it is desirable to construct a simulation that runs as quickly as possible. The state-of-the-art simulations utilize an analog electronic model of the system of interest which is then driven by a zero-mean quasi-white noise generator [120, 124]. However, such simulations cannot incorporate any importance sampling of interesting events. The total simulation time necessarily increases with the rarity of the event under study, which typically increases exponentially as the noise intensity decreases.

6.2 Path sampling Langevin dynamics*

The transition path sampling methodology has been developed to efficiently sample rare events in equilibrium systems. The main innovation is to sample path space directly using a Monte Carlo algorithm. Instead of passively waiting for the dynamics to generate an interesting trajectory, a Markov chain of trajectories is constructed, each member of which incorporates the event of interest. This path ensemble Monte Carlo is completely analogues to conventional Monte Carlo algorithms acting on configurational ensembles. A trial trajectory is generated by a small, random change in the previous trajectory; it is immediately rejected if the desired boundary conditions are not met (typically that the path starts in region A and ends in region B); and it is accepted with a probability that generates the correct distributions of trajectories.

Unfortunately, the standard methods for efficiently sampling path space can not be directly applied to nonequilibrium Langevin dynamics. Perhaps the most obvious method for generating new trajectories in a stochastic dynamics is the local algorithm [125, 59]. The path is represented by a chain of states, $\mathbf{x} = \big(x(0), x(1), \ldots, x(L)\big)$, and the probability of the path, $\mathcal{P}[\mathbf{x}]$, is written as a product of single time step transition probabilities, $P\big(x(t) \to x(t+1)\big)$;

$$\mathcal{P}[\mathbf{x}] = \rho\big(x(0)\big) \prod_{t=0}^{L-1} P\big(x(t) \to x(t+1)\big)\,. \tag{6.7}$$

*Correspondence with D. G. Luchinsky, P. V. E. McClintock and R. Mannella greatly aided development of path sampling for these driven Langevin dynamics.

Here, $\rho\big(x(0)\big)$ is the probability of the initial state of the path. A trial trajectory, $\mathbf{x}'$, is generated by changing the configuration at a single time slice, it is immediately rejected if the the desired boundary conditions are not fulfilled, and it is accepted with the Metropolis probability,

$$P_{\rm acc}(\mathbf{x}\to\mathbf{x}') = \min\Big[1, \frac{\mathcal{P}[\mathbf{x}']P_{\rm gen}(\mathbf{x}'\to\mathbf{x})}{\mathcal{P}[\mathbf{x}]P_{\rm gen}(\mathbf{x}\to\mathbf{x}')}\Big], \tag{6.8}$$

which ensures a correctly weighted ensemble of paths. Here, $P_{\rm gen(\mathbf{x}\to\mathbf{x}')}$ is the probability of generating the trial configurations $\mathbf{x}'$.

Although effective [126], the local algorithm suffers from several deficiencies, the most serious of which is that the relaxation time of the path scales as L^3, where L is the total number of time steps [127]. The driven Duffing oscillator and the Maier-Stein system require on the order of thousands of time steps to make the large, rare excursions from the stable states that are of interest, which renders the local algorithm impractical.

Several simple and efficient methods of generating trial trajectories (shooting and shifting [24]) have been developed for equilibrium dynamics. Unfortunately, they are not directly applicable to nonequilibrium dynamics, since they assume a knowledge of the initial state probability. Statistical mechanics provides simple expressions for equilibrium probabilities, but no such simple expression exists for nonequilibrium steady states.

Fortunately, there is an alternative representation of a stochastic path that admits a simple and efficient path sampling algorithm. A stochastic trajectory can be defined by the chain of states that the system visits, but it can also be represented by the initial state and the set of random numbers, the noise history, that was used to generate the trajectory. The probability of the path can then be written as

$$\mathcal{P}[\mathbf{x}] = \rho\big(x(0)\big)\prod_{t=0}^{L-1}\frac{1}{\sqrt{2\pi\epsilon}}\exp\{-\xi(t)^2/2\epsilon\} \tag{6.9}$$

where each ξ is a Gaussian random number of zero mean and ϵ variance. This is a convenient representation, since we normally generate a stochastic trajectory from a set of random numbers, and not random numbers from a trajectory.

Suppose that we have the initial state and the noise history of a relatively short path that undergoes the rare event in which we are interested. (We will return to the problem of creating this initial path shortly.) A trial path can be created by replacing the noise at a randomly chosen time step with a new set of Gaussian random numbers. This trial trajectory is accepted as a new member of the Markov chain of paths if it still undergoes the event of interest. Since high friction Langevin dynamics is highly dissipative, nearby trajectories converge rapidly, and a small change in the

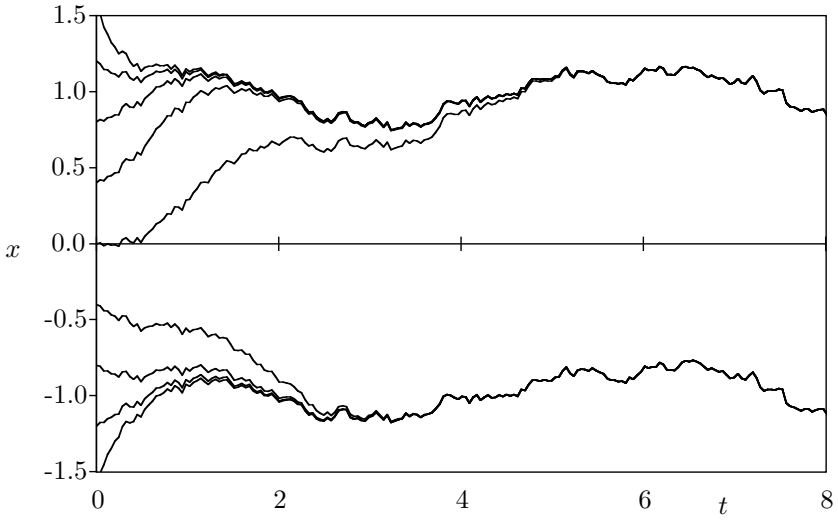


Figure 6.4: Convergence of paths in the driven Duffing oscillator with $\epsilon = 0.012$, $A = 0.264$ and $\omega = 1.2$. Each trajectory is generated with the same noise history but a different initial state. Due to the highly dissipative nature of the dynamics, all the paths rapidly converge into two trajectories, one in each metastable state.

noise generally produces a small change in the trajectory. This phenomena is illustrated in Fig. 6.4. Therefore, most trial trajectories are accepted. Only rarely does the change in the noise produce a path that no longer executes the event under study.

This noise sampling algorithm does not suffer from the poor scaling of relaxation time with path length that renders the local algorithm impractical, since a local move in noise space induces a small but nonlocal move in path space. Consider, for a moment, an unconstrained path. Then every change in the noise history is accepted. After $\mathcal{O}(L)$ moves almost all of the random numbers used to generate the path will have been replaced, and an entirely new path will have been generated, one that is uncorrelated with the previous path. Generating a trial trajectory from the noise history is very fast since the required random numbers needed to generate the path have already been created and stored. The amount of information that must be stored scales with the number of time steps, but this is a trivial amount of memory for the low dimensional systems considered here.

Unlike the local path sampling algorithm, sampling the noise history allows a choice of methods for integrating the dynamics. For compatibility

with previous digital simulations [128] we used the second order Runge-Kutta [129] method. Compared to a simple finite difference equation, this integrator is more stable and allows longer time steps. The maximum total time of the trajectories was $\tau = 16$, with a time step of $\Delta t = 1/512$, for a total of 8192 time slices. This time step is small enough to ensure better than 90% trial move acceptance rate at any one time slice for the noise intensities studied. It requires about 10 seconds of computer time (67 MHz IBM RS/6000 3CT) to generate a statistically independent path, and about a day to generate a typical data set of approximately 8000 samples. Unlike simulations (digital or analogue) without importance sampling these simulation times are largely independent of the noise intensity. There is a logarithmic increase of the transition time with decreasing noise [123], which would eventually require longer paths. The smallest noise intensities used to generate trajectories are significantly lower than the smallest values that can be practically studied with an analogue simulation, where interesting events are generated only about once a day [130].

An initial path can be generated using the following procedure. The initial point of the path is fixed, an entirely random initial noise history is generated and the end point of the corresponding trajectory is computed. A small change is then made in the noise history, and this move is accepted only if the new end point of the trajectory is closer to the desired final region than the previous path. In this manner the final point of trajectory can be dragged into the desired region, and a valid initial path obtained. It is then necessary to relax this initial path so that the correct ensemble of transition paths is generated.

A separate Monte Carlo move is used to sample the initial configuration. A trial configuration is selected from an entirely separate, non-path sampling simulation that has been relaxed to the steady state. A long trajectory ensures that the final state is largely insensitive to the initial state, and therefore that this trial move is often accepted, even if the change in the initial state is large. Alternatively, the initial state can simply be fixed at some representative point of the steady state ensemble. The simulation results will not be altered if the trajectory is significantly longer than the relaxation time of the system.

Further details and explicit implementation of this path sampling algorithm can be found in Appendix A.

6.3 Results

The driven Duffing oscillator has been extensively studied via analogue simulation [120, 111], which allows a direct test of path sampling. Results from a path sampling simulation are shown in Fig. 6.5 for exactly the same parameters and conditions previously published in Ref. [120]. Results from

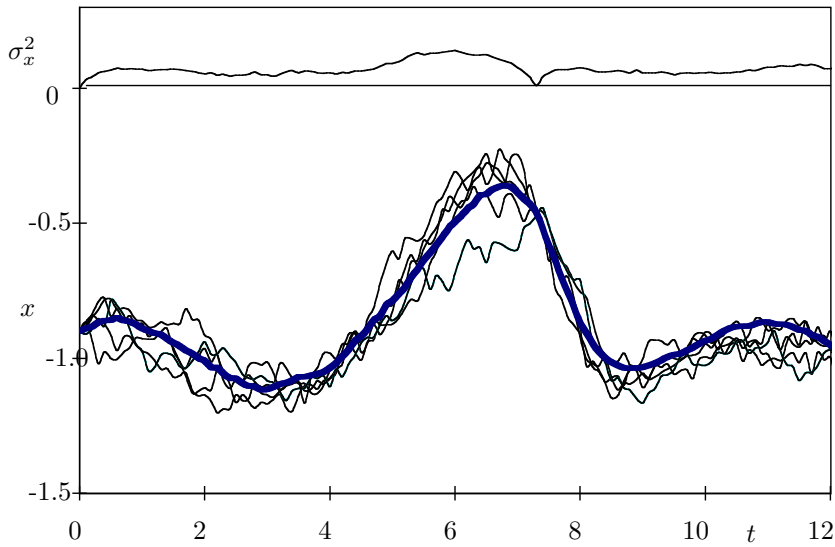


Figure 6.5: A selection of paths for the driven Duffing oscillator with $A = 0.264$, $\omega = 1.2$ and $\epsilon = 0.014$. The trajectories start at $x = -0.9$ at time $t = 0$, and make a rare excursion to the remote state $-0.478 < x < -0.442$ at time $t = 7.3$. Thin lines are individual paths, and the thick line is the average path. The upper graph presents the position variance, $\sigma_x^2(t)$, as a function of time.

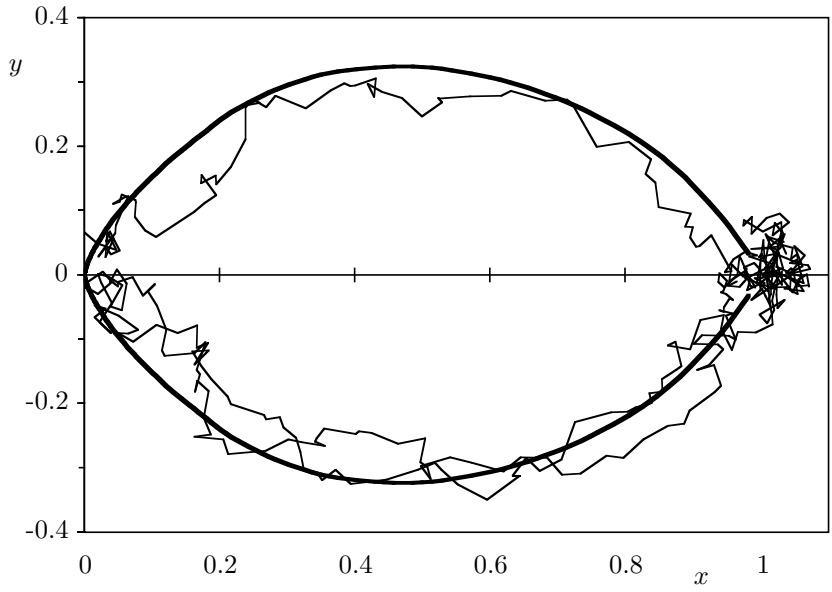


Figure 6.6: A representative sample of exit paths (thin lines) for the Maier-Stein system with $\alpha = 6.67$, $\mu = 1.0$ and $\epsilon = 0.005$, generated from a path sampling simulation. These trajectories cluster around the most probable exit paths (MPEPs) (thick lines). The MPEPs where calculated via simulated annealing of the transition paths.

path sampling and conventional simulations are indistinguishable. For this simulation, the initial point of the path was fixed. The variance of position is plotted in the upper part of the figure, and demonstrates just how quickly the system forgets its initial fixed state. The imposed excursion to a remote state at $t = 7.3$ produces a marked increase in the position variance over a much longer time scale. Finally, it should be noted that this simulation does not involve transitions between basins of attraction. Instead, the final point of the path is confined to a small window, which is located in the same basin of attraction as the initial point. Therefore, it should be possible to combine path sampling and umbrella sampling [49] to scan a window across phase space, and thereby calculate a nonequilibrium probability distribution at the chosen time slice.

The Maier-Stein system has also been extensively studied using numerical and analogue simulations [120, 122, 123]. Figure 6.3 shows several representative Maier-Stein trajectories that carry the system from the sta-

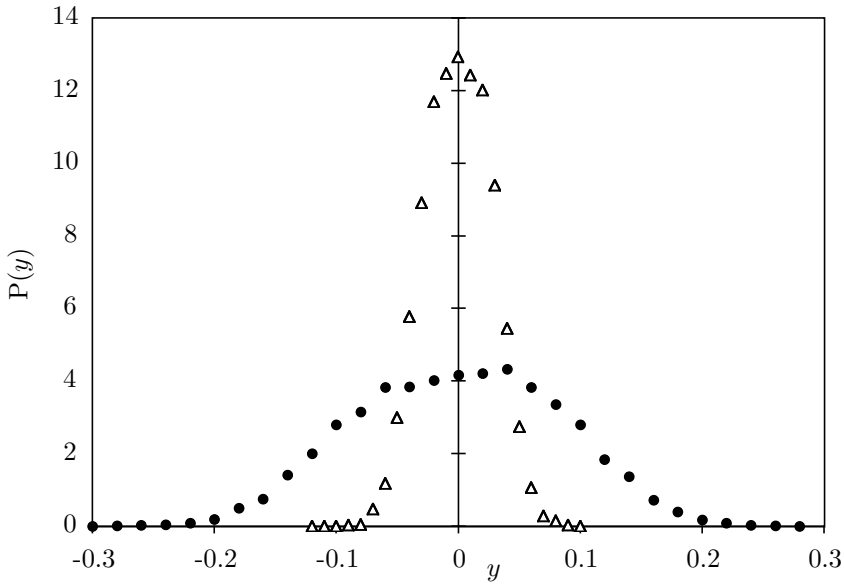


Figure 6.7: Exit location distributions for the Maier-Stein system with $\alpha = 6.67$, $\mu = 1.0$ and $\epsilon = 0.005\,(\bullet)$, or $\epsilon = 0.0005\,(\triangle)$. Data points are averages from a path sampling simulation (8192 samples). For these parameters the exit location is a relatively flat, broad distribution that has not been well characterized.

ble region around $(1, 0)$ to the separatrix at $x = 0$. For $\mu = 1.0$ and $\alpha > 4$ the set of exit paths bifurcates [108,110]. Instead of following the x axis to the transition state, trajectories instead make large excursions away from the axis, and approach the transition state from the top or bottom right quadrants. For weak noise, a single path sampling simulation of this system would lock into either the top or bottom set of trajectories and equilibration in path space would be very slow. This is analogous to the behavior of glasses and procedures developed to study such systems could be used to aid path sampling. For the current system this is not an issue, since this bifurcation is known to exist and the paths are symmetric about the x axis.

The finite noise trajectories cluster around the most probable exit paths, which are the transition paths in the zero noise limit. These can be calculated directly from theory, but here they where generated via gradually annealing the system to very weak noise intensities ($\epsilon = 10^{-5}$) [24]. The acceptance rate for parts of the path approached 0% at $\epsilon \approx 0.0005$, effectively freezing the trajectory in place. This represents the lower noise

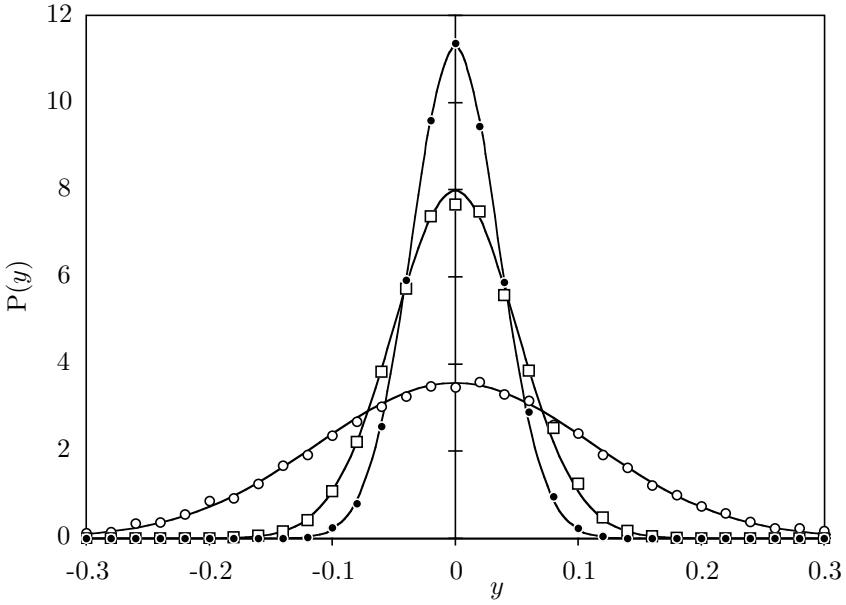


Figure 6.8: Exit location distributions for the Maier-Stein system with $\alpha = 6.67$, $\mu = 2.0$ and $\epsilon = 0.05\,(\circ)$, $0.01\,(\square)$ or $0.005\,(\bullet)$. Symbols are averages from a path sampling simulation (8192 samples) and lines are the theoretical predictions, $P(y) \propto \exp(-2y^2/\epsilon)$ [112, 122].

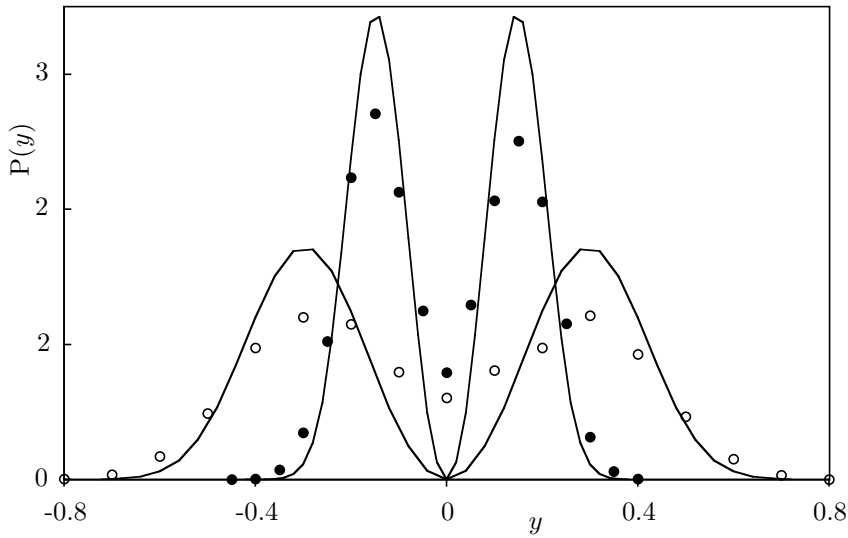


Figure 6.9: Exit location distributions for the Maier-Stein system with $\alpha = 10$, $\mu = 0.67$ and $\epsilon = 0.04\,(\circ)$ or $0.005\,(\bullet)$. Symbols are averages from path sampling simulations (8192 samples) and lines are the symmetrized Weibull distribution, $P(y) = |y|^{(2/\mu)-1} \exp(-|y/A|^{2/\mu}/\epsilon)$ [112, 123]. The parameter $A \approx 1$ is determined from the behavior of the most probable exit path near the saddle point, $y = \pm A x^{\mu}$.

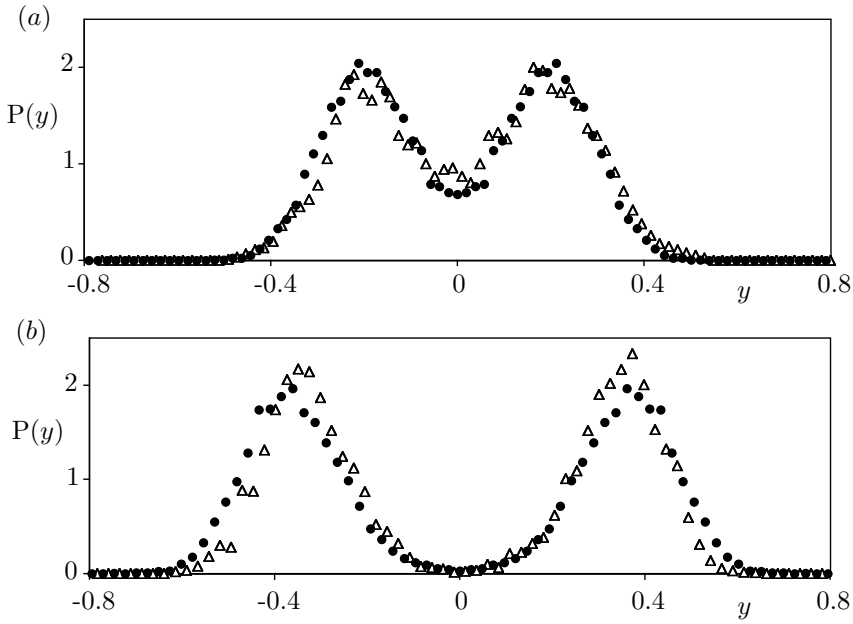


Figure 6.10: Comparison of exit location distributions generated from path sampling (•) and analogue simulation (△) at (a) $\alpha = 10$, $\mu = 0.67$ and $\epsilon = 0.011$ (b) $\alpha = 10$, $\mu = 0.2$ and $\epsilon = 0.009$. This analogue simulation data, kindly donated by D. G. Luchinsky, is the same data as was used to generate Fig. 2 of Ref. [123].

limit for the current implementation. To study weaker noise it would be necessary to use smaller time steps (which would increase the path length) or smaller changes in the noise.

There are a variety of predictions regarding the distribution of exit locations [112, 123], the point on the y axis where the transition path first crosses from one stable state to the other. Figure 6.8 shows path sampling simulation results and theoretical predictions for parameters where it is known that the theoretical predictions are accurate. Excellent agreement is observed, validating the path sampling algorithm. Figure 6.9 displays exit location distributions and theoretical predictions for parameters $(\alpha = 10, \mu = 0.67, \epsilon = 0.04)$ where the agreement between the theory and simulation is known to be poor [123]. Path sampling was used to study the exit location distribution at a noise intensity ($\epsilon = 0.005$) approximately 3 times smaller than previously possible. Even at this very low noise intensity, the agreement between theory and simulation remains unsatisfactory. Finally, Fig. (6.10) demonstrates excellent agreement between exit location distributions calculated from path sampling, and that collected from analogue experiments [123].

7. Pathways to evaporation

Will you keep to the old path that evil men have trod?
— Job 22:15 [131]

7.1 Introduction

If two macroscopic hydrophobic surfaces are brought toward one another in liquid water, a simple thermodynamic calculation indicates that the plates will destabilize the liquid phase relative to the vapor phase when they are closer than about 1000 Å [133, 134]. However, experimentally, evaporation of the liquid is not observed to occur until the plate separation is reduced below about 100 Å [132, 133]. The explanation for this discrepancy is that the liquid is metastable, and the rate of evaporation is controlled by a rare nucleation event.

This phenomena is very challenging from a computational perspective. It would be necessary to simulate at least $\frac{1}{2}$ million water molecules in a box $100 \times 400 \times 400$ Å^3. (To get the geometry correct the side length of the surfaces should be significantly greater than the intersurface separation.) This would require at least 1 CPU day of computation per picosecond on a circa 1999 workstation. (This is a very optimistic estimation, based on a linear extrapolation from a small system.) Our difficulties do not end with the large size of the system. Although evaporation at a plate separation of 100 Å is fast on a macroscopic time scale, it is still a rare event on a microscopic time scale.

The first problem, the vast size of the system, can be ameliorated by studying a much simpler model, a lattice gas with a grand canonical Monte Carlo dynamics. This model does not conserve the energy, nor the number of particles, but it does contain liquid–gas phase coexistence and surface tension.

A recent computer simulation of this lattice gas model [134,135] serves to illustrate the second difficulty, that the evaporation rate is controlled by a rare event. A $12 \times 512 \times 512$ lattice gas, confined between hydrophobic walls, was initiated in the metastable liquid state. The bulk liquid rapidly dries away from the walls, forming narrow vapor layers. The resulting liquid-vapor surfaces fluctuate, and eventually a large, rare fluctuation causes the two interfaces to touch, resulting in a vapor tube bridging the gap between

the surfaces. This event occurred after $\approx 2 \times 10^4$ Monte Carlo time steps*, and another 10^4 time steps were needed for complete evaporation of the liquid. In contrast, the important nucleation event that carried the system from one basin of attraction to another occurred in significantly less than 700 time steps.

Clearly, a direct simulation of surface induced evaporation is very inefficient. As an alternative, this chapter develops a local path sampling algorithm for this model. Since the lattice gas model of a surface confined fluid is isomorphic to the Ising model of a magnet, this work may have wider utility [137].

7.2 Path sampling the kinetic Ising model

The Ising model [138] consists of a lattice of spins, labeled by the index i. Each spin can have the values $s(i) = \pm 1$. (These values are frequently referred to as up and down, in deference to the their origin as states of magnetic spin.) The system is assumed to be in contact with a constant temperature heat bath, and the equilibrium probability of a spin configuration ν is

$$\rho_\nu = \exp\left(+\beta F + \beta H \sum_{\{i\}} s(i) + \tfrac{1}{2}\beta J \sum_{\{j:<i,j>\}} s(i)s(j) \right) . \qquad (7.1)$$

Here, J is a parameter that controls the strength of the spin-spin interactions, H is an external field that controls the relative energy of up versus down spins, F is the Helmholtz free energy, and β is the inverse temperature of the heat bath. The notation $\{j :<i, j>\}$ indicates the set of all spins j that are nearest neighbors of spin i. This set is determined by the chosen lattice. This verbose, but flexible notation is used here in anticipation of more complex situations that will arise shortly.

The Ising magnet is isomorphic to a lattice gas, a simple model of coexisting liquid-gas phases on a lattice [49]. We make the transformation by setting $n_i = (s_i + 1)/2$, where $n_i = 0, 1$ indicates whether the chosen lattice site is empty (vapor) or occupied (fluid). The equilibrium probability of a configuration ν becomes

$$\rho_\nu = \exp\left(+\beta \tilde{F} + \beta\mu \sum_{\{i\}} n(i) + 2\beta J \sum_{\{j:<i,j>\}} n(i)n(j) \right) . \qquad (7.2)$$

*The standard Monte Carlo time step is 1 attempted move for each particle or spin in the system [136].

Here, I have introduced the chemical potential $\mu = 2(H - J)$, and various constants have been absorbed into the free energy, $\tilde{F}$. Since both the temperature and the chemical potential are now fixed, this lattice gas is in a grand canonical ensemble. Arbitrarily, I will continue to speak of spins and the Ising model, instead of particles and the lattice gas.

Since the Ising model does not have any intrinsic dynamics it is necessary to construct a Monte Carlo dynamics for the system, the standard choice being Glauber dynamics [139]. A spin is picked at random and that spin is allowed to relax to equilibrium while the rest of the system is held fixed. Regrettably, it is difficult to implement path sampling for this dynamics. The chief reason is that, as with many Monte Carlo methods [59, 23], given a trajectory it is computationally expensive to calculate the probability of that trajectory.

Fortunately, Glauber dynamics is not the only possibility. We are free to pick whatever dynamics is most convenient, so long as that dynamics reproduces the fundamental physics of the problem at hand. For equilibrium properties it is sufficient that the dynamics are balanced [42] so that the correct equilibrium distribution is generated. However, for dynamical properties it is also necessary to get certain broader aspects of the physics correct [140]. Specifically, the dynamics should be homogeneous in space and time, and should be rotationally invariant for quarter-turn (on a hypercubic lattice) rotations. (This is generally sufficient to ensure full rotational symmetry at large length scales.) In other words, the dynamics must treat all spins and all lattice directions equally.

The following, particularly elegant dynamics, originally developed to study critical phenomena in 3 dimensions [141], turns out to have a tractable expression for the trajectory probability. The dynamics employs a (hyper) cubic lattice, a checkerboard update scheme, and a heat bath acceptance probability. A cubic lattice has the advantage that it can be readily generalized to an arbitrary number of dimensions, unlike, for example a 2D hexagonal lattice. Moreover, a cubic lattice is bipartite; the lattice sites can be separated into two disjoint sets such that no two sites in the same set are adjacent. In 2 dimensions this leads to a checkerboard pattern. We can refer to these disjoint sets as even and odd, depending on whether the sum of the site coordinates, $x + y + z + \cdots$, are even or odd. It should be apparent that all of the even or all of the odd spins can be subjected to a Monte Carlo move simultaneously. Thus each spin of, for example, the even set can be relaxed to equilibrium given the values of the neighboring fixed odd spins. This trial acceptance probability is referred to as a heat bath dynamics [136], and it is identical, bar a minor and almost irrelevant technicality [143], to the Glauber acceptance probability. A short, illustrative trajectory for a 1D system is shown in Fig. 7.1. Note that if periodic boundaries are used then the side length must be even to preserve

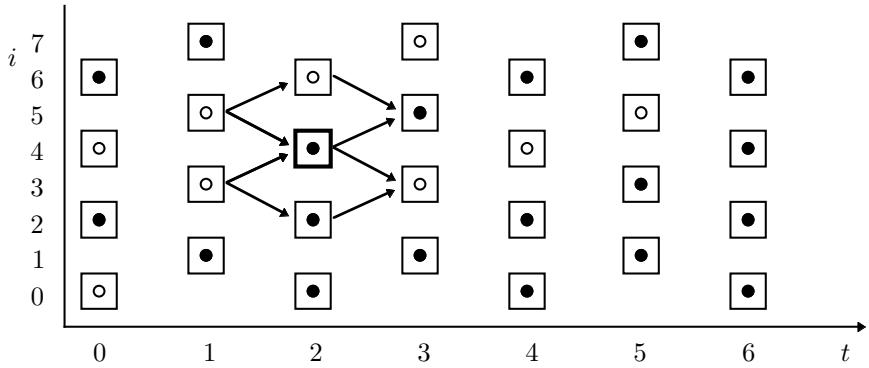


Figure 7.1: An illustrative space-time configuration of a 1D Ising model using the checkerboard/heat bath dynamics. Filled circles (•) indicate up spins, unfilled (◦) down spins. At each time slice either all of the odd or all of even spins are subjected to a Monte Carlo move. (Therefore the time units used here are 1/2 of the standard Monte Carlo unit [136].) In the figure above, only those spins that have just been updated at a particular time step are drawn. The interdependencies of spin $(2, 4)$ are shown by arrows. The state probabilities of this spin are determined by the states of spins $(1, 5)$ and $(1, 3)$, i.e., the neighboring spins in the previous time slice. The heat bath acceptance rule ensures that the state of spin $(2, 4)$ does not depend on the state of that same spin at the previous time slice, $(0, 4)$. In a path sampling simulation an entire trajectory is stored. A Monte Carlo update of the path proceeds by selecting a spin, and resetting its value with probability proportional to the probability of the resulting trajectory. Apart from the two spins in the previous time slice, the state of spin $(2, 4)$ also influences the path probability via its influence on $(3, 5)$ and $(3, 3)$, the nearest neighbor spins in the next time slice. Spin $(2, 4)$ is also indirectly coupled to spins $(2, 2)$ and $(2, 6)$ via the the single spin partition function of their common descendant in the next time slice. See Eq. (7.4).

the bipartite nature of the lattice.

The probability of a particular spin state at a particular time step, $\mathrm{P}\big(s(t,i)\big)$, is determined by the states of the neighboring spins in the previous time slice, and can be explicitly written as

$$\mathrm{P}\big(s(t,i)\big) = \frac{\exp\Big(-\beta H s(t,i) - \beta J \sum\limits_{\{j:<i,j>\}} s(t,i)s(t\text{-}1,j)\Big)}{2\cosh\Big(-\beta H - \beta J \sum\limits_{\{j:<i,j>\}} s(t\text{-}1,j)\Big)}. \tag{7.3}$$

The probability of a trajectory, $\mathbf{s}$, is given by a product of many single spin probabilities

$$\mathcal{P}[\mathbf{s}] = \rho_0 \prod_t \prod_{\{i:\mathrm{even}(t+x+y+\cdots)\}} \mathrm{P}\big(s(t,i)\big)$$

Here, ρ_0 is the probability of the initial state, and the construct even(c) is true if c is even. Thus the second product is only over the even space-time coordinates, as illustrated in Fig. 7.1.

It is now straightforward to implement a local path sampling algorithm for this dynamics. Given a trajectory, $\mathbf{s}$, a trial trajectory is chosen by choosing a spin at a particular location and time slice, and resetting it to its equilibrium value, given the surrounding space-time path. This is accomplished by calculating $\mathcal{P}[s(t,i) = +1, \mathbf{s}]/\mathcal{P}[s(t,i) = -1, \mathbf{s}]$, the ratio of the probability of the given path with $s(t,i)$ up, against the same path except that now $s(t,i)$ is down. The explicit statement of this ratio is

$$\frac{\mathcal{P}[s(t,i) = +1, \mathbf{s}]}{\mathcal{P}[s(t,i) = -1, \mathbf{s}]} = \exp\Big(+2\beta H + 2\beta J \sum_{\substack{\{j:<i,j>\},\\ \theta=\pm 1}} s(t+\theta,j)\Big) \times \prod_{\{j:<i,j>\}} \left[\frac{\cosh\Big(\beta H + \beta J \sum\limits_{\{k:<j,k>,k\neq i\}} \big(s(t,k)\text{-}1\big)\Big)}{\cosh\Big(\beta H + \beta J \sum\limits_{\{k:<j,k>,k\neq i\}} \big(s(t,k)\text{+}1\big)\Big)}\right] \tag{7.4}$$

The second term couples spin (i,t) to its next nearest neighbors in the same time slice. This is an indirect interaction, mediated by the nearest neighbor spins in the next and previous time slices. The direct influence of these nearest neighbors is encapsulated in the first term. It is interesting to note that this expression is time symmetric.

A site on a square lattice has 4 nearest and 8 next nearest neighbors in 2D. (see Fig. 7.2) Therefore, a single spin update of the trajectory requires information from 16 $(= 2 \times 4 + 8)$ spins. In 3D the total number of spins

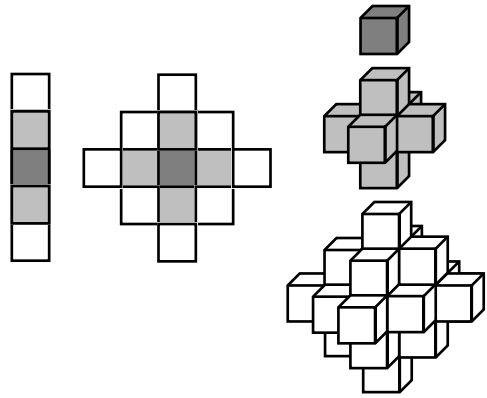


Figure 7.2: The nearest and next nearest neighbors on a hypercubic lattice in 1, 2 and 3 dimensions.

that must be examined is 28 ($= 2 \times 6 + 16$), which is large, but tractable. Fortunately it is rarely necessary to simulate the Ising model in 4 or more dimensions.

7.3 Surface induced evaporation

As a proof of principle, the path sampling algorithm developed in the previous section was used to study surface induced evaporation in 2 dimensions. The system consists of a 17×128 Ising model, with $T = 1/\beta = 1.4$, $J = 1$, and a position dependent field, H. Down spins (-1) represent vapor, and up spins ($+1$) liquid. For the bottom and top rows the field is effectively set to $H = -\infty$ so that the spins are fixed in the -1 state. This represents the hydrophobic surfaces. The remaining spins feel a small positive field, $H = 0.04$, which favors the liquid phase.

In an infinitely large system the walls would have a negligible effect, and the liquid would be the stable phase. However, under these condition the liquid between the plates is metastable. If the system is initiated in the liquid phase, then the liquid rapidly dries away from the walls, producing a vapor layer and a fluctuating liquid-gas interface. The following figure illustrates a typical configuration of this metastable state.

Black square represent vapor, and white liquid. The interfaces fluctuate, but remain close to the walls. Eventually a large fluctuation causes the

interfaces to touch, creating a vapor gap connecting the top and bottom surfaces. The vapor gap then rapidly grows and the liquid evaporates. This next figure illustrates a configuration with a large gap. With overwhelming probability, any path initiated from this configuration will rapidly relax into the stable, vapor phase.

These two states were sampled from a standard, very long Monte Carlo simulation of this system. These states where then used as initial and final states of a relatively short (256 time steps) path, and the transition regime was studied using the local path sampling algorithm. It required ≈ 0.1 seconds of computer time (on a circa 1999 workstation) for one Monte Carlo update of the trajectory. (i.e., $\approx 128,000$ single spin updates, one for each spin in the space-time configuration.) Independent paths where generated in less than 100 steps. One such path is shown in Fig. 7.3.

The transition state was located by taking each configuration of this path in turn, and shooting many trajectories from it. The transition state is the configuration from which half the trajectories undergo evaporation (defined as a vapor gap at least 8 cells wide) within 256 time steps. The transition state for the path illustrated in Fig. 7.3 occurs at $t = 112$:

It is interesting to note that in this transition state the surfaces have not quite met.

This 2D system demonstrates the practicality of path sampling, but the real interest is surface induced evaporation in 3 dimensions. By adjusting J, H and β it is possible to match the chemical potential and surface tension of water under ambient conditions [134]. Assuming (not unreasonably) that the path length used here remains sufficient, then extending the 2D model into the 3rd dimension increase the size of the system by a factor of 128. Since each spin requires only 1 bit, the total memory storage is a very modest 4 megabytes. The simulation time will increase by a larger amount due to the larger number of neighbors in 3 dimensions. However, even a conservative estimate suggests that a path update should take less than 1 minute. In short, path sampling of the full three dimensional problem [134] is eminently practical. This will allow precise characterization of the transition state of this model, as well as calculation of rate constants [23,28] for a range of temperatures and plate separations.

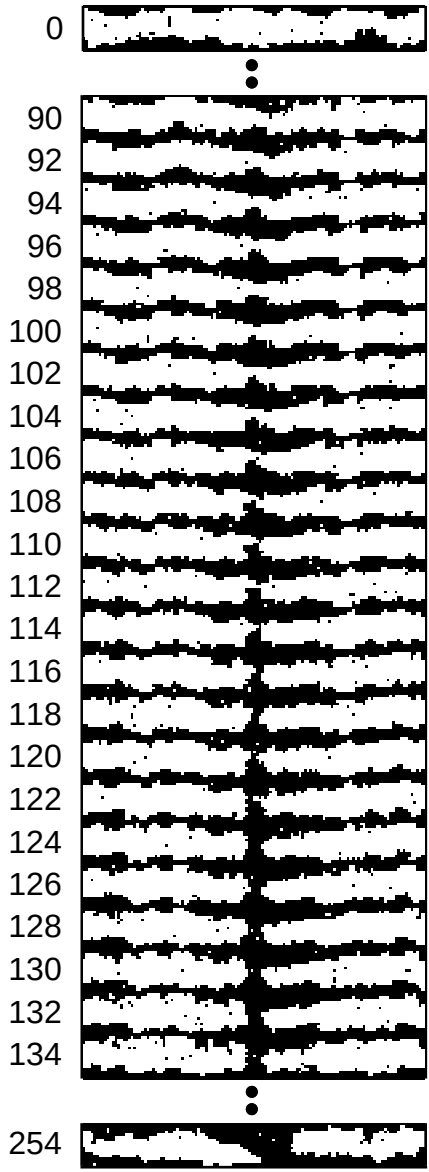


Figure 7.3: Surface induced evaporation in a 17×256 2D kinetic Ising model, with $T = 1.4$, $J = 1$ and $H = 0.04$. Down spins (-1, black cells) represent vapor and up spins ($+1$, white cells) liquid. The top and bottom rows are constrained to be vapor. This transition path was generated using the path sampling algorithm detailed in the text. The transition state occurs at $t = 112$.

8. Postscript

An alternative view is that this feeling that understanding is just a few steps away is a recurring and necessary delusion that keeps scientists from dwelling on the extent of the complexity they face and how much more remains to be discovered. — Martin Raff

A variety of exact far-from-equilibrium relations have been discussed in this thesis, all of which are derived from the principle of microscopic reversibility (1.10), and all of which can be summarized by the nonequilibrium path ensemble average, Eq. (2.1). I suspect that this relation encapsulates all of the consequences that can be obtained from the preceding principle. Very recent developments appear to support this view. Another variant of the fluctuation theorem has been proposed [144], one that I did not envisage. Yet it is contained within Eq. (2.1).

It is certainly possible that other interesting, maybe even useful, specializations of Eq. (2.1) await discovery. However, I believe that in the immediate future a more fruitful area of investigation will be the derivation, and demonstration of useful approximations to the now know exact relations, approximations that remain valid far-from-equilibrium.

Physics often proceeds by such not-entirely-rigorous approximations, the validity of which are confirmed (or repudiated) by either computer simulation or experiment. Path ensemble Monte Carlo will, in all probability, prove a useful tool for studying these approximations, as well as studying the general behavior of nonequilibrium systems.

A. Simulation of Langevin Dynamics.

For simplicity, consider a one dimensional system evolving according to the high friction Langevin equation. (See Sec. 1.6)

$$\dot{x} = F(x,t) + \xi(t), \qquad \langle \xi(t) \rangle = 0,\ \langle \xi(0)\xi(t) \rangle = \epsilon \Delta t. \tag{A.1}$$

Here, $F(x,t)$ is the systematic force and $\xi(t)$ is the stochastic force resulting from δ correlated white noise with variance ϵ. The Langevin equation can be integrated numerically by discretizing time into segments of length Δt and assuming that the systematic forces vary linearly during that time interval. This leads to the following finite difference equation [57];

$$x(t + \Delta t) = x(t) + \Delta t\, F(x,t) + \sqrt{\epsilon \Delta t}\, R(t), \tag{A.2}$$

where $R(t)$ is a Gaussian random number of unit variance.

A second order Runge-Kutta method [129] gives a more accurate and stable integrator [128], allowing larger time steps;

$$\begin{aligned} x' &= x(t) + \Delta t\, F(x) + \sqrt{\epsilon \Delta t}\, R(t)\,, \\ x'' &= x(t) + \Delta t\, F(x') + \sqrt{\epsilon \Delta t}\, R(t)\,, \\ x(t + \Delta t) &= (x' + x'')/2\,. \end{aligned}$$

This algorithm requires more computation per time step, but this is irrelevant since the computational effort is dominated by the time required to generate good pseudo-random numbers. For example, in a simulation of the Maier-Stein system (see Chapter 6) using this Runge-Kutta integrator and the `ran2` generator [145], the random number generation consumed over 90% of the simulation time.

The Gaussian random numbers where generated from the uniform distribution of `ran2` using the Box-Muller transformation [146, 57]. There is another, widely cited, method for generating Gaussian random numbers due to Knuth [147, 57]. However, on a modern computer the Box-Muller transformation is over 5 times faster than the Knuth method.

```
/*********************************************************
Standard simulation of the Maier-Stein nonequilibrium system

For background on this system, and further references see
"Irreversibility of classical fluctuations studied in
     analogue electrical circuits"
D.G. Luchinsky, P.V.E. McClintock, Nature v389 p463 (1997)

The integrator is based on example code donated by Dmitrii
G. Luchinsky.

8/98 Gavin Crooks   gavinc@garnet.berkeley.edu
                    http://gold.cchem.berkeley.edu/~gavinc
*********************************************************/

#include <stdio.h>
#include <math.h>

/* Random Number Generator.
 *  genRanGaussian() returns a double drawn from
 *  a Gaussian distribution of unit varience.
 */
#include "genRan.h"

/* System paremeters */
#define epsilon  (0.05)     /*Variance of fluctuating force*/
#define alpha    (1.)
#define mu       (1.)

#define TSTEPSPER   (512)         /*Time steps per unit time*/
#define TSTEP       (1./TSTEPSPER)/*Time per time step*/

#define SAMPLES     1000          /*Number of samples */

/******************************/
main()
{
    double  x,y;
    double  grx,gry;
    double  x2,xa,ya,xf,yf;
    int     t,samples;

    for(samples=0;samples<SAMPLES;samples++)
    {
        x=1.0;
        y=0.0;
        t=0;
```

```
        do { /*Second order Runge-Kutta integrator*/

            t++;
            grx =sqrt(epsilon*TSTEP)*genRanGaussian();
            gry =sqrt(epsilon*TSTEP)*genRanGaussian();

            x2 =x*x;
            xa = x + TSTEP*x*(1.-x2-alpha*y*y) + grx;
            ya = y - TSTEP*mu*(1.+x2)*y +gry;

            x2 = xa*xa;
            xf = x + TSTEP*xa*(1.-x2-alpha*ya*ya) + grx;
            yf = y - TSTEP*mu*(1.+x2)*ya +gry;

            x =(xf+xa)/2.;
            y =(yf+ya)/2.;

        } while(x>0.0); /* Stop when path crosses barrier */

        printf("%d\t%lg\t%lg\n",t,x,y);
        fflush(stdout);
    }
}
/*-------------------------------------------------------------*/
```

The following code implements path sampling for the Maier-Stein system, using the algorithms detailed in Chapter 6. Efficiency has been deliberately sacrificed for clarity and brevity.

```
/*************************************************************
Path Sampling Simulation of the Maier-Stein Nonequilibrium System

    8/98-9/99 Gavin Crooks  gavinc@garnet.berkeley.edu
                http://gold.cchem.berkeley.edu/~gavinc
*************************************************************/

#include <stdio.h>
#include <math.h>

/* Random Number Generator.
 *  genRanf() returns a double [0,1)
 *  genRanGaussian() returns a double drawn from
 *  a Gaussian distribution of unit varience.
 */
#include "genRan.h"

int     propagate();    /*Generate a trajectory*/
```

```
/* System paremeters */
#define epsilon     (0.05)      /*Variance of fluctuating force*/
#define alpha       (1.)
#define mu          (1.)

#define TSTEPSPER   (512)            /*Time steps per unit time*/
#define TSTEP       (1./TSTEPSPER)   /*Time per time step*/
#define TIME        (16)             /*Total time */
#define TBINS       (TIME*TSTEPSPER) /*Total time steps*/

#define SAMPLES     1000            /*Number of samples */
#define RELAX       TBINS           /*MC moves per sample */


/* These variables specify the history of the system */
double  x[TBINS];               /* x at time t  */
double  y[TBINS];               /* y at time t  */
double  grx[TBINS];             /* Random force at time t   */
double  gry[TBINS];             /* Random force at time t   */

/*******************************************************/
main()
{
    int t;                  /*time step index*/
    int transition=TBINS;   /*time step of first trasition*/
    double oldX, oldgrX, oldgrY;
    int relax, samples;

    /**** Initilisation *****/
    x[0]=1.0;
    y[0]=0.0;

    for(t=0;t<TBINS;t++) {
         grx[t] = sqrt(epsilon*TSTEP)*genRanGaussian();
         gry[t] = sqrt(epsilon*TSTEP)*genRanGaussian();
    }

    /* Generate an inital trasition path */
    while(transition==TBINS)
    {
        oldX = x[TBINS-1];
        /** Generate a trial trajectory **/
        t = (int) floor(genRanf()*TBINS);
        oldgrX = grx[t];
        oldgrY = gry[t];
        grx[t] = sqrt(epsilon*TSTEP)*genRanGaussian();
```

```
        gry[t] = sqrt(epsilon*TSTEP)*genRanGaussian();
        transition=propagate();

        /* If final point is further from the y axis than
            the previous tracectroy, reject the trial move */
        if( oldX < x[TBINS-1] )
        {
            grx[t]=oldgrX;
            gry[t]=oldgrY;
            transition=propagate();
        }
    }

    /**** Main data collection loop *****/
    for(samples=0;samples<SAMPLES;samples++)
    {
        for(relax=0;relax<RELAX;relax++)
        {
            /** Generate a trial trajectory **/
            t = (int) floor(genRanf()*TBINS);
            oldgrX = grx[t];
            oldgrY = gry[t];
            grx[t] = sqrt(epsilon*TSTEP)*genRanGaussian();
            gry[t] = sqrt(epsilon*TSTEP)*genRanGaussian();
            transition=propagate();

            /*Reject trial trajectory if there is no transition*/
            if(transition==TBINS)
            {
                grx[t] = oldgrX;
                gry[t] = oldgrY;
                transition = propagate();
            }
        }

        printf("%d\t%lg\n",transition,y[transition]);
        fflush(stdout);
    }
}

/* Generate a path using the inital state stored in x[0] & y[0],
 * and the noise history stored in grx[t] & gry[t]. Return the
 * first time step at which x[t] is negative. Else return TBINS
 */
int propagate()
{
    int t;
```

```
    int transition=TBINS;

    double x2,xa,ya,xf,yf;

    for(t=0;t<(TBINS-1);t++)
    {
        x2 =x[t]*x[t];
        xa = x[t] + TSTEP*x[t]*(1.-x2-alpha*y[t]*y[t]) + grx[t];
        ya = y[t] - TSTEP*mu*(1.+x2)*y[t] +gry[t];

        x2 = xa*xa;
        xf = x[t] + TSTEP*xa*(1.-x2-alpha*ya*ya) + grx[t];
        yf = y[t] - TSTEP*mu*(1.+x2)*ya +gry[t];

        x[t+1] =(xf+xa)/2.;
        y[t+1] =(yf+ya)/2.;

        if((transition == TBINS) && (x[t+1]<0.0)) transition=t+1;
    }
    return transition;
}
/*---------------------------------------------------------------*/
```

Beware of bugs in the above code; I have only proved it correct, not tried it. — Donald E. Knuth [148]

Bibliography

I got another quarter hundred weight of books on the subject last night. I have not read them all through.
W. Thompson, Lecture IX, p87

[1] Denis J. Evans, E. G. D. Cohen, and G. P. Morriss, "Probability of second law violations in shearing steady states" *Phys. Rev. Lett.* **71**, 2401–2404 (1993).

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

[3] G. Gallavotti and E. G. D. Cohen, "Dynamical ensembles in nonequilibrium statistical mechanics" *Phys. Rev. Lett.* **74**, 2694–2697 (1995).

[4] G. Gallavotti and E. G. D. Cohen, "Dynamical ensembles in stationary states" *J. Stat. Phys.* **80**, 931–970 (1995).

[5] Denis J. Evans and Debra J. Searles, "Causality, response theory, and the second law of thermodynamics" *Phys. Rev. E* **53**, 5808–5815 (1996).

[6] G. Gallavotti, "Chaotic hypothesis: Onsanger reciprocity and fluctuation-dissipation theorem" *J. Stat. Phys.* **84**, 899–925 (1996).

[7] E. G. D. Cohen, "Dynamical ensembles in statistical mechanics" *Physica A* **240**, 43–53 (1997).

[8] Giovanni Gallavotti, "Chaotic dynamics, fluctuations, nonequilibrium ensembles" *Chaos* **8**, 384–392 (1998).

[9] Jorge Kurchan, "Fluctuation theorem for stochastic dynamics" *J. Phys. A* **31**, 3719–3729 (1998) eprint: cond-mat/9709304.

[10] David Ruelle, "Smooth dynamics and new theoretical ideas in nonequilibrium statistical mechanics" *J. Stat. Phys.* **95**, 393–468 (1999) eprint: mp_arc/98-770.

[11] Joel L. Lebowitz and Herbert Spohn, "A Gallavotti–Cohen-type symmetry in the large deviation functional for stochastic dynamics" *J. Stat. Phys.* **95**, 333–365 (1999) eprint: cond-mat/9811220.

[12] Christian Maes, "The fluctuation theorem as a Gibbs property" *J. Stat. Phys.* **95**, 367–392 (1999) eprint: mp_arc/98-754.

[13] Gavin E. Crooks, "Entropy production fluctuation theorem and the nonequlibrium work relation for free energy differences" *Phys. Rev. E* **60**, 2721–2726 (1999) eprint: cond-mat/9901352.

[14] E. G. D. Cohen and G. Gallavotti, "Note on two theorems in nonequilibrium statistical mechanics" *J. Stat. Phys.* **96**, 1343–1349 (1999) eprint: cond-mat/9903418.

[15] Gary Ayton and Denis J. Evans, "On the asymptotic convergence of the transient and steady state fluctuation theorems" eprint: cond-mat/9903409.

[16] C. Maes, F. Redig, and A. Van Moffaert, "On the definition of entropy production via examples" eprint: mp_arc/99-209.

[17] C. Jarzynski, "Hamiltonian derivation of a detailed fluctuation theorem" eprint: cond-mat/9908286. Submitted to *J. Stat. Phys.*

[18] Gavin E. Crooks, "Path ensembles averages in systems driven far from equilibrium" eprint: cond-mat/9908420. In press *Phys. Rev. E.*

[19] Debra J. Searles and Denis J. Evans, "Fluctuation theorem for stocastic systems" *Phys. Rev. E* **60**, 159–164 (1999) eprint: cond-mat/9901258.

[20] C. E. Shannon, "A mathematical theory of communication" *Bell Syst. Tech. J.* **27**, 379–423, 623–656 (1948). Reprinted in [21] and [22].

[21] Claude E. Shannon and Warren Weaver, *The Mathematical Theory of Communication*. University of Illinois press, Urbana 1963. (Mostly a reprint of [20]. The additional material is widely regarded as superfluous. Note the subtle, but significant change in title.)

[22] Claude Elwood Shannon, *Claude Elwood Shannon : Collected papers*. IEEE Press, 1993. Edited by N. J. A. Sloane and Aaron D. Wyner.

[23] Christoph Dellago, Peter G. Bolhuis, Félix S. Csajka, and David Chandler, "Transition path sampling and the calculation of rate constants" *J. Chem. Phys.* **108**, 1964–1977 (1998).

[24] Christoph Dellago, Peter G. Bolhuis, and David Chandler, "Efficient transition path sampling: Application to Lennard-Jones cluster rearrangments" *J. Chem. Phys.* **108**, 9236–9245 (1998).

[25] Peter G. Bolhuis, Christoph Dellago, and David Chandler, "Sampling ensembles of deterministic transition pathways" *Faraday Discussion Chem. Soc.* **110**, 421–436 (1998).

[26] Félix Csajka and David Chandler, "Transition pathways in a many-body system: Application to hydrogen-bond breaking in water" *J. Chem. Phys.* **109**, 1125–1133 (1998).

[27] D. Chandler, Finding transition pathways: Throwing ropes over rough mountain passes, in the dark In B. J. Berne, G. Ciccotti, and D. F. Coker, editors, *Computer Simulation of Rare Events and Dynamics of Classical and Quantum Condensed-Phase Systems – Classical and Quantum Dynamics in Condensed Phase Simulations* pages 51–66 Singapore 1998. World Scientific.

[28] Christoph Dellago, Peter G. Bolhuis, and David Chandler, "On the calculation of reaction rate constants in the transition path ensemble" *J. Chem. Phys.* **110**, 6617–6625 (1999).

[29] Philip L. Geissler, Christoph Dellago, and David Chandler, "Chemical dynamics of the protonated water trimer analyzed by transition path sampling" *Phys. Chem. Chem. Phys.* **1**, 1317–1322 (1999).

[30] Philip L. Geissler, Christoph Dellago, and David Chandler, "Kinetic pathways of ion pair dissociation in water" *J. Phys. Chem. B* **103**, 3706–3710 (1999).

[31] Gavin E. Crooks and David Chandler, "Efficient transition path sampling for nonequilibrium stochastic dynamics".

[32] R. P. Feynman, *Statistical Mechanics: A set of lectures.* Addison-Wesley, 1972.

[33] J. Willard Gibbs, *Elementary principles in statistical mechanics, developed with especial reference to the rational foundation of thermodynamics.* Yale University press, 1902.

[34] C. Jarzynski, "Nonequilibrium equality for free energy differences" *Phys. Rev. Lett.* **78**(14), 2690–2693 (1997).

[35] Gavin E. Crooks, "Nonequilibrium measurements of free energy differences for microscopically reversible Markovian systems" *J. Stat. Phys.* **90**, 1481–1487 (1998).

[36] E. T. Jaynes, "Information theory and statistical mechanics" *Phys. Rev.* **106**, 620–630 (1957). Reprinted in [38].

[37] E. T. Jaynes, "Information theory and statistical mechanics. II" *Phys. Rev.* **108**, 171–190 (1957). Reprinted in [38].

[38] R. D. Rosenkrantz, editor *E. T. Jaynes: Papers on Probability, Statistics, and Statistical Physics*. Reidel, Dordrecht, Holland 1983.

[39] J. R. Norris, *Markov Chains*. Cambridge University Press, 1997. (Useful reference for Markov chains.)

[40] J. L. Doob, *Stochastic processes*. John Whiley, New York 1953.

[41] John G. Kemeny, J. Laurie Snell, and Anthony W. Knapp, *Denumerable Markov Chains*. Springer-Verlag, New York 2nd. edition 1976.

[42] Vasilios I. Manousiouthakis and Michael W. Deem, "Strict detailed balance is unnecessary in Monte Carlo simulation" *J. Chem. Phys.* **110**, 2753–2756 (1999).

[43] P. C. G. Vassiliou, "The evolution of the theory of non-homogeneous Markov systems" *Appl. Stoch. Model. D. A.* **13**, 59–176 (1997). (Useful reference for non-homogeneous Markov Chains.)

[44] Rudolf Clausius, *Die mechanische Warmetheorie (The mechanical theory of heat)*. F. Vieweg und Sohn, 1879.

[45] P. A. M. Dirac, "The conditions for statistical equilibrium between atoms, electrons and radiation" *Proc. Rol. Soc. A* **106**, 581–596 (1924). (Definition and discussion of the "Principle of Detailed Balance")

[46] S. R. de Groot and P. Mazur, *Non-equilibrium thermodynamics*. North-Holland, Amsterdam 1962.

[47] Richard C. Tolman, "Duration of molecules in upper quantum states" *Phys. Rev.* **23**, 693–709 (1924). (First use of the term "principle of microscopic reversibility".)

[48] Richard C. Tolman, *The Principles of Statistical Mechanics*. Oxford University Press, London 1938. pp. 163 and 165. (Early reference for the "principle of microscopic reversibility".)

[49] David Chandler, *Introduction to Modern Statistical Mechanics*. Oxford University Press, New York 1987.

[50] R. Kubo, M. Toda, and N. Hashitsume, *Statistical Physics II: Nonequilibrium Statistical Mechanics.* Springer-Verlag, New York 1985.

[51] Jerzy Luczka, Mariusz Niemiev, and Edward Piotrowski, "Exact solution of evolution equation for randomly interrupted diffusion" *J. Math. Phys.* **34**, 5357–5366 (1993).

[52] S. Carnot, *Réflexions sur la puissance motrice du feu (Reflexions on the motive power of fire).* , 1824.

[53] John H. Kalivas, editor *Adaption of simulated annealing to chemical optimization problems.* Elsevier, New York 1995.

[54] I. R. McDonald, "NpT-ensemble Monte Carlo calculations for binary liquid mixtures" *Mol. Phys.* **23**, 41–58 (1972). (First implementation of an isothermal-isobaric simulation.)

[55] Michael Creutz, "Microcanonical Monte Carlo simulation" *Phys. Rev. Lett.* **50**, 1411–1414 (1983).

[56] M. P. Langevin, *Comptes Rend. Acad. Sci. (Paris)* **146**, 530 (1908). (Original reference for Langevin dynamics)

[57] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids.* Oxford University press, 1987.

[58] F. W. Wiegel, *Introduction to Path-Integral Methods in Physics and Polymer Science.* World Scientific, 1986.

[59] Félix Stéphane Csajka, *Transition Pathways in Complex Systems.* PhD thesis University of California, Berkeley 1997.

[60] L. Onsager and S. Machlup, "Fluctuations and irreversible processes" *Phys. Rev.* **91**, 1505–1512 (1953).

[61] S. Machlup and L. Onsager, "Fluctuations and irreversible processes. II. Systems with kinetic energy" *Phys. Rev.* **91**, 1512–1515 (1953).

[62] D. R. Axelrad, *Stochastic mechanics of discrete media.* Springer-Verlag, 1993.

[63] C. Jarzynski, "Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach" *Phys. Rev. E* **56**, 5018–5035 (1997) eprint: cond-mat/9707325.

[64] C. Jarzynski, "Microscopic analysis of Clausius-Duhem processes" *J. Stat. Phys.* **96**, 415–427 (1999) eprint: cond-mat/9802249.

[65] Tomoji Yamada and Kyozi Kawasaki, "Nonlinear effects in shear viscosity of critical mixtures" *Prog. Theor. Phys.* **38**, 1031–1051 (1967). (Origional reference for the Kawasaki responce function.)

[66] Gary P. Morriss and Denis J. Evans, "Isothermal response theory" *Mol. Phys.* **54**, 629–636 (1985). (A Kawasaki responce paper.)

[67] Denis J. Evans and Gary P. Morriss, *Statistical Mechanics of Nonequilibrium Liquids.* Academic Press, London 1990. http://rsc.anu.edu.au/ evans/Book Contents.pdf.

[68] Denis J. Evans and Debra J. Searles, "Steady states, invariant measures, and response theory" *Phys. Rev. E* **52**, 5839–5848 (1995). (Bare and renormalized forms of the Kawasaki response function, and numerical confirmation.)

[69] Janka Petravic and Denis J. Evans, "The Kawasaki distribution function for nonautonomous systems" *Phys. Rev. E* **58**, 2624–2627 (1998).

[70] C. Jarzynski, "Equilibrium free energies from nonequilibrium processes" *Acta Phys. Pol. B* **29**, 1609–1622 (1998) eprint: cond-mat/9802155.

[71] Daan Frenkel and Berend Smit, *Understanding Molecular Simulation: From Algorithms to Applications.* Academic Press, 1996.

[72] C. Jarzynski, 1998. private communication.

[73] C. Jarzynski, 1997. private communication.

[74] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller, "Equation of state calculations by fast computing machines" *J. Chem. Phys.* **21**, 1087–1092 (1953).

[75] Herbert B. Callen and Theodore A. Welton, "Irreversibility and generalized noise" *Phys. Rev.* **83**, 34–40 (1951).

[76] Debra J. Searles and Denis J. Evans, "The fluctuation theorem and Green-Kubo relations" eprint: cond-mat/9902021.

[77] Rudolf Clausius, "Ueber verschiedene für die anwendung bequeme formen der hauptgleichungen der mechanischen wärmetheorie" *Annalen der Physik und Chemie* **125**, 353 (1865). (Introduction of the term entropy, and statement of Clausius's form of the second law of thermodynamics.)

[78] William Francis Magie, *A Source Book in Physics.* McGraw-Hill, 1935.

[79] C. H. Bennett, "Efficient estimation of free energy differences from Monte Carlo data" *J. Comput. Phys.* **22**, 245–268 (1976).

[80] K. K. Mon and Robert B. Griffiths, "Chemical potential by gradual insertion of a particle in Monte Carlo simulation" *Phys. Rev. A* **31**, 956–959 (1985).

[81] T. P. Straatsma and J. A. McCammon, "Computational alchemy" *Annu. Rev. Phys. Chem.* **43**, 407–435 (1992).

[82] Robert W. Zwanzig, "High-temperature equation of state by a perturbation method: I. Nonpolar gases" *J. Chem. Phys.* **22**, 1420–1426 (1954). (This is the standard reference for thermodynamic perturbation because it was the oldest paper on TP referenced in [79]. Apparently the method predates even this paper.)

[83] D. L. Beveridge and F. M. DiCapua, "Free energy via molecular simulation: Applications to chemical and biomolecular systems" *Annu. Rev. Biophys. Biophys. Chem.* **18**, 431–492 (1989).

[84] Randall J. Radmer and Peter A. Kollman, "Free energy calculation methods: A theoretical and empirical comparison of numerical errors and a new method for qualitative estimates of free energy changes" *J. Comput. Chem.* **18**, 902–919 (1997).

[85] Ryogo Kubo, "Generalized cumulant expansion method" *J. Phys. Soc. Jpn* **17**, 1100–1120 (1962).

[86] Benoit B. Mandelbrot, *The fractal geometry of nature.* W. H. Freeman, New York 1977.

[87] Melville S. Green, "Markoff random processes and the statistical mechanics of time-dependent phenomena" *J. Chem. Phys.* **20**, 1281–1295 (1952). (Linear response.)

[88] Melville S. Green, "Markoff random processes and the statistical mechanics of time-dependent phenomena. II. Irreversible processes in fluids" *J. Chem. Phys.* **22**, 398–413 (1954). (Linear response.)

[89] Ryogo Kubo, "Statistical-mechanical theory of irreversible processes, I. General theory and simple applications to magnetic and conduction problems" *J. Phys. Soc. Jpn.* **12**, 570 (1957). (Linear response)

[90] Robert Zwanzig, "Time-correlation functions and transport coefficients in statistical mechanics" *Annu. Rev. Phys. Chem.* **16**, 67–102 (1965). (Linear response)

[91] Munir S. Skaf and Branka M. Ladanyi, "Molecular dynamics simulation of solvation dynamics in methanol-water mixtures" *J. Phys. Chem.* **100**, 18258–18268 (1996). (An example of a system for which linear response does not work.)

[92] Y. Pomeau and P. Résibois, "Time dependent correlation functions and mode-mode coupling theories" *Phys. Rep.* **19**, 63–139 (1975).

[93] Kyozi Kawasaki and James D. Gunton, "Theory of nonlinear transport processes: Nonlinear shear viscosity and normal stress effects" *Phys. Rev. A* **8**, 2048–2064 (1973).

[94] Lars Onsanger, "Reciprocal relations in irreversible processes. I." *Phys. Rev.* **37**, 405–426 (1931).

[95] Lars Onsanger, "Reciprocal relations in irreversible processes. II." *Phys. Rev.* **38**, 2265–2279 (1931).

[96] Robert Graham, Onset of cooperative behavior in nonequilibrium steady states In G. Nicolis, G. Dewel, and J. W. Turner, editors, *Order and fluctuations in equilibrium and nonequilibrium statistical mechanics* pages 235–273 New York 1981. Wiley.

[97] Gregory L. Eyink, "Action principle in statistical dynamics" *Prog. Theor. Phys. Suppl.* **130**, 77–86 (1998).

[98] Yoshitsugu Oono and Marco Paniconi, "Steady state thermodynamics" *Prog. Theor. Phys. Suppl.* **130**, 29–44 (1998).

[99] William M. Visscher, "Transport processes in solids and linear-response theory" *Phys. Rev. A* **10**, 2461–2472 (1974). (Transient time correlation function.)

[100] James W. Dufty and Michael J. Lindenfeld, "Nonlinear transport in the Boltzmann limit" *J. Stat. Phys.* **20**, 259–301 (1979). (Transient time correlation function.)

[101] E. G. D. Cohen, "Kinetic-theory of non-equilibrium fluids" *Physica* **118**, 17–42 (1983). (Transient time correlation function. See also [104])

[102] Gary P. Morriss and Denis J. Evans, "Application of transient correlation functions to shear flow far from equilibrium" *Phys. Rev. A* **35**, 792–797 (1987). (Transient time correlation function)

[103] Janka Petravic and Denis J. Evans, "Nonlinear response for nonautonomous systems" *Phys. Rev. E* **56**, 1207–1217 (1997). (Transient time correlation function)

[104] E. G. D. Cohen, "Correction" *Physica* **120**, 1–2 (1983). (Of [101].)

[105] Marko Petkovsek, Herbert S. Wilf, and Doron Zeilberger, *A=B*. Wellesley, 1996.

[106] M. I. Dykman, P. V. E. McClintock, V. N. Smelyanski, N. D. Stein, and N. G. Stocks, "Optimal paths and the prehistory problem for large fluctuations in noise-driven systems" *Phys. Rev. Lett.* **68**, 2718–2721 (1992).

[107] Robert S. Maier and D. L. Stein, "Transition-rate theory for nongradient drift fields" *Phys. Rev. Lett.* **69**, 3691–3695 (1992).

[108] Robert S. Maier and D. L. Stein, "Effect of focusing and caustics on exit phenomena in systems lacking detailed balance" *Phys. Rev. Lett.* **71**, 1783–1786 (1993).

[109] Robert S. Maier and D. L. Stein, "Escape problem for irreversible systems" *Phys. Rev. E* **48**, 931–938 (1993). (Proposal of the 2-dimensional Maier-Stein system.)

[110] Robert S. Maier and D. L. Stein, "A scaling theory of bifurcations in the symmetric weak-noise escape problem" *J. Stat. Phys.* **83**, 291–357 (1996).

[111] M. I. Dykman, D. G. Luchinsky, P. V. E. McClintock, and V. N. Smelyanskiy, "Corrals and critical behavior of the distribution of fluctuational paths" *Phys. Rev. Lett.* **77**, 5229–5232 (1996). (Theoretical results and analogue simulations of the driven Duffing oscillator.)

[112] Robert S. Maier and Daniel L. Stein, "Limiting exit location distributions in the stochastic exit problem" *SIAM J. Appl. Math.* **57**, 752–790 (1997).

[113] V. N. Smelyanskiy, M. I. Dykman, and R. S. Maier, "Topological features of large fluctuations to the interior of a limit cycle" *Phys. Rev. E.* **55**, 2369–2391 (1997).

[114] Marcelo O. Magnasco, "Forced thermal ratchets" *Phys. Rev. Lett.* **71**, 1477–1481 (1993).

[115] Jong-Tae Lim and Semyon M. Meerkov, "Theory of Markovian access to collision channels" *IEEE Trans. Commun.* **35**, 1278–1288 (1987).

[116] John Ross, Katharine L. C. Hunt, and Paul M. Hunt, "Thermodynamic and stochastic theory for nonequilibrium systems with multiple reactive intermediates: The concept and role of excess work" *J. Chem. Phys.* **96**, 618–629 (1992).

[117] M. I. Dykman, Eugenia Mori, John Ross, and P. M. Hunt, "Large fluctuations and optimal paths in chemical kinetics" *J. Chem. Phys.* **100**, 5735–5750 (1994).

[118] M. I. Dykman, V. N. Smelyanskiy, R. S. Maier, and M. Silverstein, "Singular features of large fluctuations in oscillating chemical systems" *J. Phys. Chem.* **100**, 19197–19209 (1996).

[119] E. E. Sel'kov, "Self-oscillations in glycolysis" *Eur. J. Biochem.* **4**, 79–86 (1968).

[120] D. G. Luchinsky and P. V. E. McClintock, "Irreversibility of classical fluctuations studied in analogue electrical circuits" *Nature* **389**, 463–466 (1997).

[121] D. G. Luchinsky, "On the nature of large fluctuations in equilibrium systems: observation of an optimal force" *J. Phys. A* **30**, L577–L583 (1997).

[122] D. G. Luchinsky, R. S. Maier, R. Mannella, P. V. E McClintock, and D. L. Stein, "Experiments on critical phenomena in a noisy exit problem" *Phys. Rev. Lett.* **79**, 3109–3112 (1997).

[123] D. G. Luchinsky, R. S. Maier, R. Mannella, P. V. E McClintock, and D. L. Stein, "Observation of saddle-point avoidance in noise-induced escape" *Phys. Rev. Lett.* **82**, 1806–1809 (1999).

[124] D. G. Luchinsky, P. V. E. McClintock, and M. I. Dykman, "Analogue studies of nonlinear systems" *Rep. Prog. Phys.* **61**, 889–997 (1998).

[125] Lawrence R. Pratt, "A statistical method for identifying transition states in high dimensional problems" *J. Chem. Phys.* **85**, 5045–5048 (1986).

[126] Marco Paniconi and Michael F. Zimmer, "Statistical features of large fluctuations in stochastic systems" *Phys. Rev. E* **59**, 1563–1569 (1999). (Path ensemble Monte Carlo ("space-time Monte Carlo") of Langevin dynamics using a local algorithm.)

[127] D. M. Ceperley, "Path integrals in the theory of condensed helium" *Rev. Mod. Phys.* **67**, 279–355 (1995).

[128] D. G. Luchinsky, 1998. private communication.

[129] Riccardo Mannella, Numerical integration of stochastic differential equations In Luis Vàzquez, Francisco Tirando, and Ignacio Martin, editors, *Supercomputation in nonlinear and disordered systems* pages 100–129 Singapore 1997. World Scientific.

[130] P. V. E. McClintock, 1998. private communication.

[131] *The Holy Bible: New International Version.* Zondervan publishing house, 1984.

[132] Hugo K. Christenson and Per M. Claesson, "Cavitation and the interaction between macroscopic hydrophobic surfaces" *Science* **239**, 390–392 (1988).

[133] John L. Parker, Per M. Claesson, and Phil Attard, "Bubbles, cavities and the long-ranged attraction between hydrophobic surfaces" *J. Phys. Chem.* **98**, 8468–8480 (1994).

[134] Ka Lum and Alenka Luzar, "Pathway to surface-induced phase transition of a confined fluid" *Phys. Rev. E* **56**, 6283–6286 (1997).

[135] Ka Lum, *Hydrophobicity at Small and Large Length Scales.* PhD thesis University of California, Berkeley 1998.

[136] M. E. J. Newman and G. T. Barkema, *Monte Carlo methods in statistical physics.* Clarendon Press, 1999.

[137] Michael F. Zimmer, "Monte Carlo updating of space-time configurations: New algorithm for stochastic, classical dynamics" *Phys. Rev. Lett.* **75**, 1431–1434 (1995). (This paper presents a path sampling simulation of the Ising model using a local algorithm. Unfortunately, the implementation is completely wrong. An admission of this can be found at the top of page 1433, along with a kludge that probably makes things worse.)

[138] Ernst Ising, "Beitrag zur theorie des ferromagnetismus" *Z. Physik* **31**, 253–258 (1925).

[139] Roy J. Glauber, "Time-dependent statistics of the Ising model" *J. Math Phys.* **4**, 294–307 (1963). (Glauber dynamics for the Ising model.)

[140] Tommaso Toffoli, "Occam, Turing, von Neumann, Jaynes: How much can you get for how little? (A conceptual introduction to cellular automata)" *InterJournal of Complex Systems* (1994).

[141] Eytan Domany, "Exact results for two- and three-dimensional Ising and Potts models" *Phys. Rev. Lett.* pages 871–874 (1984). Also see [142].

[142] Eytan Domany, "Correction" *Phys. Rev. Lett.* **52**, 1731–1731 (1984). Of [141].

[143] Haye Hinrichsen and Eytan Domany, "Damage spreading in the Ising model" *Phys. Rev. E.* **56**, 94–98 (1997).

[144] Debra J. Searles, Gary Ayton, and Denis J. Evans, "Generalised fluctuation formula" eprint: cond-mat/9910470.

[145] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery, *Numerical Recipes in C: the art of scientific computing.* Cambridge University press, Cambridge 2nd edition 1992.

[146] G. E. P. Box and Mervin E. Muller, "A note on the generation of random normal deviates" *Ann. Math. Stat.* **29**, 610–611 (1958). (Original reference for the standard Box-Muller method for generating Gaussian random numbers. Note that there are several squares missing from the last displayed equation.)

[147] Donald E. Knuth, *The art of computer programming.* Addison-Wesley, 2nd edition 1973.

[148] Donald E. Knuth, "Notes on the van Emde Boas construction of priority deques: An instructive use of recursion" Private letter to Peter van Emde Boas. 1977. (The final sentance of this memo reads "Beware of bugs in the above code; I have only proved it correct, not tried it.")

Index