UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title
The Statistical Dynamics of Nonequilibrium Control

Permalink
https://escholarship.org/uc/item/2hj568q6

Author
Rotskoff, Grant Murray

Publication Date
2017-01-01

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

The Statistical Dynamics of Nonequilibrium Control

by

Grant Murray Rotskoff

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Biophysics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Phillip L. Geissler, Chair
Professor Carlos J. Bustamante
Professor Alan Hammond
Professor Daniel A. Fletcher

Spring 2017

The Statistical Dynamics of Nonequilibrium Control

Copyright 2017
by
Grant Murray Rotskoff

Abstract

The Statistical Dynamics of Nonequilibrium Control

by

Grant Murray Rotskoff

Doctor of Philosophy in Biophysics

University of California, Berkeley

Professor Phillip L. Geissler, Chair

Living systems, even at the scale of single molecules, are constantly adapting to changing environmental conditions. The physical response of a nanoscale system to external gradients or changing thermodynamic conditions can be chaotic, nonlinear, and hence difficult to control or predict. Nevertheless, biology has evolved systems that reliably carry out the cell's vital functions efficiently enough to ensure survival. Moreover, the development of new experimental techniques to monitor and manipulate single biological molecules has provided a natural testbed for theoretical investigations of nonequilibrium dynamics. This work focuses on developing paradigms for both understanding the principles of nonequilibrium dynamics and also for controlling such systems in the presence of thermal fluctuations.

Throughout this work, I rely on a perspective based on two central ideas in nonequilibrium statistical mechanics: large deviation theory, which provides a formalism akin to thermodynamics for nonequilibrium systems, and the fluctuation theorems which identify time symmetry breaking with entropy production. I use the tools of large deviation theory to explore concepts like efficiency and optimal coarse-graining in microscopic dynamical systems. The results point to the extreme importance of rare events in nonequilibrium dynamics. In the context of rare dynamical events, I outline a formal approach to predict efficient control protocols for nonequilibrium systems and develop computational tools to solve the resulting high dimensional optimization problems. The final chapters of this work focus on applications to self-assembly dynamics. I show that the yield of desired structures can be enhanced by driving a system away from equilibrium, using analysis inspired by the theory of the hydrophobic effect. Finally, I demonstrate that nanoscale, protein shells can be modeled and controlled to robustly produce monodisperse, nonequilibrium structures strikingly similar to the microcompartments observed in a variety of bacteria.

To my parents.

Contents

List of Figures

Acknowledgments

In the second year of my Ph.D., I had the pleasure of meeting Attila Szabo. When he learned that I was being advised by both Phill Geissler and Gavin Crooks, he told me that he thought it wise decision—"Those are two of the smartest people, and they're smart in very different ways." His comment has stuck with me because, as I reflect on my time in Berkeley, I am struck not only by how much Phill and Gavin have taught me, but also how little overlap there is between what I have learned from each.

Phill has shown me what it means to be careful. Writing it in those words trivializes what I have come to understand as a deep, ethical instinct for scientific praxis. His willingness to engage on a diverse set of topics has also afforded me tremendous freedom, which has allowed me to carve out my own distinctive set of interests. Gavin's approach to science, on the other hand, can be manic, searching always for deep connections between disparate fields. Working with Gavin has heavily influenced the topics that I intend to pursue after leaving Berkeley. I am indebted to him for keeping my perspective upbeat, the enthusiasm and wonder he brings to a problem, and his willingness to think freely.

David Chandler was among the strongest influences on me during my time at Berkeley. David taught me, among many other things scientific and otherwise, that science remains a noble endeavor. Because he believed so deeply in our enterprise, he held his colleagues to the highest possible standard, never relenting until he felt things were clear, correct, and complete. He was a singular steward in our scientific community. I will miss his erudition, his frankness, and his insight.

My colleagues played a major role in shaping my work here, as well. I had the pleasure of working quite closely with Todd Gingrich, who, in addition teaching me many things about sampling and large deviation theory, has become a great friend. Suri Vaikunanathan also guided my work and kept the office working at a high intensity with his unending enthusiasm. Eric Vanden-Eijnden, with whom I will be working after leaving Berkeley, has provided me with a much deeper technical understanding of stochastic calculus. I would also like to thank the numerous others that I've had scientific projects with over the years, Fred van Wijland, James Dama, and Ramin Khajeh, in particular. I have also benefited from conversations and interactions with Robert Jack, Hugo Touchette, and Juan Garrahan.

My friends and house mates over the years kept me entertained outside of scientific pursuits. I had a great time living with Ryan Olf, Kayte Fischer, Sydney Schreppler, Joe Gardner, and Lucy. And I remain close to a supportive group of friends, Arjun Nayini, Seth Pree, Liam Hession, and Kareem Sayegh.

My family has been loving and supportive of my academic pursuits from an early juncture. I am most grateful for their confidence in me, not now, but in 2006, when I left home for the Illinois Mathematics and Science Academy. While 2016 was a tough year, watching our family band together to help and support Samantha, Gretchen, and Nancy after the loss of Steve and Mark gave me incredible confidence in us.

And finally, I want to thank Sengyeon Lee, who has endured with me a separation of thousands of miles for, now, roughly a thousand days.

Chapter 1

The exceptional importance of rare events

The complexity of biophysical dynamics leads to a fascinating, but unmanageably diverse array of physical phenomena. While we have documented a vast collection of perplexing systems and behaviors that inspire further study, physicists are equipped with few, if any, universal guiding principles for biophysical systems. The curious importance of rare events is one of the few unifying themes that has emerged from our still-developing understanding of equilibrium dynamics in condensed matter systems. A careful assessment of rare behavior has influenced our theoretical understanding of much chemical physics: reaction rates, binding events, dynamical phase transitions, self-assembly processes, among many other topics. Rare does not mean unimportant. Events that occur extremely frequently from our perspective can be unimaginably rare relative to the timescales of typical molecular fluctuations. For example, the relaxation of a protein on the microsecond to second timescale can feature dramatic transitions between long-lived configurations, roughly 10^{12} times slower than its typical thermal motion. Studying events that span large scales in time and space requires resolution at *all* scales, a demand that has thwarted efforts to develop simple, general theories of protein folding. In the case of biophysical and chemical dynamics, we currently lack a systematic and tractable approach for analyzing the confluence of timescales implicated in many important phenomena. Even so, large deviations away from the average behavior encompass some of the most intricate and important processes in condensed matter physics.

Theoretical work on biophysical systems has largely focused on equilibrium dynamics. Protein folding and conformational heterogeneity, as prime examples, have been extensively studied using equilibrium molecular dynamics simulations and free energy calculations. Techniques for sampling rare events have aided these computational and theoretical studies immeasurably. The tools of the trade for equilibrium statistical mechanics, despite their manifold and well-documented limitations, can be robustly applied to a wide variety of models. No such toolkit exists for studying nonequilibrium dynamics.

Increasingly, statistical physicists are embracing techniques from probability theory to probe rare events in nonequilibrium systems. This is a timely and necessary effort: Many of

the cell's most important processes, even at the molecular scale, consume energy. Molecular motors and ratchets, in particular, have been a topic of intense research [1–3]. They are implicated in crucial biological functions—actin networks actively remodel membranes [4–11], kinesin and dynein proteins transport cargo through the cell [12], and viral motors package the genome for replication and subsequent infection [13].

These dissipative processes rely on directional chemical reactions, which push energy consuming proteins far away from equilibrium. This thesis focuses, both in general and in the context of specific models, on the nonequilibrium dynamics of driven systems. Studying the statistical dynamics of nonequilibrium control gives us a detailed picture of the relationship between perturbations and response when we push a system far from its equilibrium state. The general prescriptions that we offer for understanding the relationship between dissipation and control provide design principles for fabricating biomimetic machines and, at the very least, a deeper appreciation of the syzygy between fluctuations and control.

The tools developed in this thesis focus on finding paradigms for external control that minimize the work dissipated to the environment. However, the approaches taken throughout are not confined to this narrow goal. Applications to self-assembly, motivated by the final chapters of this thesis, will require adapting the ideas herein to new types of objective functions.

1.1 Nonequilibrium dynamics

Markov processes

Physical systems, on the shortest timescales, evolve deterministically. On mesoscopic timescales, the dynamics can be modeled as a random process, abstracting the chaotic interactions with the fluctuating environment into noise. Einstein was among the first to point out that dynamics can be approximated by a Markov process on timescales longer than the systems' autocorrelation time [14]. Fragments of a physical trajectory are statistically independent when the observation time is sufficiently long. A stochastic process possesses the Markov property when it lacks a memory of past events [15]. Mathematically, a Markovian system can be analyzed much more tractably than a system for which subsequent portions of a trajectory are not statistically independent.

Our ability to model a system's dynamics as a Markov process assumes a separation of timescales between molecular motions and transitions between states on mesoscopic scales. For example, a protein is buffeted by its solvent and evolves deterministically on timescales in the femtosecond regime. A large scale transition between distinct mesostates of the protein, however, might only be expected once every microsecond. This gives the system ample time to become decorrelated from its initial state through random, thermal motions. When memory of the initial state of the system is erased by chaotic collisions with the solvent, the apparent statistics become memoryless, as well.

Throughout, we will be concerned with Markov processes on a finite state space. It is convenient to represent the state space as a directed graph $G = (V, E)$, where V denotes the set of states $\{X_i\}_{i=1}^{N_v}$ and E denotes set of possible transitions between mesostates, $\{e_{ij}\}_{i,j=1}^{N_e}$, which forms of the edges of the graph. Associated with each directed edge e_{ij} is a weight W_{ij} that determines the rate of a transition from state X_j to state X_i. Physically, the weight W_{ij} represents the statistical distribution of waiting times for transitions from X_j to X_i

$$P(t_{\text{wait}} = t) \propto e^{-W_{ij}t}. \tag{1.1}$$

The assumption of Poisson statistics is a natural one for Markov processes because any process for which each transition event is statistically independent of past or future transitions must obey Poisson statistics [15]. Of course, describing a physical process in terms of pure Poisson processes is not always possible.

With the graph and transition rates in hand, we can describe the stochastic time evolution of the system, which we represent as a trajectory taking on values $X(t_i) = X_j \in V$ for times $t \in [t_0, t_k]$ and transitions at times $\{t_i\}_{i=0}^{k}$

$$P[X(t)] = p(X(t_0))p(X(t_1)|X(t_0)) \dots p(X(t_k)|X(t_{k-1})). \tag{1.2}$$

This expression takes on a simple form due to the fact that each transition depends only on the state of the system at time t_i. In principle, to learn about dynamics of the system that we are modeling, we can collect a large number of trajectories $X(t)$ and study that ensemble of trajectories.

A much more powerful approach relies on analyzing the master equation of the process. Rather than the dynamics of an single-walker trajectory, the master equation formalism describes the evolution of a vector of probabilities on V. Let $\boldsymbol{\rho}(t)$ be a vector in $\mathbb{R}^{N_v}$ with components that represent the probability density at each site. The master equation is an explicit formula for the time-evolution of this density in terms of generator W_{ij}

$$\partial_t \boldsymbol{\rho}(t) = W\boldsymbol{\rho}(t). \tag{1.3}$$

We assume that all the transition rates are strictly non-negative. The matrix W must additionally satisfy the property that

$$W_{ii} = -\sum_{i \neq j} W_{ij}, \tag{1.4}$$

meaning that we can write the time evolution of the probability density ρ_i at a single site $v_i \in V$ as

$$\partial_t \rho_i(t) = \sum_{j \neq i} W_{ij}\rho_j(t) + W_{ii}\rho_i(t). \tag{1.5}$$

The first term on the right hand side represents the flow of probability into state i, whereas the second term accounts for escape from state i. A matrix W satisfying these assumptions

we call a stochastic Metzler matrix and we note that its dynamics conserves probability. We will refer to W as a "rate matrix."

The dynamics will eventually reach a "steady state" in which the probability distribution is no longer changing, regardless of its initial value. An eigendecomposition of the master equation reveals this limiting behavior. We first write a formal solution to Eq. (1.3)

$$\boldsymbol{\rho}(t) = e^{Wt}\boldsymbol{\rho}(0), \tag{1.6}$$

noting that, in general, diagonalizing e^{Wt} will be prohibitively difficult. Expanding the matrix exponential over some short time Δt, we get a discrete time transition matrix

$$T = e^{W\Delta t} = I + W\Delta t + \mathcal{O}(\Delta t^2) \tag{1.7}$$

Because the columns of W sum to zero, it is clear that the columns of T must sum to unity. Equivalently, T has a left eigenvector with with eigenvalue 1. Because T is additionally a positive matrix, we can apply the Perron-Frobenius theorem to the formal solution, a result that states

$$\lim_{t\to\infty} \boldsymbol{\rho}(t) = \boldsymbol{\rho}^{\text{ss}}. \tag{1.8}$$

Note that in order for the theorem to apply, the dynamics must also be ergodic, i.e. each state must be accessible for every other state through a finite number of transitions. The steady state density is given by the right eigenvector with unit eigenvalue, because the propagator leaves the density invariant.

Spectral decompositions of rate matrices are not feasibly computed for many models. The dimensionality of the rate matrix W grows like the square number of vertices N_{v}^2. When the number of states is large, even computing $\boldsymbol{\rho}^{\text{ss}}$ becomes computationally infeasible. This fact should come as no surprise to condensed matter physicists, because the problem outlined here is mathematically identical to solving for the ground state wavefunction of a many-body quantum system. For a nonequilibrium dynamical system with a time-dependent rate matrix $W(t)$, this challenge is even further amplified. Studying the dynamics of complex systems requires sophisticated numerical strategies, as we will discuss in later chapters.

Diffusions

When we track the state of a system through continuous space, it becomes cumbersome to represent each state as the vertex of a graph. The configuration of a system of particles, for example, would be better represented as a coordinate $X \in \mathbb{R}^{3n}$, where n is the number of particles. The time evolution of the state of the system X_t may again satisfy the Markov property, but continuous state spaces demand a distinct formalism. The representation of the master equation (1.3) in continuous space defines the diffusion process [16].

In diffusion processes, the role of short-timescale microscopic events is abstracted into a random force that influences the dynamics of the system. The effect of the surrounding medium on the system described by the coordinate X is represented a Gaussian spectrum

of forces. The magnitude of the random force is determined by the temperature of the environment and must be consistent with the fluctuation-dissipation theorem, cf. Chapter 3.

The mathematical representation of a diffusion process involves some subtleties. We again represent the time-evolution of the system as a differential equation, akin to Eq. (1.3). However, the interpretation of the precise meaning of the equation is sensitive to a choice of stochastic calculus. A Langevin equation at inverse temperature β is a typical example:

$$dX_t = -\nabla V(X_t, t)dt + f_{\text{ext}}(X_t, t)dt + \sqrt{2\beta^{-1}}b(X_t, t)dW_t, \tag{1.9}$$

where V is the potential energy of the system, f_{ext} is an external force, $D \equiv \frac{1}{2}bb^{\text{T}}$ is a spatially and time-dependent diffusion tensor, and W_t is the Wiener process. Note that, in nonequilibrium systems, V can be time-dependent and f_{ext} need not be a conservative force.

The Wiener process should be viewed as a source of randomness. Its differential

$$dW_t = \xi(t)dt, \tag{1.10}$$

is a random process that generates Gaussian "white noise",

$$\xi(t) \sim \mathcal{N}(0, 1), \tag{1.11}$$

$$\langle \xi_i(t)\xi_j(t')\rangle = \delta_{ij}\delta(t - t'). \tag{1.12}$$

Further details of this process are outlined in Appendix A.

When $b(X_t, t)$ does not depend on X_t, Eq. (1.9) is defined unambiguously. However, if D must be evaluated at a particular value X_t, we must make a choice as to how we discretize the equation. The two most common conventions are known as the Itō and Stratonovich calculi. Let X_i denote the configuration at some point in time. After propagating the dynamics for a short time Δt, the configuration is X_{i+1}. The Itō convention is "non-anticipating" and evaluates all functions at X_i, whereas the Stratonovich interpretation uses the midpoint of the interval, $X_m = (X_{i+1} + X_i)/2$. These choices specify the following two equations

$$\text{Itō} \rightarrow X_{i+1} = X_i - \nabla V(X_i, t) + f_{\text{ext}}(X_i, t) + \sqrt{2\beta^{-1}}b(X_i, t)\Delta W, \tag{1.13}$$

$$\text{Stratonovich} \rightarrow X_{i+1} = X_i - \nabla V(X_m, t) + f_{\text{ext}}(X_m, t) + \sqrt{2\beta^{-1}}b(X_m, t)\Delta W. \tag{1.14}$$

The distinction would be largely immaterial for studying physics were it not that the choice results in distinct Fokker-Planck equations for the dynamics. If we interpret Eq. (1.9) using the Stratonovich calculus, a choice common in the physics literature, then the equivalent Itō equation has an additional drift

$$dX_t = \left(-\nabla V(X_t, t) + f_{\text{ext}}(X_t, t) + \frac{1}{2}b(X_t, t)\partial_x b(X_t, t)\right)dt + \sqrt{2\beta^{-1}}b(X_t, t)dW_t. \tag{1.15}$$

The formalism discussed here influences the definition of observables that we will use to study nonequilibrium dynamics.

Nonequilibrium observables

Detailed balance is the signature of equilibrium dynamics. An equilibrium probability distribution does not evolve in time, so each microscopic transition must satisfy

$$p(X_i)p(X_i \to X_j) = p(X_j)p(X_j \to X_i). \tag{1.16}$$

The consequence of this mathematical constraint is that probability does not flow through the system on average. For a detailed balance dynamics, we find that the probability of an individual trajectory is invariant under time-reversal

$$\frac{p(X_0)p(X_0 \to X_1)\dots p(X_{k-1} \to X_k)}{p(X_k)p(X_1 \to X_0)\dots p(X_k \to X_{k-1})} = e^{-\beta E_0 + \beta E_k - \sum_{i=1}^{k} \beta E_i - \beta E_{i-1}} \equiv 1. \tag{1.17}$$

When a system is out of equilibrium, its dynamics does not satisfy detailed balance. As a result, the trajectories that a nonequilibrium system executes are not invariant under time-reversal. To quantify the extent to which detailed balance is broken, we compare a trajectory subject to the external protocol $\Lambda(t)$ with its time-reversal. When considering the time-reversed trajectory it is extremely important that we also reverse the protocol. Then, we can define a quantity $\omega[X(t), \Lambda(t)]$, which we will later see has a deep thermodynamic meaning

$$\ln \frac{P_{\rm F}[X(t), \Lambda(t)]}{P_{\rm R}[\tilde{X}(t), \tilde{\Lambda}(t)]} \equiv \omega[X(t), \Lambda(t)], \tag{1.18}$$

where the subscripts F and R denote forward-time and reverse-time distributions of trajectories. The symbol $\tilde{\cdot}$ represents the time-reversal operation. The lack of time-reversal symmetry in a nonequilibrium system means that $\omega[X(t), \Lambda(t)]$ can be nonzero.

For a Markov process, we can write explicitly the ratio of the forward and reverse probabilities in Eq. (1.18)

$$\ln \frac{P_{\rm F}[X(t), \Lambda(t)]}{P_{\rm R}[\tilde{X}(t), \tilde{\Lambda}(t)]} = \ln \frac{p(X_0)p(X_0 \to X_1)\dots p(X_{k-1} \to X_k)}{p(X_k)p(X_1 \to X_0)\dots p(X_k \to X_{k-1})}. \tag{1.19}$$

While the trajectory as a whole need not be reversible, each microscopic transition is associated with an external reservoir to which it is coupled

$$\frac{p(X_i \to X_{i+1})}{p(X_{i+1} \to X_i)} = \frac{p(X_{i+1})}{p(X_i)}. \tag{1.20}$$

If, at the time of the transition, the system is coupled to a thermal reservoir at inverse temperature β, the logarithm of the ratio of the probabilities quantifies the amount of heat that has flowed into the system from the reservoir

$$\ln \frac{p(X_i \to X_{i+1})}{p(X_{i+1} \to X_i)} = -\beta\big(E(X_{i+1}) - E(X_i)\big) \tag{1.21}$$

The assumption, sometimes known as "local detailed balance" stipulates that energy required for a spontaneous transition is drawn from a single reservoir in the form of heat. However, over the course of the trajectory, the system might differentially be coupled to multiple reservoirs. This notion is not restricted to thermal couplings; for example, in a grand canonical ensemble, material could be exchanged with multiple chemical reservoirs.

The local detailed balance assumption relates the extent of broken detailed balance over the course of a trajectory to the heat absorbed by the system from the environment. Explicitly, the total entropy production in a single trajectory can be quantified as

$$-\beta\mathcal{Q}[X(t),\Lambda(t)] = \omega[X(t),\Lambda(t)] = \sum_{i=0}^{k-1} \ln \frac{p(X_i \to X_{i+1})}{p(X_{i+1} \to X_i)} = -\beta\big(E(X_{i+1}) - E(X_i)\big). \quad (1.22)$$

This relation, known as the Crooks Fluctuation Theorem, establishes a direct correspondence between time-reversal symmetry and entropy production in microscopically reversible dynamics [17–21]. Its importance cannot be overstated: this correspondence is the cornerstone of the modern approach to nonequilibrium statistical physics. The fluctuation theorem has been extended to many types of dynamics, including Langevin dynamics [21, 22] feedback-measurement protocols [23, 24], and transformations between nonequilibrium steady states [25–27]. It should be noted that results mathematically identical to the fluctuation theorems were independently discovered in the statistics literature and have found applications in machine learning [28].

Remarkably, the equilibrium free energy difference between two thermodynamic states can be estimated using the Crooks Fluctuation Theorem [19]. Averaging an ensemble of nonequilibrium trajectories

$$\Delta F = -\beta^{-1} \ln \left\langle e^{-\beta W} \right\rangle . \quad (1.23)$$

This relation, known as the Jarzynski equality, historically precedes—and provided an impetus for the development of—the fluctuation theorem [29].

Experimental tools cannot accurately resolve the minute energy differences necessary to measure the entropy production for an isolated trajectory. This limitation has led to systematic explorations of a wide range of alternative "dynamical" observables to characterize nonequilibrium systems. Unlike the static observables common in equilibrium systems, a dynamical observable depends on the entire history of a trajectory. Mathematically, a dynamical observable is a functional, hence the use of square brackets for the arguments to ω.

The generic form of a nonequilibrium observable for a Markov jump process can be written as a sum over all of the transitions in a dynamical trajectory

$$A[X(t)] = \sum_{i=0}^{k} A(X_i \to X_{i+1}). \quad (1.24)$$

One example is the activity, which simply counts the total number of transitions in a trajectory. Currents, which, unlike the activity resolve the direction of the dynamics on the

graph, will play a significant role in Chapter 3. The work, heat, and entropy production are also dynamical observables.

In the case of diffusions, the form of a dynamical observable is the continuous generalization of the expression for jump processes, Eq. (1.24). We accumulate the "transitions" in the value of X using a Stratonovich product, denoted by $\circ$,

$$A[X_t] = \int_0^{t_{\rm obs}} A(X_t) \circ dX_t. \tag{1.25}$$

The mathematical details of this formula can be found in Appendix A.

A unifying theme of the functionals considered here is that they grow linearly in time, on average. The "time-extensive" growth of dynamical observables underpins a thermodynamic formalism for nonequilibrium systems, based on large deviation theory [30–32]. We will make frequent use of the mathematical tools and physical insights provided by this approach throughout.

Dynamical observables other than the dissipation can also act as a signature of nonequilibrium dynamics. Currents, which vanish on average in equilibrium systems, can be a persistent feature of a nonequilibrium steady state [33]. Current-like observables have been used in recent experiments to demonstrate violations of detailed balance in biological systems [11]. In addition to currents, alternative dynamical order parameters can be used to quantify the rate of relaxation. In classical systems, work on the glass transition has relied on the activity as a measure of dynamical heterogeneity [34–37]. A similar approach may find applications in quantum systems to characterize the many-body localized phase [38, 39], with a dynamical order parameters like the entanglement growth that tracks the increase in von Neumann entropy as a function of time [40].

1.2 Large deviations

The formalism of statistical mechanics has deep and pervasive analogies to the theory of large deviations in probability. Perhaps the most direct incarnation of the inherent similarities between the two disciplines appears in the partition function. This function, which is all-but-impossible to compute for complex systems, contains complete information about the set of microscopic states and their statistical weights in the specified environment. The fundamental postulate of statistical mechanics is, in fact, a statement about the partition function: At constant energy, volume, and number density, every microscopic state is equally likely, meaning that the partition function is simply

$$Z = \sum_{\nu} 1, \tag{1.26}$$

where the sum is taken over all states of the system ν consistent with the specified thermodynamic variables. The nomenclature commonly encountered in statistical mechanics can obscure the fact that Z is simply the normalization constant of a probability distribution over

the space of all microstates. With this parallel in mind, we leverage tools from probability theory to analyze the partition function. This perspective confers mathematical advantages when studying nonequilibrium systems.

Cumulant generating functions and the partition function

The well-established theory of equilibrium statistical mechanics defines a set of statistical ensembles each of which describes a system coupled to a fixed thermodynamic environment. Consider a system with phase space coordinates $X \in \mathbb{R}^{3n}$ occupying a fixed volume in real space. Assuming that the energy is held constant and that the dynamics is ergodic, each accessible microstate is postulated to be equally likely. Under these conditions, we can define the microcanonical ensemble, which is the volume of the submanifold $U \subseteq \mathbb{R}^{3n}$ at which the energy is constant

$$Z(E, V, N) = \int d^{3n}X \ \delta(E(X) - E). \tag{1.27}$$

Physically, a system governed by microcanonical statistics must be perfectly isolated from a thermal, chemical, or barostatic environment, lest it exchange energy, volume, or particles with the external world.

Preparing isolated systems is challenging in laboratory conditions, so it is often more useful to fix an intensive parameter that sets the degree of coupling to the environment and allows fluctuations. This can be achieved by introducing a statistical bias that sets the average of one or more of the thermodynamic variables. In the case of the canonical ensemble, we write,

$$\begin{aligned} Z(\beta, V, N) &= \sum_{\nu} e^{-\beta E(\nu)}, & (1.28)\\ &= \sum_{E} e^{-\beta E} Z(E, V, N), & (1.29)\\ &\propto \left\langle e^{-\beta E} \right\rangle. & (1.30) \end{aligned}$$

The parameter $\beta^{-1} = k_{\mathrm{B}}T$ is conjugate to the energy and defines the inverse temperature up to the Boltzmann constant k_{B}.

When suitably normalized, the canonical partition function is transparently related to the cumulant generating function for the energy distribution. Let $\epsilon = E/N$ and then a function of the form

$$\psi_E(\beta, V, N) = \lim_{N\to\infty} N^{-1} \ln \left\langle e^{-N\beta\epsilon} \right\rangle, \tag{1.31}$$

is called a cumulant generating function because its nth derivative with respect to β, evaluated at $\beta = 0$, is the nth cumulant of the per-particle energy distribution. Averaging over the unnormalized distribution does not affect the computation of cumulants, e.g.,

$$\left(\frac{\partial \ln \langle e^{-\beta E} \rangle}{\partial \beta} \right)_{V,N} = \left(\frac{\partial \ln Z(\beta, V, N)}{\partial \beta} \right)_{V,N}, \tag{1.32}$$

and thus the Helmholtz free energy $A(\beta, V, N) \equiv -\beta^{-1} \ln Z(\beta, V, N)$ can be viewed as a cumulant generating function for the energy.

The complete sequence of cumulants for an observable fully characterizes its full probability distribution. If the observable, energy in the present discussion, grows extensively with the size of the system, it satisfies a large deviation principle [32], meaning that,

$$\lim_{N\to\infty} N^{-1} \ln P(\epsilon) = I(\epsilon) + o(1), \tag{1.33}$$

$$P(N\epsilon) \asymp \exp\left(-NI(\epsilon)\right). \tag{1.34}$$

The symbol $\asymp$ connotes the asymptotic equality of the terms up to logarithmic corrections. The function $I(\epsilon)$ describes the exponential rate of decay of probability away from the average and is therefore called the large deviation rate function.

We can compute the rate function directly from the cumulant generating function

$$\left\langle e^{-\beta N\epsilon} \right\rangle = \int \exp\left[-NI(\epsilon) - \beta N\epsilon\right]. \tag{1.35}$$

Because N is large, the integral can be evaluated using the Laplace saddle-point method, setting

$$\psi_E(\beta) = \inf_{\epsilon} \left\{ I(\epsilon) + \beta\epsilon \right\}, \tag{1.36}$$

we obtain

$$\left\langle e^{-\beta N\epsilon} \right\rangle = e^{N\psi_E(\beta)}. \tag{1.37}$$

Mathematically, Eq. (1.36) states that $I(\epsilon)$ is the Legendre-Fenchel transform of the scaled cumulant generating function (1.31), a result known as the Gärtner-Ellis Theorem [32].

We can carry out an analogous program for the nonequilibrium case, where thermodynamic state variables are insufficient to characterize the system. In this more general setting, we must compute partition functions over the space of trajectories of some duration t, rather than configurations. The relevant order parameters in this context are the dynamical order parameters discussed in Section 1.1. The dynamical equivalent of the microcanonical partition function, denoted $\mathcal{Z}(A, t_{\text{obs}})$, counts all accessible trajectories of duration t_{obs} with $A[X(t)] = A$.

Just as in the equilibrium case, it is often inconvenient to have a rigid constraint on the value of the dynamical observable. Performing a Laplace transformation as above, we can derive a statistical ensemble in which A fluctuates and its average is determined by the value of a statistical bias conjugate to A. For example, we can transform $\mathcal{Z}(A, t_{\text{obs}})$ to

$$\mathcal{Z}(s, t) = \int \mathcal{D}[X(t)]\ \mathcal{Z}(A[X(t)], t_{\text{obs}}) e^{-sA[X(t)]} \tag{1.38}$$

with a cumulant generating function

$$\psi_A(s) = \lim_{t_{\text{obs}}\to\infty} -{t_{\text{obs}}}^{-1} \ln \mathcal{Z}(s, t_{\text{obs}}), \tag{1.39}$$

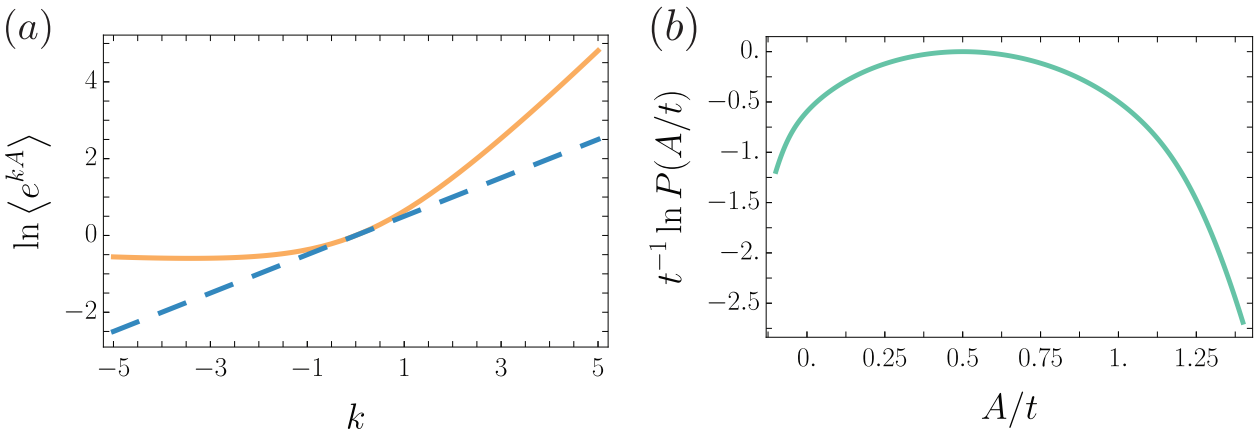

Figure 1.1: (a) A schematic of a time-scaled cumulant generating function for a dynamical observable, the slope at zero, which determines the zero value of the large deviation rate function, is shown as a blue dashed line. (b) The Legendre-Fechnel transform of the cumulant generating function, which is the large deviation rate function.

where the subscript denotes the time-extensive dynamical order parameter and the argument is a conjugate intensive parameter. In light of the analogy we have been following, $\psi_A(s, t_{\rm obs})$ is a dynamical free energy, related to the probability distribution of $A/t_{\rm obs}$ via a Legendre-Fenchel transform. A crucial observation is that the large parameter in the nonequilibrium setting is the duration of a trajectory, not the number of particles. While a thermodynamic free energy requires an $N \to \infty$ limit to rigorously show a phase transition, the dynamical free energy can develop singularities with a finite spatial extent.

1.3 Fluctuations and response away from equilibrium

The mathematical approach we have laid out establishes a natural equivalence between fluctuations and the response to an externally tuned field. If we view the parameter s as an external field conjugate to the dynamical observable A, then we can compute the change in the expectation of A to a change in s as

$$\frac{\partial \langle A \rangle}{\partial s} = \frac{\partial^2 \ln \mathcal{Z}(s, t_{\rm obs})}{\partial s^2}. \tag{1.40}$$

The response to small changes in s is encoded in the dynamical free energy $\psi_A(s, t_{\rm obs})$, reinforcing the analogy with equilibrium statistical mechanics. Because $\psi_A(s, t_{\rm obs})$ is a cumulant generating function, its higher derivatives provide information about the nonlinear response to changes in s. When evaluated at $s = 0$, the second derivative of the logarithm of the dynamical partition function is the variance of A

$$\frac{\partial^2 \ln \mathcal{Z}(s, t_{\rm obs})}{\partial s^2} = \langle A^2 \rangle - \langle A \rangle^2. \tag{1.41}$$

This asserts the correspondence between dynamical fluctuations and response.

The field s can seem abstract because it is a time-nonlocal parameter that might be difficult to tune physically. Nevertheless, like the statistical biases commonly used in standard free energy calculations [41], the fictitious field plays an important role in our effort to study rare features of the trajectory ensemble. In order to compute the scaled cumulant generating function, we construct an effective process in which rare values of the dynamical order parameter A become typical. We carry this out by imposing a statistical bias on A throughout the dynamics, which can be accomplished by choosing a non-zero value of s, and then appropriately reweight the biased trajectories.

Mathematically, we can view this procedure as a change of measure on the space of trajectories. Suppose that we would like to collect trajectories for which some value $A^* \neq \langle A \rangle$ is typically observed. To do so, we construct a probability measure on the space of trajectories $P^*[X(t)]$ in which A^* is the average value of $A[X(t)]$, that is,

$$\int \mathcal{D}[X(t)]\ P^*[X(t)]A[X(t)] = A^*. \tag{1.42}$$

To compute the probability of the rare value A^*, we use the standard trick of importance sampling

$$P(A^*) = \int \mathcal{D}[X(t)]\ \frac{P[X(t)]}{P^*[X(t)]} P^*[X(t)]\delta(A[X(t)] - A^*). \tag{1.43}$$

While the identity above seems trivial, it is of significant practical consequence both computationally and theoretically. The simplest demonstration of that advantage is that it might be much easier to draw trajectories with $A[X(t)] = A^*$ from the biased distribution, which will help reduce the variance an estimate of $P(A^*)$.

In order to bias the dynamics, we exponentially "tilt" the path measure

$$P^*[X(t)] = P[X(t)]e^{-sA[X(t)]} \equiv P_{\text{tilt}}[X(t)]. \tag{1.44}$$

The generator associated with the tilted measure does not conserve probability. In the case of a Markov jump process, the tilted generator, which we denote $W(s)$, is related to the rate matrix W of the unbiased process. In particular, we just bias the transition rates with a weight proportional to the contribution a transition from j to i makes to the value of the observable A

$$W_{ij}(s) = W_{ij}e^{-sA(X_j \to X_i)}. \tag{1.45}$$

We can justify the choice of an exponential bias with a simple calculation. The minimally invasive way of biasing the distribution will minimize the Kullback-Liebler divergence between the original and the distribution in which a rare value A^* is typical

$$D_{\text{KL}}\big(P^*[X(t)] \| P[X(t)]\big) = \int \mathcal{D}[X(t)]\ P^*[X(t)] \ln \frac{P^*[X(t)]}{P[X(t)]}. \tag{1.46}$$

By abuse of language, the divergence is sometimes called a distance because it, roughly, computes the overlap or distance between the two distributions. It is not, however, a distance in the mathematical sense of the word.

To see the emergence of an exponential bias, we minimize the divergence over all functions $P^*[X(t)]$ while enforcing a normalization constraint and a constraint on the average,

$$1 = \int \mathcal{D}[X(t)]\ P^*[X(t)], \tag{1.47}$$

$$A^* = \int \mathcal{D}[X(t)]\ A[X(t)]P^*[X(t)]. \tag{1.48}$$

This calculation can easily be carried out using the method of Lagrange multipliers

$$\begin{aligned} 0 = \frac{\delta}{\delta P^*[X(t)]}\Bigg[\alpha\left(\int \mathcal{D}[X(t)]\ P^*[X(t)] - 1\right) \\ - s\left(\int \mathcal{D}[X(t)]\ A[X(t)]P^*[X(t)] - A^*\right) \\ - D_{\mathrm{KL}}\big(P^*[X(t)]\|P[X(t)]\big)\Bigg]. \end{aligned} \tag{1.49}$$

The result of the functional minimization is

$$P^*[X(t)] = P[X(t)]e^{-sA[X(t)]}e^{\alpha}, \tag{1.50}$$

and the optimal value of s is given by the Legendre transform

$$s_{\mathrm{opt}} = \mathrm{argmin}_s\left(sA^* + \ln\int \mathcal{D}[X(t)]\ P[X(t)]e^{sA[X(t)]}\right). \tag{1.51}$$

The maximization defining the optimal value of s, Eq. (1.51), if evaluated at s_{opt} is the Legendre-Fenchel transform $I(A^*)$ of the cumulant generating function $\ln\langle\exp(-sA)\rangle$.

The long-time behavior of dynamics generated by $W_{ij}(s)$ can be assessed by studying the spectrum of the tilted generator. The maximum eigenvalue of $W_{ij}(s)$, which we denote $\boldsymbol{\lambda}(s)$, is related to the cumulant generating function for the asymptotic statistics

$$\boldsymbol{\lambda}(s) = \lim_{t_{\mathrm{obs}}\to\infty}\frac{1}{t_{\mathrm{obs}}}\ln\langle e^{-sA/t_{\mathrm{obs}}}\rangle. \tag{1.52}$$

We know from the Gärtner-Ellis theorem that $\boldsymbol{\lambda}(s)$ is related to the large deviation function $I(A/t_{\mathrm{obs}})$ by a Legendre-Fenchel transform, as depicted in Fig. 1.1 and given by Eq. (1.36).

Hence, the tilted path measure gives us a means to compute large deviation function for an arbitrary dynamical observable. However, the bias that we have added to the generator makes the process unphysical in the sense that probability is no longer conserved [42]. This observation led Chetrite and Touchette [42] to consider a generalization of Doob's h-transform that produces a conservative generator for the rare dynamics. We call the generator of the "driven process" $\mathbb{W}(s)$. In the context of importance sampling rare trajectories, knowing this generator amounts to having the "perfect" importance sampler: no reweighting

is needed because each sample is generated without any statistical bias. Nevertheless, there is a significant caveat. The h-transform requires knowledge of the right-eigenvector $\boldsymbol{r}(s)$ with maximum eigenvalue, because the unique form of the generator for the driven process is

$$\mathbb{W}_{ij}(s) = \boldsymbol{r}_i^{-1}(s) W_{ij}(s) \boldsymbol{r}_j(s). \tag{1.53}$$

Thus, computing $\boldsymbol{r}(s)$ for each value of s is mathematically equivalent to computing $\psi_A(s)$.

While computations using these techniques remain challenging, large deviation theory has become an important tool for extending statistical mechanics beyond the near-equilibrium approximations that dominated the field for most of the 20th century [43–47]. Its connection with thermodynamic formalism makes it a natural language for connecting fluctuations in dynamical order parameters with response. We will employ this perspective frequently in what is to come.

Chapter 2

Large deviations in thermodynamic efficiency

So you would be violating the Second Law of Thermodynamics, getting something for nothing...
Thomas Pynchon. The Crying of Lot 49.

As engineering capabilities reach molecular scales, design principles must account for the large fluctuations inherent in the behavior of nanoscale machines [24, 48–51]. Familiar thermodynamic quantities, such as the heat absorbed, work extracted, and efficiency do not realize a single value for small scale machines, and require the use of stochastic thermodynamics [18, 52–54]. Instead, each quantity varies from one measurement to the next according to a probability distribution; rare excursions from the average operation of the device can have dramatically counterproductive consequences. In order to analyze the thermodynamic efficiency in this regime, an understanding of the statistics of the relevant quantities is crucial. This section uses material from Gingrich, Rotskoff, Vaikuntanathan, and Geissler [55].

Knowledge of the fluctuations in thermodynamic efficiency, a measure of the work extracted per amount of heat absorbed defined by Carnot as $\eta = -\mathcal{W}/\mathcal{Q}_{\rm h}$, provides a powerful viewpoint from which to assess a nanoscale machine [56]. Verley *et al.* employed large deviation theory to highlight some surprising features of the statistical properties of a fluctuating efficiency [57]. They argued that in a long-time limit the efficiency distribution would attain the form $P(\eta) \sim e^{t_{\rm obs} J(\eta)}$, where $t_{\rm obs}$ denotes the observation time, and that the large deviation rate function, $J(\eta)$ has a global minimum at the "reversible" efficiency for time-symmetric driving. At this efficiency, sufficiently long trajectories produce no entropy. In the case of a heat engine, the reversible efficiency is the Carnot efficiency $\eta_{\rm C} = 1 - T_{\rm c}/T_{\rm h}$, the maximum efficiency that can be attained by a heat engine on average [56, 58]. Here, we address the more practical case of devices operating under time-asymmetric non-equilibrium protocols. Standard engine protocols such as the Carnot cycle and the Stirling cycle are time asymmetric for both macroscopic and microscopic realizations [59].

For a two-level model of a nanoscale heat engine, we examine the statistics and large deviation scaling for the work, heat, and thermodynamic efficiency. By computing a large deviation rate function for joint observations of $\mathcal{W}$ and $\mathcal{Q}_{\rm h}$ we in turn calculate the long-time behavior of the probability distribution for η [32]. In particular, we determine the rate function for η, which resembles rate functions analyzed in [57]. However, we show that time-asymmetric driving shifts the location of the minimum away from the Carnot efficiency. As such, the result of Verley et al. [57] does not generalize to encompass common engine protocols. Furthermore we note an important caveat pertaining to the relationship between the probability distribution of η and its asymptotic representation as $e^{t_{\rm obs}J(\eta)}$. As a consequence the efficiency distribution sampled over long but finite times may not reveal the minimum predicted by an analysis of the efficiency rate function. We obtain the form of these finite-time distributions, which we expect to be significantly more relevant to the understanding of efficiency statistics in experiments.

We aim to explore generic features of efficiency distributions in non-equilibrium engines, but physical systems can be held out-of-equilibrium in two fundamentally different ways. A system can be out-of-equilibrium in a time-independent manner if it is held in contact with multiple reservoirs maintained at different thermodynamic conditions (e.g. two unequal temperature baths can induce a temperature gradient across a system). Verley et al. constructed a model of one such system, photoelectric cells, for which the temperature of Earth served as one bath and the temperature of the Sun as another [57]. The non-equilibrium protocol driving these systems is time independent and therefore time-reversal symmetric, which we will show has significant implications for the efficiency distribution. More generally, systems can be out of equilibrium if they are driven in a time dependent manner, which need not be a time-symmetric protocol.

We investigate this scenario by constructing a two-state model of a stochastic engine. The temperature and energy levels are varied cyclically through four consecutive stages (see Fig. 2.1),

$$\text{Stage 1: } T^{(1)} = T_{\rm c}, E_{\rm L}^{(1)} = 0, E_{\rm R}^{(1)} = 0$$

$$\text{Stage 2: } T^{(2)} = T_{\rm c}, E_{\rm L}^{(2)} = 0, E_{\rm R}^{(2)} = -\Delta E$$

$$\text{Stage 3: } T^{(3)} = T_{\rm h}, E_{\rm L}^{(3)} = 0, E_{\rm R}^{(3)} = -2\Delta E$$

$$\text{Stage 4: } T^{(4)} = T_{\rm h}, E_{\rm L}^{(4)} = 0, E_{\rm R}^{(4)} = -\Delta E,$$

with $E_{\rm L}$ and $E_{\rm R}$ the energies of the left and right states. $T_{\rm h}$ and $T_{\rm c}$ are the high and low temperatures respectively achieved by alternately coupling the two-state system to hot or cold baths. The superscript on temperatures and energies acts as an index for the stage of the cycle.

We carry out each cycle in time τ, with each stage lasting for $\tau/4$ units of time. During each stage the particle can hop between the left and right states with Arrhenius rates given

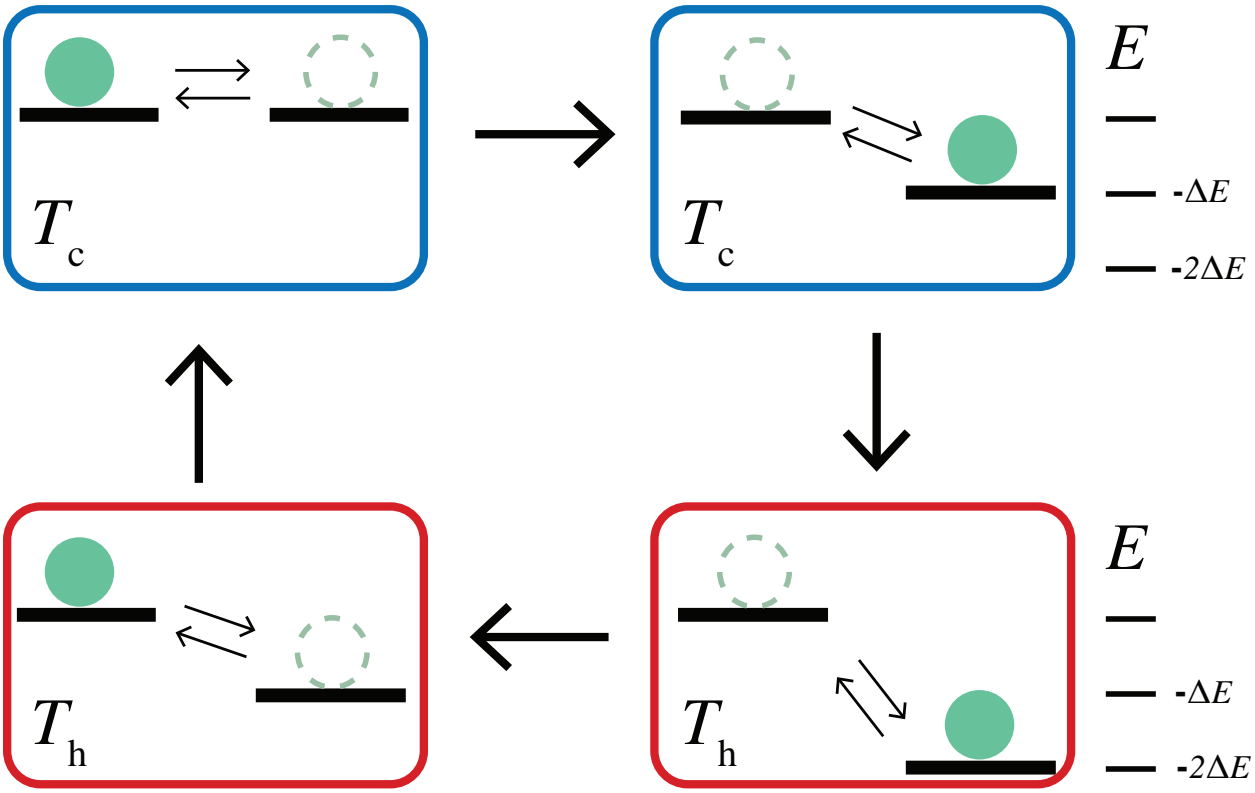

Figure 2.1: A single particle can occupy one of two energy levels with thermal transitions between the states. The energy of the right state, E_{R} and the temperature, T_{h} or T_{c} are instantaneously switched in a cycle among four stages. These switches occur in multiples of $\tau/4$, where τ is the period of the cycle.

by a tunable barrier height, B. The continuous time rate matrix for the i^{th} stage of the cycle is therefore

$$\mathbb{W}^{(i)} = \begin{pmatrix} -e^{-\beta^{(i)}B} & e^{-\beta^{(i)}(B-E_{\mathrm{R}}^{(i)})} \\ e^{-\beta^{(i)}B} & -e^{-\beta^{(i)}(B-E_{\mathrm{R}}^{(i)})} \end{pmatrix}, \tag{2.1}$$

with $\beta^{(i)} \equiv (k_{\mathrm{B}}T^{(i)})^{-1}$ and the Boltzmann constant, k_{B}, set equal to unity throughout.

Work is extracted from the system when the right energy level is occupied while being instantaneously lowered. Each transition between the energy levels requires heat absorbed from the reservoir equal to the energy difference between the levels. We adopt the conventions that positive heat flow corresponds to heat flowing into the system and that positive work is performed on the system [18, 29]. The simplicity of our four stage, two-state model lends itself to formal analysis of these fluctuating quantities as well as exhaustive computational study.

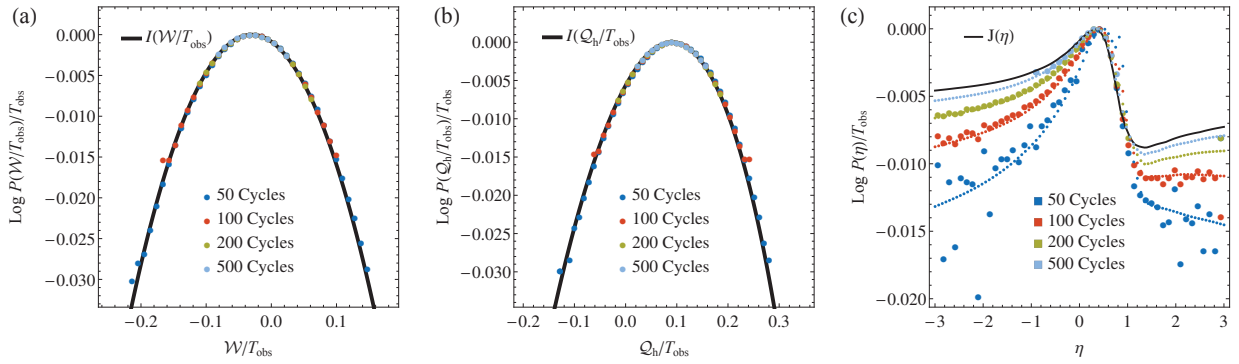

Figure 2.2: (a) Work, (b) heat, and (c) efficiency sampling from 3×10^7 trajectories of the four stage, two-state engine with $\Delta E = 2.375, T_{\rm c} = 2, T_{\rm h} = 14, B = 0.05$, and period $\tau = 10$. The large deviation rate functions are shown as a solid black line. Histograms of sampled values are shown as large circles. The small circles in (c) plot the result of integrating out ΔS in Eq. (2.16).

2.1 Simulations

While the particle dynamics occurs in continuous time, the heat and work are solely functions of the position of the particle at the end of a stage. We advance time in units of $\tau/4$ by drawing the state at the end of each stage in proportion to its probability, computed using the matrix exponential $\mathbb{T}^{(i)} = e^{\tau \mathbb{W}^{(i)}/4}$. We collected statistics on the work extracted, heat absorbed, and efficiency of steady state stochastic trajectories evolved over many repeated engine cycles for a choice of parameters for which the protocol acted as an engine: $\Delta E = 2.375, T_{\rm c} = 2, T_{\rm h} = 14, B = 0.05$, and $\tau = 10$. The histograms, shown in Fig. 2.2, illustrate that the system acts as engine on average — the mean work supplied to the system is negative while mean heat absorbed from the hot bath is positive, resulting in a positive mean efficiency. Furthermore, direct sampling of trajectories with varying lengths, $t_{\rm obs}$, numerically illustrates that the heat and work distributions tend to a large deviation form [21, 32, 60]. The probability of observing a total work $\mathcal{W}$ and total heat absorbed from the hot bath $\mathcal{Q}_{\rm h}$ in a trajectory of length $t_{\rm obs}$ can thus be written as

$$p(\mathcal{W}, \mathcal{Q}_{\rm h}) \sim e^{t_{\rm obs} I(\mathcal{Q}_{\rm h}/t_{\rm obs}, \mathcal{W}/t_{\rm obs})}, \tag{2.2}$$

where $\sim$ denotes an equality in the asymptotic limit and $I(\mathcal{Q}_{\rm h}/t_{\rm obs}, \mathcal{W}/t_{\rm obs})$ is the large deviation rate function. The large deviation scaling is robust even for a modest $t_{\rm obs}$ of only 50 engine cycles. The heat and work statistics of trajectories evolved for at least 50 engine cycles can therefore be well-described by a large deviation rate function.

The statistics of efficiency are similarly shown in Fig. 2.2(c) along with the efficiency rate function, $J(\eta)$. The sampled distributions for our finite time measurements show no minimum. As expected, the efficiency distribution tends toward $P(\eta) \sim e^{t_{\rm obs} J(\eta)}$ at long times,

but in practice even a large $t_{\rm obs}$ is not sufficient for the minimum of the large deviation rate function to be relevant to the efficiency distribution. In marked contrast to the statistics of heat and work, the large deviation form is not predictive of the sampled efficiency distributions for the reported values of $t_{\rm obs}$.

2.2 Large deviation rate functions for $\mathcal{W}$, $\mathcal{Q}_{\rm h}$,and η

The large deviation function, $I(\mathcal{W}/t_{\rm obs}, \mathcal{Q}_{\rm h}/t_{\rm obs})$, can be calculated using standard methods of large deviation theory [21, 32]. We introduce two fields, $\lambda_{\mathcal{W}}$ and $\lambda_{\mathcal{Q}_{\rm h}}$ and construct a scaled cumulant generating function for $\mathcal{W}$ and $\mathcal{Q}_{\rm h}$,

$$\psi(\lambda_{\mathcal{W}}, \lambda_{\mathcal{Q}_{\rm h}}) = \lim_{t_{\rm obs}\to\infty} \frac{1}{t_{\rm obs}} \log\langle e^{-\lambda_{\mathcal{W}}\mathcal{W} - \lambda_{\mathcal{Q}_{\rm h}}\mathcal{Q}_{\rm h}}\rangle. \tag{2.3}$$

Applying a saddle point approximation, which is exact in the long-time limit, reveals that $I(\mathcal{W}/t_{\rm obs}, \mathcal{Q}_{\rm h}/t_{\rm obs})$ can be obtained from $\psi(\lambda_{\mathcal{W}}, \lambda_{\mathcal{Q}_{\rm h}})$ by a Legendre transform [1]. In the long-time limit the scaled cumulant generating function can be found as a maximum eigenvalue of the appropriate "tilted operator," which for our model must involve a product of tilted operators stemming from each stage of the engine [32].

Recall that work is performed instantaneously between stages and only if the particle is in the right state, so we define the tilted operator

$$\mathbb{T}^{(i)}_{\mathcal{W}}(\lambda_{\mathcal{W}}) = \begin{pmatrix} 1 & 0 \\ 0 & e^{\lambda_{\mathcal{W}}\Delta E^{(i)}} \end{pmatrix}, \tag{2.4}$$

whose derivatives with respect to $\lambda_{\mathcal{W}}$ provide statistical information about the work. $\Delta E^{(i)}$ denotes the change in $E_{\rm R}$ between stages of the protocol. The heat absorbed from the hot bath differs from the entropy production only by a factor of $\beta_{\rm h}$, so the tilted rate matrix for heat absorbed during those stages is analogous to the entropy production tilted operator of Lebowitz and Spohn [21]. The matrix elements are given by

$$\mathbb{W}^{(i)}_{\mathcal{Q}_{\rm h}}(\lambda_{\mathcal{Q}_{\rm h}})_{jk} = \begin{cases} \mathbb{W}^{(i)}_{jk}, & j = k \\ \mathbb{W}^{(i)}_{jk})^{T_{\rm h}\lambda_{\mathcal{Q}_{\rm h}}}(\mathbb{W}^{(i)}_{kj})^{1-T_{\rm h}\lambda_{\mathcal{Q}_{\rm h}}}, & j \neq k. \end{cases} \tag{2.5}$$

The tilted rate matrix propagates in continuous time, but it is convenient to also define a tilted operator which accounts for the complete stage, a time of $\tau/4$.

$$\mathbb{T}^{(i)}_{\mathcal{Q}_{\rm h}}(\lambda_{\mathcal{Q}_{\rm h}}) = \exp\left(\frac{\tau}{4}\mathbb{W}^{(i)}_{\mathcal{Q}_{\rm h}}(\lambda_{\mathcal{Q}_{\rm h}})\right) \tag{2.6}$$

[1] This assumes an absence of dynamic phase transitions. Otherwise the Legendre transform returns a convex hull of the rate function.

We now construct a tilted operator, $\mathbb{T}^{cycle}$, for the entire cycle by forming a matrix product of the tilted operators for each stage (and each transition between stages).

$$\mathbb{T}^{cycle}(\lambda_{\mathcal{W}}, \lambda_{\mathcal{Q}_\mathrm{h}}) = \mathbb{T}^{(4)}_{\mathcal{W}}(\lambda_{\mathcal{W}})\mathbb{T}^{(4)}_{\mathcal{Q}_\mathrm{h}}(\lambda_{\mathcal{Q}_\mathrm{h}})\mathbb{T}^{(3)}_{\mathcal{W}}(\lambda_{\mathcal{W}})\mathbb{T}^{(3)}_{\mathcal{Q}_\mathrm{h}}(\lambda_{\mathcal{Q}_\mathrm{h}})\mathbb{T}^{(2)}_{\mathcal{W}}(\lambda_{\mathcal{W}})\mathbb{T}^{(2)}\mathbb{T}^{(1)}_{\mathcal{W}}(\lambda_{\mathcal{W}})\mathbb{T}^{(1)} \tag{2.7}$$

Note that because we do not record heat absorbed from the cold bath, the cold stages involve time propagators rather than tilted operators for $\mathcal{Q}_\mathrm{h}$. In the long-time limit, repeated application of $\mathbb{T}^{cycle}$ is well-described by its maximum eigenvalue, which we denote as $\nu(\lambda_{\mathcal{W}}, \lambda_{\mathcal{Q}_\mathrm{h}})$. The scaled cumulant generating function can then be expressed as

$$\psi(\lambda_{\mathcal{W}}, \lambda_{\mathcal{Q}_\mathrm{h}}) = \frac{1}{\tau}\log \nu(\lambda_{\mathcal{W}}, \lambda_{\mathcal{Q}_\mathrm{h}}). \tag{2.8}$$

A numerical Legendre transform gives the desired large deviation rate function, which is shown in Fig. 2.3(a).

These long-time joint statistics of $\mathcal{W}$ and $\mathcal{Q}_\mathrm{h}$ determine the statistics of efficiency. We employ the contraction principle to obtain the efficiency rate function [32]. This technique amounts to making a saddle-point approximation along each line of constant efficiency on I,

$$J(\eta) = \max_{\mathcal{Q}_\mathrm{h}} I(-\eta\mathcal{Q}_\mathrm{h}, \mathcal{Q}_\mathrm{h}). \tag{2.9}$$

The result of the contraction is shown in Fig. 2.3(b). As in the prior work of Verley et al. [57], the minimum value obtained by each curve corresponds to a class of trajectories for which the average entropy production is zero, demonstrated in Fig. 2.3(c). Our results demonstrate that there can be two values of efficiency for which the average entropy production vanishes, one of which occurs at η_C. In fact, the minimal efficiency of our engine, hereafter denoted η^*, does not occur precisely at η_C.

2.3 The location of the minimum of $J(\eta)$

The key distinguishing feature of our engine compared with those discussed in Ref. [57] is that our protocol lacks time-reversal symmetry. To emphasize the importance of time-reversal symmetry, we also plot the efficiency distribution of a time-reversed protocol, shown in red in Fig. 2.3. The minimum in the rate function for the time-reversed protocol, denoted $\tilde{\eta}^*$, is distinct from both η^* and η_C. Indeed, we can relate η^* to $\tilde{\eta}^*$ via the fluctuation theorem [18, 22], thereby illustrating that $\eta^* = \eta_\mathrm{C}$ if and only if the protocol is time-reversal symmetric.

Entropy production is defined on a single trajectory level as

$$\Delta S = \log \frac{P[x(t)|\Lambda(t)]}{\tilde{P}[\tilde{x}(t)|\tilde{\Lambda}(t)]}, \tag{2.10}$$

where $x(t)$ is a trajectory subject to a time-dependent non-equilibrium protocol $\Lambda(t)$ and a tilde denotes the time reversal of a function [61]. Assuming that the dynamics of the system

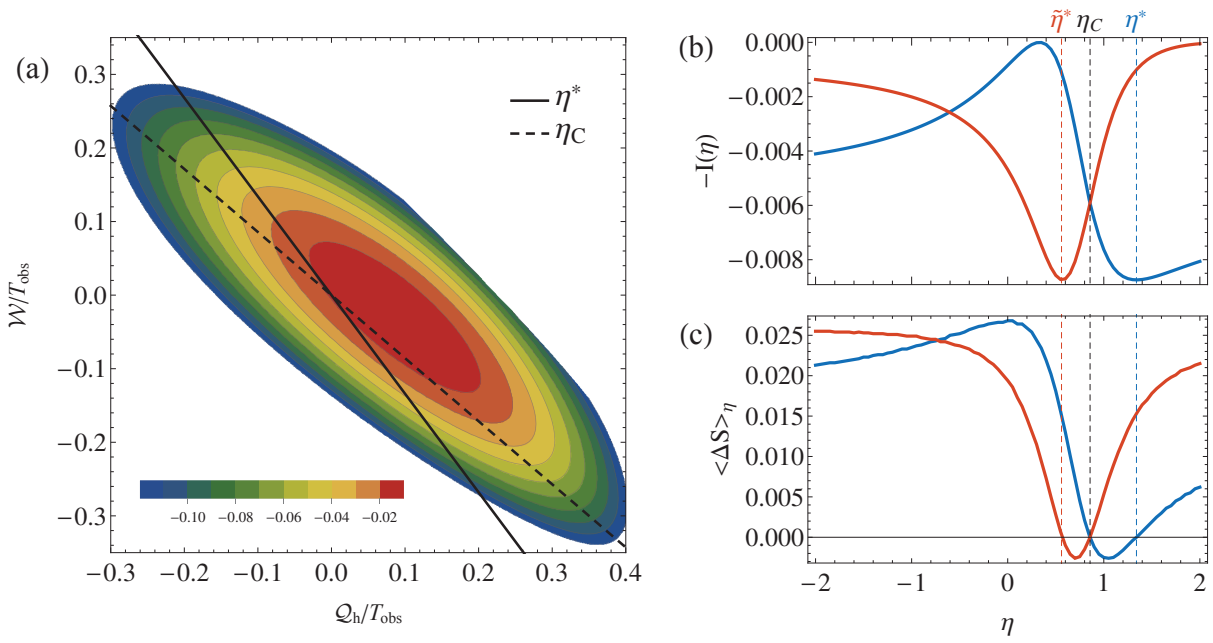

Figure 2.3: (a) The large deviation rate function, I, for the joint observations of $\mathcal{W}$ and $\mathcal{Q}_\mathrm{h}$. Parameters for the engine protocol are the same as in Fig. 2.2. The black lines have slopes $-\eta^*$ and $-\eta_\mathrm{C}$, corresponding to the set of $(W, \mathcal{Q}_\mathrm{h})$ observations consistent with the respective values of efficiency. (b) The efficiency rate function evaluated at η is given by the maximum of I attained along a line through the origin with slope -η. Red and blue lines show the rate function for the forward and reverse protocols, respectively. Note that the minimal efficiency, η^* corresponds to the slope of the tangent line through the origin in (a). (c) Plot of the most likely value of entropy production in the long-time limit conditioned upon a given value of efficiency. Only in time-reversal-symmetric protocols will this function be strictly non-negative.

are microscopically reversible, we can interpret the entropy production in terms of the heat absorbed from the thermal reservoirs. In particular,

$$\Delta S = -\beta_\mathrm{h}\mathcal{Q}_\mathrm{h} - \beta_\mathrm{c}\mathcal{Q}_\mathrm{c} + \log\frac{p_\mathrm{ss}(x(0)|\Lambda(0))}{\tilde{p}_\mathrm{ss}(\tilde{x}(0)|\tilde{\Lambda}(0))}, \tag{2.11}$$

with p_ss denoting the steady state probability [18]. The heat is extensive in time while the contribution due to the steady states probability distributions is not and can therefore be neglected in the long-time limit. By the first law, $\mathcal{Q}_\mathrm{c} = \Delta E - W - \mathcal{Q}_\mathrm{h}$, but we can similarly neglect ΔE as it too does not grow extensively with the length of the trajectory. Thus in the long-time limit $\Delta S = \beta_\mathrm{c}\mathcal{Q}_\mathrm{h}\,(\eta_\mathrm{C} - \eta)$. It follows from Eq. 2.10 that

$$I(\mathcal{W}, \mathcal{Q}_\mathrm{h}) = \tilde{I}(-\mathcal{W}, -\mathcal{Q}_\mathrm{h}) + \beta_\mathrm{c}\,(\mathcal{Q}_\mathrm{h}\eta_\mathrm{C} + \mathcal{W})\,, \tag{2.12}$$

where $\tilde{I}$ is the large deviation rate function for the time-reversed protocol. The minima of the efficiency rate functions are determined from the gradients of I and $\tilde{I}$ at $(\mathcal{W}, \mathcal{Q}_\mathrm{h}) = (0, 0)$.

$$\eta^* = \frac{\left(\frac{\partial I}{\partial \mathcal{Q}_\mathrm{h}}\right)\Big|_{(0,0)}}{\left(\frac{\partial I}{\partial \mathcal{W}}\right)\Big|_{(0,0)}} \qquad \tilde{\eta}^* = \frac{\left(\frac{\partial \tilde{I}}{\partial \mathcal{Q}_\mathrm{h}}\right)\Big|_{(0,0)}}{\left(\frac{\partial \tilde{I}}{\partial \mathcal{W}}\right)\Big|_{(0,0)}} \tag{2.13}$$

Differentiation of Eq. (2.12) with respect to $\mathcal{W}$ and $\mathcal{Q}_\mathrm{h}$ implies, after some simplifying algebra,

$$\eta^* - \eta_\mathrm{C} = T_\mathrm{c}\,(\eta^* - \tilde{\eta}^*)\left(\frac{\partial \tilde{I}}{\partial \mathcal{W}}\right)\Bigg|_{(0,0)}. \tag{2.14}$$

Note that $\frac{\partial \tilde{I}}{\partial \mathcal{W}}\Big|_{(0,0)}$ is non-zero for systems that have non-zero mean work extraction by the convexity of the rate function. We therefore see that $\eta^* = \eta_\mathrm{C}$ precisely when $\eta^* = \tilde{\eta}^*$. The minimum of $J(\eta)$ occurs at the Carnot efficiency if the protocol is time-reversal symmetric since the symmetry enforces $\eta^* = \tilde{\eta}^*$. In the more generic case of time-asymmetric engines, however, distinct values of η^* and $\tilde{\eta}^*$ imply that neither of the minima occur at η_C.

2.4 Finite-time efficiency distributions

As demonstrated by numerical simulation, there is a significant regime of observation times for which the efficiency distribution is not well-described by the large deviation form and no local minimum is evident. The minimum in the efficiency rate function therefore may not be apparent for actual experimental measurement of efficiency statistics. Nevertheless, we may leverage the large deviation form for work and heat, Eq. (2.2), to construct an approximation for $P(\eta)$ that is much more faithful to finite-time statistics.

Consider the coordinate transformation from $(\mathcal{W}, \mathcal{Q}_\mathrm{h})$ to $(\eta, \Delta S)$, where

$$\eta = -\frac{\mathcal{W}}{\mathcal{Q}_\mathrm{h}}, \qquad \Delta S = -\beta_\mathrm{C}(\mathcal{W} + \eta_\mathrm{C}\mathcal{Q}_\mathrm{h}). \tag{2.15}$$

The Jacobian for this transformation contributes negligibly to the distribution $P(\eta, \Delta S)$ in the long-time limit since it does not vary exponentially with t_obs. At long but finite times, however, it can strongly shape statistics of η and ΔS. Retaining this Jacobian, while exploiting the large deviation form of $P(\mathcal{W}, \mathcal{Q}_\mathrm{h})$, we estimate

$$P(\eta, \Delta S) \sim \left|\frac{-T_\mathrm{c}^2 \Delta S}{(\eta - \eta_\mathrm{C})^2}\right| e^{t_\mathrm{obs} I\left(\frac{\eta T_\mathrm{c} \Delta S}{(\eta_\mathrm{C} - \eta)}, \frac{T_\mathrm{c} \Delta S}{(\eta - \eta_\mathrm{C})}\right)}. \tag{2.16}$$

A saddle point approximation along the ΔS axis then gives a one dimensional representation of the probability distribution for efficiency. This procedure produces very accurate predictions for $P(\eta)$ at finite t_obs (small dotted lines in Fig. 2.2(c)), demonstrating that the form of $P(\eta)$ may be reliably extracted from the large deviation form for work and heat fluctuations.

Efficiency is meant to provide an assessment of how much work can be extracted from a machine relative to the expense of operating it. For macroscopic heat engines, fluctuations in work and heat are vanishingly small from one cycle to the next such that efficiency for the engine takes on a single value. Because fluctuations in a microscopic engine are large, the efficiency also fluctuates. An understanding of the finite-time and long-time statistics of the efficiency provides a lens through which to assess the design of a microscopic engine.

We have confirmed that the long-time statistics of time-reversible engines will exhibit a minimum at the Carnot efficiency [57], but this minimum does not necessarily lie at the Carnot efficiency for the ubiquitous case of time-asymmetric driving. While it is tempting to use these long-time statistics as a representation of the efficiency distributions, in practice efficiency distributions are collected from finite time measurements. Taking the finite time sampling into account, we have illustrated that the universal global minimum in the rate function is absent in efficiency distributions. Despite the fact that heat and work are well described by a large deviation form for the observation times we examined, these times are still too short for the efficiency rate function to be representative of the efficiency distribution. Our results highlight that it is nevertheless possible to understand and predict efficiency distributions on finite timescales.

Chapter 3

Fluctuations in currents far from equilibrium

In Section 1.1, we emphasized the difficulty of measuring dissipation directly. Currents, on the other hand, are an easily measurable alternative signature of nonequilibrium dynamics. However, unless detailed microscopic currents can be monitored, the available coarse-grained information provides only an indirect probe of the amount of dissipation of the underlying process. Remarkably, by studying the fluctuations of coarse-grained currents, we can estimate the microscopic dissipation rate [62].

Near-equilibrium fluctuations correspond quantitatively to the amount of dissipation [43, 45, 63, 64]. The diverse set of relations between the fluctuations in an equilibrium system and its dissipative response fall under the framework of the fluctuation-dissipation theorem [45]. The equality between equilibrium fluctuations and near-equilibrium response breaks down when a system is driven beyond the linear response regime.

Extending fluctuation-response relations beyond linear response has constituted a major effort in nonequilibrium statistical physics [65–71]. In this chapter, we discuss one such extension which asserts a bound on the dissipation rate based on current fluctuations. Demonstrating this bound relies on mathematical tools from the theory of large deviations for Markov processes, which we discuss thoroughly in Section 3.1. We then consider this bound under a coarse-graining procedure that maps a time-periodic nonequilibrium steady state to a time-independent steady state. Some of the work in this chapter appeared in Rotskoff [72] and Gingrich, Rotskoff, and Horowitz [62].

3.1 Large deviations at level-2.5

An effort to extend the fluctuation-dissipation theorem to systems arbitrarily far from equilibrium requires a comprehensive understanding of current fluctuations. This seemingly extreme requirement turns out to be surmountable for some systems: a complete characterization of current fluctuations, at the level of large deviations, can be derived for Markovian

dynamical systems [67, 73–75]. Our knowledge of the spectrum of current fluctuations can then be leveraged to study dissipation.

To compute the spectrum of current fluctuations, we start with a more general large deviation function. This approach is common in the theory of large deviations and relies on the "contraction principle" [32, 76]. Suppose that we know the large deviation function I for a dynamical observable A and let $a \equiv A/t_{\rm obs}$, then

$$P_{t_{\rm obs}}(a) \asymp \exp\left(-t_{\rm obs} I(a)\right). \tag{3.1}$$

If we want to compute the asymptotic fluctuations in some other observable B which we can express as a function of A, i.e., $B = f(A)$ then the contraction principle says that we can compute $I(b)$ directly from $I(a)$, where we have defined $b \equiv B/t_{\rm obs}$. A contraction stipulates that the most probable value of A such that $B = f(A)$ determines the asymptotic probability of B. At the level of the rate functions, this means that we can determine $I(b)$ by evaluating $I(a)$ at the smallest value of $I(a)$ satisfying the constraint that $f(A) = B$

$$I(b) = \inf_{A \text{ s.t. } B=f(A)} I(a). \tag{3.2}$$

Donsker and Varadhan [76] introduced the contraction principle in a slightly more general context, which gave rise to a "level" formalism. Their focus on functions of the density $\boldsymbol{\rho}$ led them to introduce a "process" level rate function, which gives the large deviations of $\boldsymbol{\rho}$ [76, 77]. Ellis referred to the process level as level II, because any function of $\boldsymbol{\rho}$, a level I rate function, could be computed from this more general function via contraction [78]. The categorization also includes large deviations at level III, which, in this context, is most simply described as a large deviation functional for distribution of dynamical trajectories.

The key observables of nonequilibrium statistical mechanics do not depend on the empirical density alone, as we described in Section 1.1. Understanding fluctuations in currents and dissipation, for example, requires information about the probability of fluctuations in the *dynamics*. The most direct way of probing dynamical fluctuations involves computing the rate function for joint fluctuation in the empirical density

$$\rho_x = \frac{1}{t_{\rm obs}} \int_0^{t_{\rm obs}} dt\ \delta_{z(t),x}, \tag{3.3}$$

and the empirical flow

$$q_{yx} = \frac{1}{t_{\rm obs}} \int_0^{t_{\rm obs}} dt\ \delta_{z(t^-),x}\delta_{z(t^+),y}. \tag{3.4}$$

These quantities are "empirical" in the sense that they are measured over some long trajectory.

In order to be explicit, let us restrict to the case of Markov jump processes on a finite state space graph, as introduced in Section 1.1 to compute a joint rate function $I(\boldsymbol{\rho}, \boldsymbol{q})$. In this case, it is important to note that the empirical flow contains more information than the empirical current; the latter specifies only the difference between the flow in the forward and

reverse directions $j_{xy} = q_{xy} - q_{yx}$. For example, the empirical current would not distinguish between a trajectory in which there are 100 $x \to y$ hops and 80 $y \to x$ hops from one in which there are 20 $x \to y$ hops and 0 in the opposite direction, while the flows would be dramatically different.

To compute asymptotic probability of a deviation $\boldsymbol{\rho}$ and $\boldsymbol{q}$ away from the expected steady state values $\boldsymbol{\rho}^{\text{ss}}$ and $\boldsymbol{q}^{\text{ss}}$, we follow a standard approach, the Cramér tilting procedure described in Section 1.3. The calculation proceeds by constructing an effective dynamics in which the rare values $\boldsymbol{\rho}$ and $\boldsymbol{q}$ are typical. We then compare the probability of the rare trajectory to a typical trajectory. A trajectory through the finite state space of the graph can be described as a sequence of states

$$X_0, \ldots, X_N \tag{3.5}$$

and transition times

$$t_1, \ldots, t_N. \tag{3.6}$$

The probability of an individual transition in the trajectory is a function of the rates W_{ij}

$$p(X_i \to X_{i+1}) = W_{X_{i+1},X_i} e^{-W_{X_i,X_i}(t_{i+1}-t_i)}, \tag{3.7}$$

and the diagonal element of the rate matrix W gives the escape rate. Because each transition is independent under the assumption of Markovian dynamics, the probability of a trajectory is simply the product of the terms above

$$P[X(t)] = \rho_{X_0} \prod_{i=0}^{N-1} W_{X_{i+1},X_i} e^{-W_{X_i,X_i}(t_{i+1}-t_i)}. \tag{3.8}$$

The path measure defined by Eq. (3.8) is expressed as a sum over transitions in a trajectory. Alternatively, for a long trajectory with empirical flow $\boldsymbol{q}$ we could represent the measure as product over all states. To do so, we need the following statement of conservation of probability

$$\sum_j q_{ij} - q_{ji} = 0. \tag{3.9}$$

Without this constraint, a single state could act as a source or a sink for probability because the net flow of probability into site i would differ from the outward flow of probability.

Bearing this fact in mind, we see write the probability of a trajectory with empirical flow $\boldsymbol{q}$ as

$$P[X(t)] = \exp\left(-\sum_{i,j} W_{ij}\rho_j\right) \prod_{i,j} W_{ij}^{q_{ij}}. \tag{3.10}$$

The exponential portion of the expression above accounts for the expected escape from each site and the product in Eq. (3.10) accumulates the probability associated with each transition.

We now compare the probability of a trajectory with empirical flow $\boldsymbol{q}$ and empirical density $\boldsymbol{\rho}$ in the original measure and a tilted measure in which $\boldsymbol{\rho}$ and $\boldsymbol{q}$ are typical. Constructing the biased measure is surprisingly simple. In fact, we need only replace the rates W_{ij} with the rates expected from the empirical flows q_{ij}, modifying the escape rates accordingly. Due to the constraint, Eq. (3.9), once we have specified $\boldsymbol{q}$, the empirical density for the trajectory is uniquely specified.

Thus, the likelihood ratio of the two measures can be written as

$$\frac{P[X(t)]}{P^{\text{tilt}}[X(t)]} = \frac{\exp\left(-\sum_{i,j} W_{ij}\rho_j\right)}{\exp\left(-\sum_{i,j} q_{ij}\right)} \prod_{i,j} \left(\frac{W_{ij}\rho_j}{q_{ij}}\right)^{q_{ij}}. \tag{3.11}$$

It should be emphasized that the expression given in Eq. (3.11) is not a mathematically rigorous one. To properly compare the path probabilities, one needs to employ more formal tools from measure theory. In particular, the precise meaning of Eq. (3.11) must be understood in terms of a Radon-Nikodym derivative. Further details of this mathematical subtlety are discussed in Appendix A.

In the long time limit, the probability of the empirical density $\boldsymbol{\rho}$ and the empirical flow $\boldsymbol{q}$ is dominated by the change of measure. That is, for trajectories $X(t)$ satisfying the constraints in Eq. (3.9),

$$P(\boldsymbol{\rho}, \boldsymbol{q}) = \int \mathcal{D}[X(t)] P^{\text{tilt}}[X(t)] \frac{P[X(t)]}{P^{\text{tilt}}[X(t)]} \delta(\boldsymbol{\rho}[X(t)] - \boldsymbol{\rho}) \delta(\boldsymbol{q}[X(t)] - \boldsymbol{q}), \tag{3.12}$$

$$= \frac{\exp\left(-\sum_{i,j} W_{ij}\rho_j\right)}{\exp\left(-\sum_{i,j} q_{ij}\right)} \prod_{i,j} \left(\frac{W_{ij}\rho_j}{q_{ij}}\right)^{q_{ij}}. \tag{3.13}$$

Because $\boldsymbol{\rho}$ and $\boldsymbol{q}$ are typical in the tilted distribution, the exponential scale fluctuations are distributed as

$$P(\boldsymbol{\rho}, \boldsymbol{q}) = e^{-t_{\text{obs}} I(\boldsymbol{\rho}, \boldsymbol{q})}, \tag{3.14}$$

where the large deviation function is just,

$$I(\boldsymbol{\rho}, \boldsymbol{q}) = \log \frac{P}{P^{\text{tilted}}}, \tag{3.15}$$

$$= \sum_{i,j} q_{ij} \log \frac{q_{ij}}{W_{ij}\rho_j} - q_{ij} + W_{ij}\rho_j. \tag{3.16}$$

The functional $I(\boldsymbol{\rho}, \boldsymbol{q})$ is known as the level-2.5 large deviation rate function [73–75].

Because the empirical flow contains all the information necessary to compute empirical currents, we can extract the spectrum of current fluctuations from level-2.5 function. Using the contraction principle Eq. (3.2), we see that

$$I(\boldsymbol{\rho}, \boldsymbol{j}) = \inf_{q_{ij} \text{ s.t. } j_{ij} = q_{ij} - q_{ji}} I(\boldsymbol{\rho}, \boldsymbol{q}). \tag{3.17}$$

Note that, by convention, currents that are not consistent with the constraints in Eq. (3.9), have $I(\boldsymbol{\rho}, \boldsymbol{j}) = \infty$, which is also the case for $I(\boldsymbol{\rho}, \boldsymbol{q})$ when the flows are not conservative.

In the absence of dynamical phase transitions, the infimum defining $I(\boldsymbol{\rho}, \boldsymbol{j})$ can be computed explicitly. In fact, the minimization involves only routine calculus; we express the rate function

$$I(\boldsymbol{\rho}, \boldsymbol{j}) = \sum_{i,j} W_{ij}\rho_j + W_{ji}\rho_i - q^*_{ij} - q^*_{ji} + q^*_{ij} \ln \frac{q^*_{ij}}{W_{ij}\rho_j} + q^*_{ji} \ln \frac{q^*_{ji}}{W_{ji}\rho_i}, \tag{3.18}$$

in terms of the minimizer

$$q^*_{ij} = \operatorname{argmin}_{q_{ij}} I(\rho_i, q_{ij}) - I(\rho_j, q_{ji} - j_{ij}) \tag{3.19}$$

$$= \frac{1}{2}\left(j_{ij} + \sqrt{j_{ij}^2 + 4W_{ij}\rho_j W_{ji}\rho_i}\right) \tag{3.20}$$

The functional form of $I(\boldsymbol{\rho}, \boldsymbol{j})$ already exposes a connection to the dissipation. This can be further emphasized by writing the expression in terms of thermodynamic affinities [74].

A similar argument can be carried out for diffusion processes, like those described in Section 1.1. In the case of diffusions, however, there is no distinction between the microscopic empirical flow and the microscopic empirical current. The resulting functional, which can also be viewed as a coarse-graining of Eq. (3.16), is a Gaussian functional

$$I[\boldsymbol{\rho}, \boldsymbol{j}] = \int d\boldsymbol{x}\ \frac{\left(\boldsymbol{j}(\boldsymbol{x}) - \boldsymbol{j}^{ss}(\boldsymbol{x})\right)^2}{4D\rho(\boldsymbol{x})}. \tag{3.21}$$

In applications to nonequilibrium systems, Eq. (3.21) is sometimes called the macroscopic fluctuation functional. It has been successfully used to study a variety of diffusive nonequilibrium steady states [79].

3.2 Extending fluctuation-dissipation relations beyond linear response

The large deviation results for current fluctuations described in the previous section hold for nonequilibrium steady states arbitrarily far from equilibrium. As seen in Eq. (3.18), the extent of nonequilibrium current fluctuations is mathematically tied to the steady-state rate of dissipation. Maes and Netočny [67] argued that the level-2.5 formalism provided a far-from-equilibrium generalization of the Onsager-Machlup approach. Indeed, the level-2.5 rate function (3.18) asserts a relationship between the average dissipation and current fluctuations, but, at first blush, it appears to be a complicated one.

Barato and Seifert [80] discovered a simple relationship between current fluctuations and dissipation in the context of a statistical estimation problem. They investigated with what certainty the average of a current-like observable could be measured in the context of

enzyme kinetics. Under the assumption of linear response regime, they proved, that for any conservative current $\boldsymbol{j}$,

$$\frac{2(\boldsymbol{j}^{\text{ss}})^2}{\text{var}(\boldsymbol{j})} \leq \sigma^{\text{ss}}, \tag{3.22}$$

where σ^{ss} is the steady-state entropy production of the process and $\boldsymbol{j}^{\text{ss}}$ is the steady-state current. Their work initiated a body of research [62, 72, 81–86] concerning what has come to be more generally understood as a generic relationship between current fluctuations and dissipation in nonequilibrium steady states.

The bound (3.22) holds in remarkable generality. Gingrich et al. [85] proved that Eq. (3.22) holds for nonequilibrium steady states arbitrarily far from equilibrium by appealing to the level-2.5 rate function formalism. From Eq. (3.18) one can prove the uncertainty relation by first evaluating $I(\boldsymbol{\rho}, \boldsymbol{j})$ at the steady state density, $\boldsymbol{\rho}^{\text{ss}}$. By the contraction principle, we know that $I(\boldsymbol{\rho}^{\text{ss}}, \boldsymbol{j}) > I(\boldsymbol{j})$. Then, the quadratic upper bound Eq. 3.22 follows by demonstrating that a quadratic bound holds on $I(\boldsymbol{\rho}^{\text{ss}}, \boldsymbol{j})$. Ample numerical evidence for a wide variety of nonequilibrium steady states was provided in Ref. [82], alongside proofs in some limiting cases.

The uncertainty relation does not necessarily provide a tight bound. Gingrich, Rotskoff, and Horowitz [62] addresses the problem of inferring the dissipation rate from coarse-grained measurements of the current and explores when accurate inference can be performed. Of course, when the thermodynamic affinities are small, the bound is tightest—consistent with Gaussian form of Eq. (3.21). Efforts have been made to tighten the quadratic bound [84] and to extend a weaker version to finite-time measurements, outside the large deviation regime [86].

3.3 Time-dependent driving and stochastic pumping

The discussion of thermodynamic uncertainty and level-2.5 large deviations applies exclusively to nonequilibrium steady states. Of course, many of the physical systems that we would like to study do not fit into this category. Experiments in single molecule biophysics often perturb proteins or DNA using time-dependent protocols to study their response [87–89]. At the cellular scale, biological environments on the cellular scale are often subject to oscillatory environmental conditions, a type of perturbation mimicked in laboratory experiments [50].

In Markovian models of physical dynamics, each transition from one state to another is modeled as a Poisson process, as described in detail in Sec. 1.1. Many physical systems can be realistically represented by this dynamics due to the separation of timescales that pervades chemical and biophysical processes. When a Markov process has a steady state that breaks detailed balance, the physical system that it represents must be dissipative. The energy flow that drives such a nonequilibrium dynamics is modeled *implicitly*—the stochastic transition rates represent some random energy transfer process, for example, the statistics of chemical reactions that catalyze a change of state and an asymmetry in time.

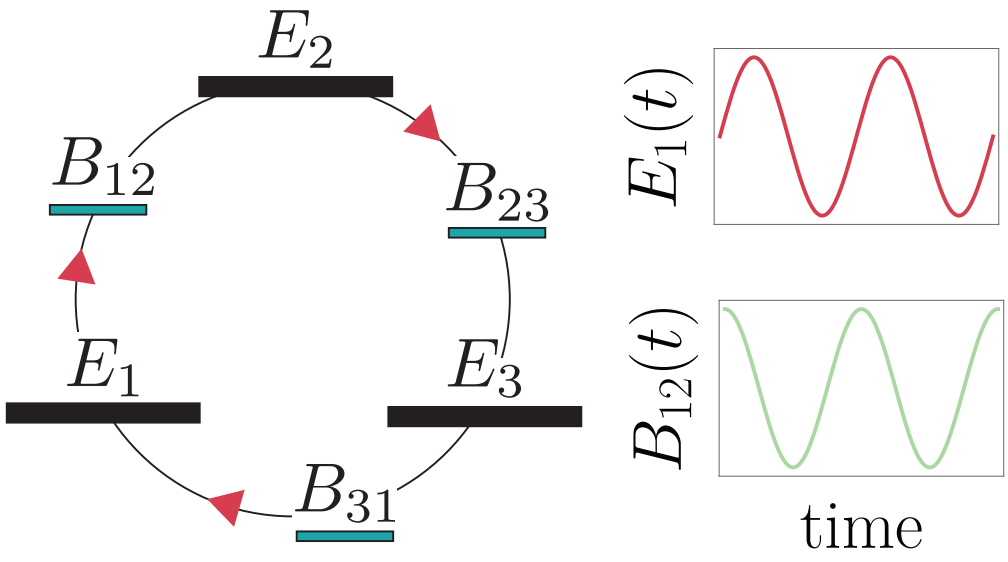

Figure 3.1: The simplest realization of the stochastic pumping phenomenon. The dynamics of a particle on the three state cycle is governed by a Poisson process with Arrhenius rates. The energy levels and barrier heights can be varied periodically in time to generate a non-zero average current through the system. The "no-pumping" theorem states that both a barrier and an energy must be modulated periodically time to produce current.

Nonequilibrium behavior, of course, can be achieved by pumping energy into a physical system with deterministic, external control. When we manipulate a physical system as a function of time, the rates of transitions among the states become time-dependent. Such systems are not purely Markovian and they do not relax to a time-independent steady state.

Consider, as an example of an externally driven system, a collection of energy levels with barriers between them, as depicted schematically in Fig. 3.1. Currents can be generated even in the case that the rates always satisfy detailed balance, a phenomenon known as stochastic pumping. To see this, we can modulate the barrier heights and energy levels periodically in time so that the Arrhenius rates are given by

$$W_{ij}(t) = \exp\left(B_{ij}(t) - E_j(t)\right). \tag{3.23}$$

Note that for any fixed time t_0, the rates satisfy detailed balance. As a result, if the protocol were stopped at any point, the system would relax to an equilibrium state with no net currents. In the case that the protocol is periodic, the system relaxes to a unique, time-periodic nonequilibrium steady state with the same period as the protocol [90].

Stochastic pumps have been studied extensively and a number of deep results characterize their dynamics. Among the most surprising is the "no-pumping" theorem [91–94], which provides the minimal conditions on the external protocol for pumping a probability current through the system. In the context of the model described above, the theorem asserts that

at least one energy level E_i and one barrier B_{ij} must be varied in time in order to achieve non-vanishing currents. Mandal and Jarzynski [91] give a beautiful, straightforward proof of this theorem.

3.4 Mapping time-dependent systems to time-independent nonequilibrium steady states

Nonequilibrium steady states are an essential paradigm for describing nanoscale biological machines, such as molecular motors that extract work from chemical gradients [95]. When a system is coupled to reservoirs with different chemical potentials, the dynamics breaks detailed balance and persistent, directed motion can be used to perform mechanical work. Such a system is typically described as a Markov process with time-independent rates that depend both on the external chemical gradient and internal dynamics.

Promising applications across many disciplines have motivated efforts to design artificial molecular machines that behave like those in biological settings. Nonequilibrium steady states, however, have proved difficult to engineer [96]. Time-dependent external perturbations offer an alternative route to breaking detailed balance. Indeed, many synthetic nanoscale machines are implemented as "stochastic pumps," in which currents are generated by periodically varying an external potential [94, 97–100]. A stochastic pump can be modeled as a non-homogeneous Markov jump process with instantaneous Arrhenius rates that are determined by time-dependent energy levels and barrier heights, as depicted in Fig. 3.1 [91, 92, 101].

Raz et al. [101] proposed a mapping between time-independent steady states and periodically driven stochastic pumps that offers a set of design principles for engineering biomimetic nanodevices. While the mapping ensures that the average properties are asymptotically equivalent in both representations, it makes no guarantees about the fluctuations. At the nanoscale, however, fluctuations play a crucial in determining characteristics like work and efficiency in finite-time measurements [55, 57]. Indeed, fluctuations in both efficiency and current have become a central focus in nonequilibrium statistical mechanics: theoretical developments have predicted universal properties of fluctuating nanoscale machines [57, 80, 85, 102] and experimental realizations of nanoscale engines have drawn particular attention to the impact of fluctuations on measurements of efficiency [100, 103].

Translating between nonequilibrium steady states and stochastic pumps relies on the so-called "dynamical equivalence principle" of Zia and Schmittmann [104]. This principle stipulates that nonequilibrium steady states are characterized by the average currents and the average density. For Markov jump processes, the asymptotic fluctuations of a nonequilibrium steady state, however, are not dictated by these average properties alone.

The large deviation formalism of Section 3.1 suggests a stricter requirement for dynamical equivalence among jump processes: if the asymptotic form of the fluctuations is to be accurately captured, then it is not the average currents, but rather the average flows that

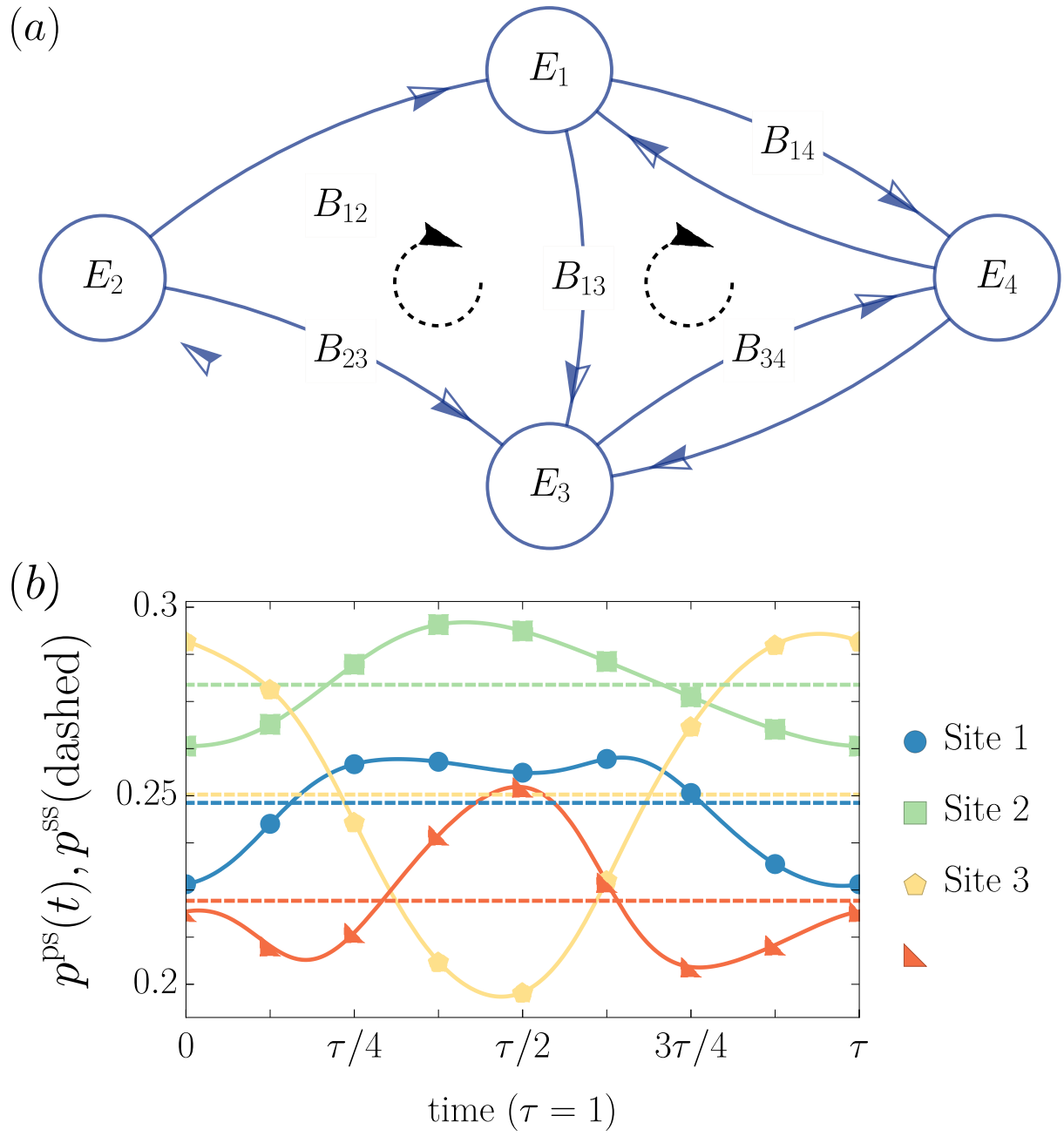

Figure 3.2: (*a*) A schematic of the stochastic pump under consideration. Symmetric barriers B_{ij} and energy levels E_i parametrize Arrhenius rates and are varied periodically in time to generate a current. The corresponding nonequilibrium steady state representation of the pump has no time-dependence, but rather rates that break detailed balance. (*b*) The time periodic steady state probabilities for each site on the graph are shown over an entire period τ. The solid lines show the time-dependent occupation of the pump. The dashed lines show the average occupancy per period, a property matched by the corresponding steady state.

must be used to describe the dynamics of a nonequilibrium steady state. This is a more rigid prescription. Further, these insights motivate a solution to the mapping problem between stochastic pumps and nonequilibrium steady states that preserves the fluctuations. Interestingly, in order to optimally describe current fluctuations of a stochastic pump, the corresponding nonequilibrium steady state must have a lower average entropy production rate than that of the pump. The origin of this "excess" entropy production can be explained with a simple decomposition of the entropy production of the stochastic pump [25, 53, 105].

The nonequilibrium steady state representation of the pump satisfies a universal lower bound on the magnitude of its current fluctuations, dictated by the total entropy production less the excess [81, 83, 85]. As a consequence of this splitting, we demonstrate that, in a perturbative limit, stochastic pumps satisfy a universal bound on their current fluctuations, dictated by the entropy production of the corresponding steady state. Taken together, these insights offer a powerful set of design principles for translating between stochastic pumps and steady states as well as a potentially useful technique for theoretical analysis of systems under time-dependent driving.

To illustrate our mapping, we consider a simple model of a stochastic pump: a single particle hopping with Arrhenius rates on a four state graph. We vary two energy levels and one barrier periodically in time, which provides a time-dependent perturbation that generates a non-vanishing current, as permitted by the no-pumping theorem [94, 98, 99]. This setup is depicted in Fig. 3.2 (a).

The pump achieves a periodic steady state, which can be calculated numerically by integrating

$$p_i^{\mathrm{ps}}(t+s) = \int_t^s W_{ij}(t+t')p_j^{\mathrm{ps}}(t+t')dt'. \tag{3.24}$$

Here $p_i^{\mathrm{ps}}(t)$ is the probability of being in state i at time t and $W(t)$ is the continuous time rate matrix for the dynamics at time t. The periodic steady state satisfies

$$p_i^{\mathrm{ps}}(t+\tau) = p_i^{\mathrm{ps}}(t), \tag{3.25}$$

where τ is the period of the pumping protocol. Note that, by construction, $W(t)$ satisfies detailed balance at each point in time. The Arrhenius rates determine the instantaneous rate matrix

$$\begin{aligned} W_{ij}(t) &= e^{-\beta\left(B_{ij}(t)-E_j(t)\right)} && \text{for } i \neq j \\ W_{ii}(t) &= -\sum_{i\neq j} W_{ij}(t) \end{aligned}$$

where $E_j(t)$ denotes the energy level of state j and $B_{ij}(t) = B_{ji}(t)$ is the barrier height. In our example, the only time-dependent quantities are

$$\begin{aligned} E_3(t) &= \sin(2\pi t/\tau) + 1, \\ E_4(t) &= \sin(4\pi t/\tau), \\ B_{13}(t) &= \sin(2\pi t/\tau). \end{aligned} \tag{3.26}$$

The periodic solution is plotted in Fig. 3.2 (b).

We aim to find a time-independent rate matrix W^{ss} that mimics the stochastic pump and matches its fluctuations. Following [104], we let $\overline{W}_{ij} = W^{\text{ss}}_{ij}\hat{p}_j$ where $\hat{p}_j$ is the average occupancy in the periodic steady state and write

$$\overline{W}_{ij} = \mathcal{S}_{ij} + \mathcal{A}_{ij}, \tag{3.27}$$

where $\mathcal{S}$ is a symmetric, stochastic matrix and $\mathcal{A}$ is an antisymmetric matrix. The symmetric part of this decomposition is related to the "activity" of a trajectory [68, 106]. The continuous time rate matrix for the dynamics is then given by

$$W^{\text{ss}} = (\mathcal{S} + \mathcal{A})\,\mathcal{P}^{-1}, \tag{3.28}$$

where $\mathcal{P}$ is a diagonal matrix with $\mathcal{P}_{ii} = \hat{p}_i$, the steady state probability of site i. If we further impose the constraint that the steady state currents agree with the periodic average current along each edge,

$$\hat{j}_{ij} = \int_0^\tau dt\; W_{ij}(t)p^{\text{ps}}_j(t) - W_{ji}(t)p^{\text{ps}}_i(t), \tag{3.29}$$

then the antisymmetric part of the rate matrix is uniquely identified,

$$\mathcal{A}_{ij} = \frac{1}{2}\hat{j}_{ij}. \tag{3.30}$$

The rate matrix W^{ss} describes a probability conserving stochastic process, and, as a result, the form of $\mathcal{S}$ is constrained, but only weakly. In particular, for $i \neq j$, it must be the case that

$$\mathcal{S}_{ij} \geq |\mathcal{A}_{ij}| \tag{3.31}$$

and

$$\sum_j \mathcal{S}_{ij} = 0, \tag{3.32}$$

which ensures that W^{ss} is a stochastic matrix.

Though the rate matrix is not uniquely specified, any valid choice of $\mathcal{S}$ results in a stochastic process with identical average currents and average occupancy statistics. The same cannot be said for the fluctuations. The freedom in $\mathcal{S}$ can be directly represented by noting that any valid off-diagonal entry in the matrix can be written,

$$\mathcal{S}_{ij} = c_{ij}|\mathcal{A}_{ij}|,\;\; c_{ij} > 1. \tag{3.33}$$

Due to symmetry, there are $N(N-1)/2$ choices to make. Indeed, the rate matrices resulting from different choices of $\mathcal{S}$ yield different average entropy production rates, given by,

$$\hat{\sigma}_{ij} = \hat{j}_{ij} \ln \frac{c_{ij}|\hat{j}_{ij}| + \hat{j}_{ij}}{c_{ij}|\hat{j}_{ij}| - \hat{j}_{ij}}. \tag{3.34}$$

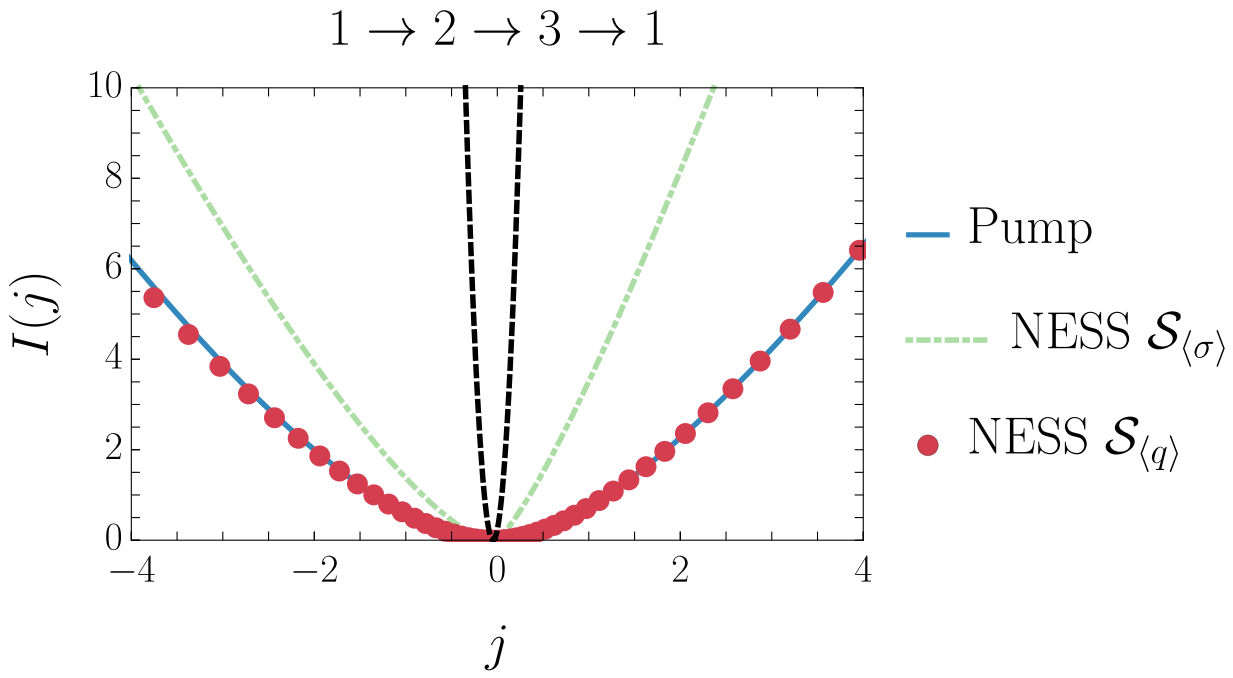

Figure 3.3: The large deviation rate function for the current around the upper cycle (see Fig. 3.2) $1 \to 2 \to 3 \to 1$ is shown for the stochastic pump (blue) the nonequilibrium steady state with the same average entropy production along each edge as the pump (green), and the nonequilibrium steady state with the same average flow along each edge as the pump (red dots). While the nonequilibrium steady state with $\mathcal{S}_{\langle\sigma\rangle}$ has the same average current, the character and extent of its fluctuations are extremely different. Choosing $\mathcal{S}_{\langle q\rangle}$ preserves even very rare fluctuations in current. The black dashed line shows the quadratic determined by Eq. (3.22) with σ^{ss} given by the steady state entropy production of the stochastic pump.

The c_{ij} values can be varied independently so long as they meet the constraint $c_{ij} \geq 1$, meaning that the total entropy production can be made arbitrarily small by taking c_{ij} large.

Raz et al. [101] suggest choosing $\mathcal{S}$ so that the average entropy production rate along each edge is the same in the stochastic pump and the nonequilibrium steady state representations. This choice, which we denote $\mathcal{S}_{\langle\sigma\rangle}$ uniquely specifies a rate matrix and also guarantees that the average current, occupancy, and entropy production rates are preserved by the map. However, the asymptotic fluctuations in entropy production and current are dramatically different.

To demonstrate this, we computed the entropy production and current large deviation rate functions for both the stochastic pump and the nonequilibrium steady state representation, shown in Figs. 3.3 and 3.6. To calculate the rate functions, we first compute the scaled

cumulant generating functions for entropy production ω and current j,

$$\psi_\omega(\lambda) = \lim_{t\to\infty} \frac{1}{t} \ln\langle e^{-\lambda\omega}\rangle, \qquad \psi_j(s) = \lim_{t\to\infty} \frac{1}{t} \ln\langle e^{-sj}\rangle. \tag{3.35}$$

For the nonequilibrium steady state representation, the cumulant generating functions can be calculated exactly by Cramér tilting [32]. In the case of the stochastic pump, the averages in (3.35) can be directly evaluated in the time-periodic steady state, meaning that the cumulant generating function can be numerically computed as,

$$\psi_\omega(\lambda) = \frac{1}{\tau} \ln \sum_j \int_0^\tau W_{ij}(t;\lambda) p_j^{\mathrm{ps}}(t), \tag{3.36}$$

where $W(t;\lambda)$ is the tilted rate matrix for entropy production [21]. We use the Gärtner-Ellis theorem to compute the large deviation rate functions by first computing the scaled cumulant generating function and then performing a Legendre-Fenchel transform [32].

Fig. 3.6 shows the entropy production rate function with $W^{\mathrm{ss}} = \left(\mathcal{S}_{\langle\sigma\rangle} + \mathcal{A}\right)\mathcal{P}^{-1}$. Note that, while the averages agree, the nature of the entropy production fluctuations is quite different. The steady state with the matching average entropy production has a notably thinner tail for large entropy production rates.

In order to match the fluctuations in current, we instead choose $\mathcal{S}$ so that the average empirical flows are accurately captured by the dynamics. In particular, we let

$$\mathcal{S}_{\langle q\rangle} = \hat{q}_{ij} - \frac{1}{2}\hat{j}_{ij} \iff \overline{W}_{ij} = \hat{q}_{ij}, \tag{3.37}$$

where $\hat{q}_{ij}$ denotes the average flow along edge ij in the periodic steady state. This choice has the additional advantage of simplicity: the dynamics produces the correct average number of hops in both directions along each edge of the network. We note that for high-dimensional networks, measuring all of the detailed edge currents or flows could be a formidable challenge. Nevertheless, biological motors have been successfully modeled as Markov jump processes with a small number of distinct ligation states [1] and engineered nanodevices typically have only a few states [107]. Because $\mathcal{S}$ does not affect the antisymmetric part of the rate matrix, the average currents along each edge are equivalent in both the stochastic pump and the nonequilibrium steady state. As illustrated by Fig. 3.3 choosing $\mathcal{S}_{\langle q\rangle}$ leads to striking agreement between the current fluctuations of the stochastic pump and the corresponding steady state.

3.5 Monte Carlo sampling

To probe the rate of convergence of the large deviation form for current fluctuations in both the pump and the nonequilibrium steady state, we simulated the dynamics using kinetic Monte Carlo (KMC) sampling. In the case of the nonequilibrium steady state, standard

algorithms can be employed [108, 109]. However, the procedure must be modified slightly to sample the pump, where the rates are time dependent. To perform the simulations, we follow Ref. [110]. We note that the probability of escape from state i in time Δt is,

$$\exp\left(-\int_t^{t+\Delta t} W_{ii}(t)dt\right). \tag{3.38}$$

Thus, we choose a random number $r \in (0,1]$ and compute Δt by numerically solving the following equation,

$$\ln(r) = -\int_t^{t+\Delta t} W_{ii}(t)dt. \tag{3.39}$$

Once Δt is determined, a new state is selected in proportion to the flow from the current state into the new state. Consider an ordered list of the rates at time $t+\Delta t$, i.e., $\{W_{ji}\}_{j\neq i}$ with $W_{ji} < W_{(j+1)i}$ for all j. We define

$$R_{ji}(t+\Delta t) = \sum_{k=1}^{j} W_{ki}(t+\Delta t). \tag{3.40}$$

Next, we choose a random number $r' \in (0,1]$ and perform a binary search to determine j such that,

$$R_{ji}(t+\Delta t) \leq r' W_{ii}(t+\Delta t) \leq R_{(j+1)i}(t+\Delta t). \tag{3.41}$$

We collected 1×10^6 independent trajectories for both the stochastic pump and the nonequilibrium steady state representations with an observation time $t_{\rm obs} = 1000$. We computed the scalar current j around the cycle $1\to 2\to 3\to 1$ and plotted $-1/t_{\rm obs}\ln p_{\rm sim}(j)$. The results, as shown in Fig. 3.4, are in good agreement with the asymptotic form of the rate function. The modest timescale over which the large deviation form is adopted emphasizes the practical implications of these predictions.

3.6 Effective stationary process

The time-periodic master equation can be written,

$$\begin{aligned} \partial_t p_i(t) &= \sum_j W_{ij}(t)p_j(t), \\ p_i(0) &= p_i^{\rm init}. \end{aligned} \tag{3.42}$$

The rates are assumed to vary periodically in time with period τ so that $W_{ij}(t+\tau) = W_{ij}(t)$. This equation admits a formal solution using the time-ordered exponential operator $\overrightarrow{\exp}$,

$$p(t+\Delta t) = \overrightarrow{\exp}\left(\int_t^{t+\Delta t} dt'\ W(t')\right) p(t), \tag{3.43}$$

$$\equiv \mathcal{G}(t, t+\Delta t)p(t). \tag{3.44}$$

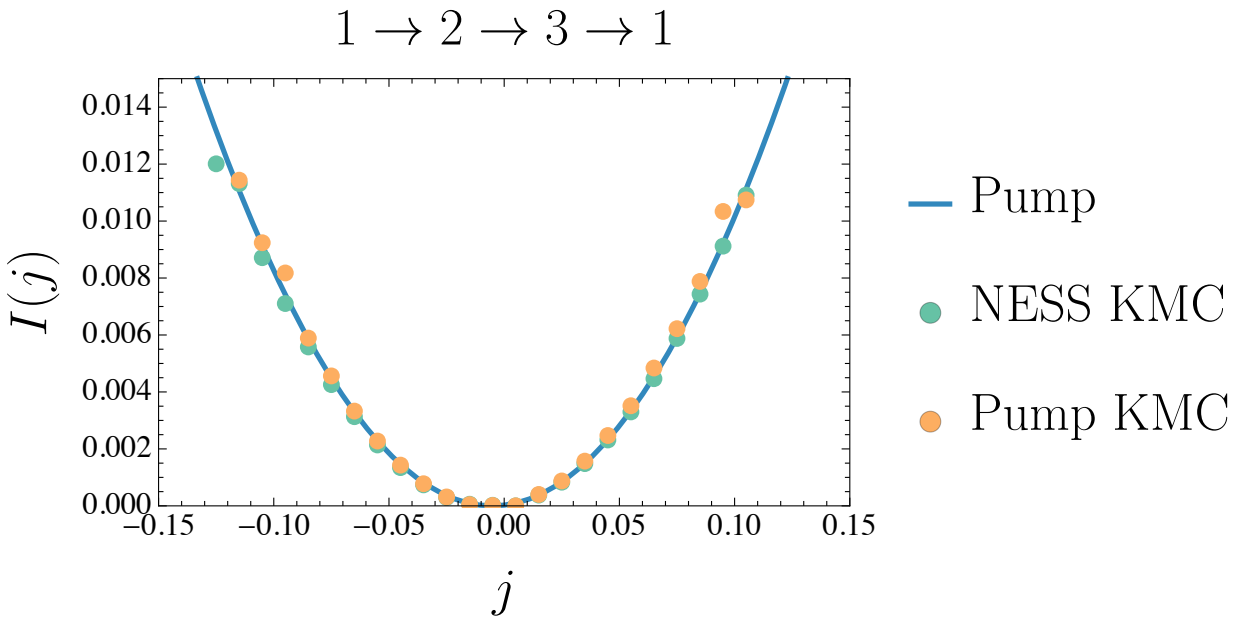

Figure 3.4: Results from kinetic Monte Carlo sampling show good agreement with the asymptotic limit for the stochastic pump. The nonequilibrium steady state simulations use the rate matrix determined by $\mathcal{S}_{\langle q \rangle}$. The KMC data for the stochastic pump was sampled with the Monte Carlo procedure described in the Section 3.5.

At long times, the solution becomes periodic, up to an exponential factor called the Floquet multiplier. In our case, there are no sources or sinks for the probability, so the exponents vanish (cf. Ref. [90]). The propagator $\mathcal{G}(t, t')$ is itself a periodic function of time by Floquet's theorem. Further, by the semi-group property,

$$\mathcal{G}(0, n\tau) = \mathcal{G}^n(0, \tau) \equiv \mathcal{G}^n(\tau). \tag{3.45}$$

In the long time limit, $\mathcal{G}(\tau)$ is discrete time rate matrix that propagates probability through the network. By the law of large numbers, the average flow q_{ij}^{eff} along each edge determined by $\mathcal{G}$ must match the average in the periodic steady state $\hat{q}_{ij}$. This correspondence is exact.

It remains to compare the current fluctuations of the non-homogeneous Markov process with current fluctuations in the effective process. A correspondence between the nonequilibrium steady state and pump will hold in the long time limit if the deviations from Poisson statistics determined by the periodic averages can be neglected. We consider the Fourier representation of the periodic dynamics in its periodic steady state; the right hand side of Eq. (3.42) becomes

$$\sum_j \sum_{k,l} \tilde{W}_{ij}(k-l)\tilde{p}_j^{\text{ps}}(l)e^{2\pi i k t}. \tag{3.46}$$

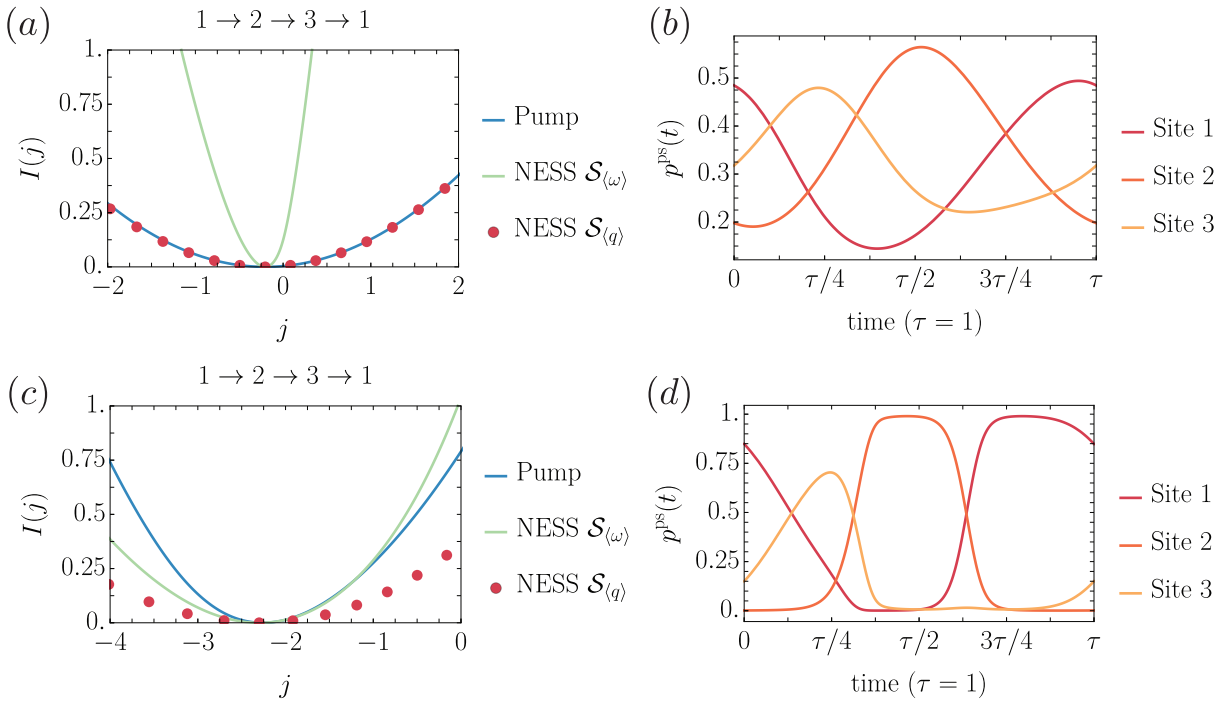

Figure 3.5: (a) The periodic steady state plotted over the duration of a single period for energy variations of amplitude $1k_{\rm B}T$. (b) The periodic steady state plotted over the duration of a single period for energy variations of amplitude $5k_{\rm B}T$. (c) The current large deviation rate function of the stochastic pump with energy variations of amplitude $1k_{\rm B}T$ compared with the current rate functions for the two coarse-grained representations. (d) The current large deviation rate function of the stochastic pump with energy variations of amplitude $5k_{\rm B}T$ compared with the current rate functions for the two coarse-grained representations.

If the time-periodic perturbation to the hopping rates is small, then we can neglect the higher Fourier coefficients and retain only the $k = 0$ contribution,

$$\sum_j \tilde{W}_{ij}(0) \sum_l \tilde{p}_j^{\rm ps}(l) e^{2\pi i l t}, \tag{3.47}$$

$$= \sum_j \hat{W}_{ij} p_j^{\rm ps}(t). \tag{3.48}$$

Because $\tilde{W}_{ij}(0)$ is the time periodic average, the transition rates are given by the average rate of hopping over the course of the period. The fluctuations, in this case, will be dominated by the long-time properties of $\mathcal{G}$. Empirically, the pump rate functions show robust agreement with the nonequilibrium steady state representation for a wide range of different pumping protocols and networks. However, this perturbative limit can be broken with strong variation in the energy.

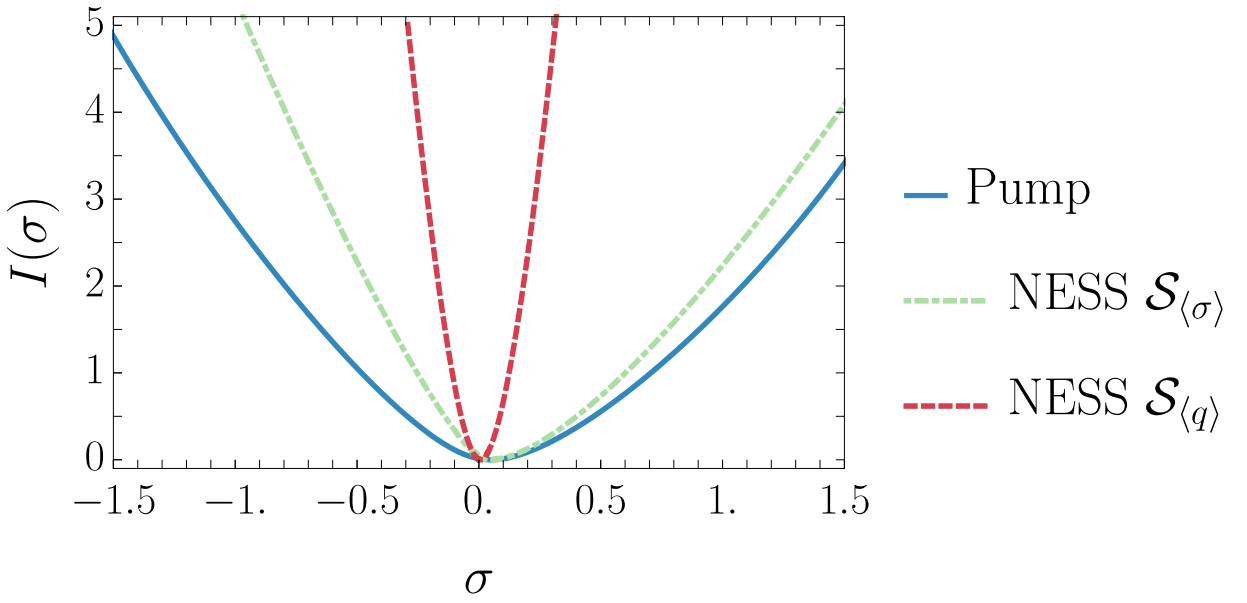

Figure 3.6: The large deviation rate functions for entropy production reveal that the steady state that recapitulates the current fluctuations has a smaller average entropy production. Furthermore, the extent of entropy production fluctuations in the corresponding steady state is much less pronounced. $\mathcal{S}_{\langle\sigma\rangle}$, on the other hand, leads to greatly enhanced entropy production fluctuations.

3.7 Excess entropy production

With the choice of $\mathcal{S}_{\langle q\rangle}$, both the average entropy production rate and its fluctuations in the nonequilibrium steady state representation differ markedly from the corresponding stochastic pump, as shown in Fig. 3.6. The "excess" entropy production has a physical origin and can be explained with a natural decomposition of the stochastic pump entropy production. Unlike nonequilibrium steady states, which can only produce entropy around closed cycles, stochastic pumps can produce entropy without completing a cycle [53]. We decompose the total stochastic pump entropy production rate into a contribution from the steady state, akin to the "housekeeping heat", and the excess associated with the pumping protocol [25, 105],

$$\sigma^{\text{pump}} = \sigma^{\text{ss}} + \sigma^{\text{ex}}, \tag{3.49}$$

where,

$$\sigma^{\text{ss}}_{ij} = \hat{j}_{ij} \ln \frac{\hat{q}_{ij}}{\hat{q}_{ji}}, \tag{3.50}$$

and

$$\sigma_{ij}^{\mathrm{ex}} = \frac{1}{\tau}\int_0^\tau j_{ij}\left(\ln\frac{q_{ij}}{q_{ji}} - \ln\frac{\hat{q}_{ij}}{\hat{q}_{ji}}\right). \tag{3.51}$$

The second law of thermodynamics ensures that both σ^{pump} and σ^{ss} are non-negative on average. This decomposition is analogous to the decomposition of entropy production used to describe the "housekeeping heat", i.e., the amount of heat required to maintain a nonequilibrium steady state [25, 53, 105].

The excess entropy produced by the stochastic pump, σ^{ex} is also non-negative. The inequality

$$\frac{1}{\tau}\int_0^\tau j_{ij}\ln\frac{q_{ij}}{q_{ji}} \geq \hat{j}_{ij}\ln\frac{\hat{q}_{ij}}{\hat{q}_{ji}}, \tag{3.52}$$

is known as the log-sum inequality and follows directly from Jensen's inequality, because $q_{ij}(t) > 0$ and $x\ln x$ is a convex function [111]. In the adiabatic limit, the system remains in the instantaneous equilibrium distribution and σ^{ex} vanishes. In this limiting case, $\mathcal{S}_{\langle\sigma\rangle} = \mathcal{S}_{\langle q\rangle}$. That is, for slow driving, entropy is only produced in the long time limit if probability is pumped through the network on average. While one might hope to match both the current and entropy production fluctuations when mapping a stochastic pump to a nonequilibrium steady state, or vice versa, this can only be achieved if the pumping protocol is adiabatic. As a consequence, in the inverse mapping problem, a pump protocol cannot generally be designed to mimic both current and entropy production fluctuations because the average entropy productions only agree in the adiabatic limit. Choosing the time-independent rate matrix so that it gives the steady state entropy production of the stochastic pump, that is, choosing $\mathcal{S}_{\langle q\rangle}$, yields a coarse-graining that is consistent with the physical mechanism by which entropy is produced in the pump.

In a stochastic pump, the hopping statistics along each edge need not be Poissonian, even in the adiabatic limit [112]. Therefore, the instantaneous dynamics of the nonequilibrium steady state, for which all transitions are purely Poissonian, may not perfectly recapitulate the behavior of the pump. The mapping determined by the choice (3.37) yields a universal bound on current fluctuations in weakly driven stochastic pumps, akin to the thermodynamic uncertainty relations recently discovered for nonequilibrium steady states [62, 81–83, 85]. Distinct behavior can be achieved with random driving: Barato and Seifert recently showed that, through the use of a driving protocol that changes at stochastic times, current fluctuations can be suppressed without incurring significant dissipation [113]. For the deterministic protocols considered here, in the perturbative limit, the rate function for any generalized current j is subject to a quadratic bound determined by the steady state entropy production rate

$$I^{\mathrm{pump}}(j) \leq \frac{(j-\hat{j})^2}{4\hat{j}^2/\sigma^{\mathrm{ss}}}. \tag{3.53}$$

The bound is tighter than the quadratic function of the full entropy production because incorporating the excess entropy production only reduces the curvature of the quadratic form. The lack of Poisson statistics for the pump suggests that the bound does not follow

from level-2.5 large deviation in all cases, but numerical evidence demonstrates that it is quite robust.

Chapter 4

The geometry of near equilibrium response

When a physical system is perturbed only slightly away from its equilibrium state, its response is determined by equilibrium fluctuations. The relationship between fluctuations and response is made explicit in the fluctuation-dissipation theorem discussed in detail in Chapter 3. When the external drive is sufficiently weak, the response function is determined by the covariance and time correlations of the perturbed system. Higher cumulants of the system's fluctuations can be neglected and the dissipation is entirely determined by Gaussian statistics. The mathematical properties of this spectrum of fluctuations enable a compelling mathematical analogy: In this limit, we can take advantage of connections with Riemannian geometry and statistical estimation theory to predict the response to complex, time-dependent external perturbations.

4.1 Geometry and statistical estimation

The free energies that we often wish to measure in physical chemistry and biophysics have a natural relation to normalization constants in statistics, as discussed in the Introduction 1. The problem of computing a free energy difference between two thermodynamic states of a physical system can be viewed, in this light, as a statistical estimation problem. The free energy difference between two thermodynamic states is

$$-\beta\Delta F = \ln \frac{Z\big(\boldsymbol{\lambda}(t)\big)}{Z\big(\boldsymbol{\lambda}(0)\big)}, \tag{4.1}$$

where $\boldsymbol{\lambda}$ is a vector of coupling parameters specifying the state, e.g., temperature, pressure, and particle number. To provide a procedure for estimating ΔF, we can rewrite this expression using the trivial equality

$$\Delta F = \int_0^t d\tau \; \frac{d}{dt} \ln Z\big(\boldsymbol{\lambda}(\tau)\big). \tag{4.2}$$

The free energy difference does not depend on the path of $\boldsymbol{\lambda}$, because it is a state function. However, an estimate of ΔF, which we denote by $\widehat{\Delta F}$, will, in general, depend on the path, which we call a "protocol". The derivative of the free energy with respect to the protocol time is related to an equilibrium average of the response to changes in the control parameters

$$\frac{d}{dt}\ln Z\big(\boldsymbol{\lambda}(\tau)\big) = -\left\langle \frac{\partial \beta E}{\partial \lambda_i} \right\rangle \dot{\lambda}_i. \tag{4.3}$$

The repeated indices are summed, a convention used throughout this chapter. This equality indicates that we can write an unbiased estimator for the free energy difference

$$\widehat{\Delta F} = -\int_0^t d\tau \; N^{-1} \sum_{j=1}^{N} \left(\frac{\partial \beta E}{\partial \lambda_i} \right)^{(j)} \dot{\lambda}_i. \tag{4.4}$$

Operationally, the scheme described here is similar to thermodynamic integration [41].

The fact that the estimator is unbiased, meaning that $\left\langle \widehat{\Delta F} \right\rangle = \Delta F$ is a desirable property, but the optimal estimator will additionally minimize the variance. The variance of the estimator

$$\left\langle \left(\widehat{\Delta F} \right)^2 \right\rangle - (\Delta F)^2 \tag{4.5}$$

depends on the protocol $\boldsymbol{\lambda}(t)$. We want to minimize the average variance over all possible choices of protocol. This quantity is strictly positive, so we seek to minimize it directly. Explicitly, the term to be minimized is

$$\left\langle \left(\widehat{\Delta F} \right)^2 \right\rangle - (\Delta F)^2 = \int_0^t d\tau \; \dot{\lambda}_i \left(\left\langle \frac{\partial \beta E}{\partial \lambda_i} \frac{\partial \beta E}{\partial \lambda_j} \right\rangle - \left\langle \frac{\partial \beta E}{\partial \lambda_i} \right\rangle \left\langle \frac{\partial \beta E}{\partial \lambda_j} \right\rangle \right) \dot{\lambda}_j. \tag{4.6}$$

The expression inside the average is symmetric under exchange of i and j and is positive semi-definite. We can rewrite it as

$$g_{ij}(\boldsymbol{\lambda}) = \langle \delta \boldsymbol{X}_i \delta \boldsymbol{X}_j \rangle_{\boldsymbol{\lambda}} \tag{4.7}$$

where the quantity $\delta \boldsymbol{X}$ defines the deviation of the response field from its average value,

$$\delta X_i = \left(\frac{\partial \beta E}{\partial \lambda_i} \right) - \left\langle \frac{\partial \beta E}{\partial \lambda_i} \right\rangle. \tag{4.8}$$

The tensor $g_{ij}(\boldsymbol{\lambda})$ defines a Riemannian metric on the space of protocols, specifying the geometric distortion of the protocol manifold at each value of the control parameters. The meaning of this tensor and its use in nonequilibrium protocol optimization will be the focus of this chapter. The choices of $\boldsymbol{\lambda}(t)$ that minimize Eq. (4.7) are geodesics on the manifold of control parameters. This metric, which endows the space of protocols with an intrinsic geometric structure, turns out to have a much deeper connection to statistical mechanics. As we will see throughout the next two Chapters, we can leverage generalizations of the minimum variance statistical estimator to determine minimum dissipation protocols for nonequilibrium control.

4.2 Thermodynamic length

The Gaussian spectrum of equilibrium fluctuations about each thermodynamic state associates with that state a metric tensor, analogous to the Fisher information metric of Eq. (4.7). The geometric interpretation of the manifold of thermodynamic states was initiated in the 1970s in Refs. [114, 115], which highlighted the mathematical parallels between thermodynamics and Riemannian geometry. Salamon and Berry [47] and Schlögl [116] later applied geometric thermodynamics to study macroscopic nonequilibrium systems. It should be noted that many of the essential ingredients necessary to analyze such problems geometrically can be found in the seminal work of Onsager and Machlup on irreversible processes [43, 44].

In applications to nonequilibrium systems, we investigate the response to changes in a vector of external control parameters. Throughout, we consider control parameters $\boldsymbol{\lambda} \in \mathbb{R}^d$ which couple linearly to observables $\boldsymbol{X} = -\nabla_{\boldsymbol{\lambda}} \beta E(\boldsymbol{x}, \boldsymbol{\lambda})$, where $\boldsymbol{x}$ is a configuration variable in $\mathbb{R}^n$. Generally, we will consider changes the control parameters in a deterministic fashion $\boldsymbol{\lambda}(t) : [0, t_{\rm obs}] \to \mathbb{R}^d$ as a means to transform the system between two thermodynamic states. This metric defined in Eq. (4.7) directly quantifies the dissipation in the "endoreversible" limit, which stipulates that the system remains at all times in an equilibrium state, though not necessarily the equilibrium state given by the instantaneous value of the control parameters [47].

The connections between the Riemannian geometry of thermodynamics and the Fisher information metric were explored in Ref. [117], which also emphasized the relationship to estimating free energies with the Bennett Acceptance Ratio method. Sivak and Crooks [118] revisited the endoreversibility assumption and argued that the metric should include additional time correlations (cf. Eq. (4.12)) to properly account for lag in the linear response regime. The formalism of thermodynamic length has been applied to control problems in analytically tractable models, where metric can be computed exactly [119]. In most applications to physical systems, however, a closed form expression for the metric tensor (4.12) cannot be derived. Chapter 5 describes computational approaches to determining the geodesic minimum dissipation protocol when the thermodynamic metric must be estimated by sampling the dynamics.

An explicit form for the thermodynamic metric can be derived by minimizing, over all possible protocols with fixed endpoints, the average excess work done in the course of the transformation. In Sec. 4.3 we prove that the minimum dissipation protocol is a geodesic on the manifold of control parameter values when the rate of driving is slow relative to the dynamics. We consider a system with coordinates $\boldsymbol{x} \in \mathbb{R}^d$ and control the system with a time-dependent, nonequilibrium protocol $\boldsymbol{\lambda}(t) = (\lambda_1(t), \ldots, \lambda_N(t)) \in \mathbb{R}^N$, $t \in [0, t_{\rm obs}]$, for some $t_{\rm obs} > 0$. As we will see below our results will be independent of $t_{\rm obs}$ provided that the protocol is slow relative to the dynamics of $\boldsymbol{x}$. We assume that we can independently tune the components of $\boldsymbol{\lambda}$, which we refer to as the "control parameters". The dynamics of the system is governed by an overdamped Langevin equation with a time-dependent potential

V that depends parametrically on the protocol

$$\dot{\boldsymbol{x}} = -\frac{1}{\epsilon}\nabla V\big(\boldsymbol{x}(t), \boldsymbol{\lambda}(t)\big) + \sqrt{\frac{2}{\beta\epsilon}}\eta(t). \tag{4.9}$$

The reciprocal temperature is denoted by β and $\eta(t)$ is Gaussian random noise with zero mean and covariance $\langle \eta_i(t)\ \eta_j(t')\rangle = \delta_{ij}\delta(t-t')$. The parameter $\epsilon \ll 1$ is proportional to the friction coefficient for the dynamics of the system and sets a separation of timescales between the system and the protocol: when ϵ is small the dynamics of the underlying system $\boldsymbol{x}(t)$ are much faster than the changes in protocol $\boldsymbol{\lambda}(t)$.

An optimal protocol $\boldsymbol{\lambda}(t)$ minimizes the *average* microscopic work $\langle \mathcal{W}_\epsilon \rangle$, where the expectation, denoted $\langle\cdot\rangle$, is performed over stochastic trajectories $\boldsymbol{x}(t)$ which begin in equilibrium. In the limit of infinitely slow driving, the system remains in equilibrium at every point in time and the transformation is thermodynamically reversible. If the system is driven by the protocol at a finite rate, then work must be done and a positive amount of energy is dissipated on average. For a fixed, deterministic protocol $\boldsymbol{\lambda}(t)$, the heat absorbed by the bath can be computed as a stochastic integral [120],

$$\tilde{\mathcal{W}}_\epsilon = -\epsilon^{-1}\int_0^{t_{\text{obs}}} \nabla V\big(\boldsymbol{x}(t), \boldsymbol{\lambda}(t)\big) \circ d\boldsymbol{x}(t), \tag{4.10}$$

where $\circ$ denotes the time-symmetric Stratonovich product, described in detail in Appendix A. The expression for the heat (4.10) can be related to the familiar stochastic thermodynamic expression for work,

$$\mathcal{W}_\epsilon = \epsilon^{-1}\int_0^{t_{\text{obs}}} \partial_{\boldsymbol{\lambda}} V\big(\boldsymbol{x}(t), \boldsymbol{\lambda}(t)\big)\cdot \dot{\boldsymbol{\lambda}}\ dt, \tag{4.11}$$

by noting that $dV\big(\boldsymbol{x}(t), \boldsymbol{\lambda}(t)\big) = \nabla V\big(\boldsymbol{x}(t), \boldsymbol{\lambda}(t)\big)\circ d\boldsymbol{x}(t) + \partial_{\boldsymbol{\lambda}} V\big(\boldsymbol{x}(t), \boldsymbol{\lambda}(t)\big)\cdot\dot{\boldsymbol{\lambda}}\ dt$ and integrating by parts. The quantity $\tilde{\mathcal{W}}_\epsilon$ differs from $\mathcal{W}_\epsilon$ by a boundary term, which does not depend on the protocol itself but only its endpoints. Its contribution to the overall cost of control is fixed, and therefore can be ignored in our optimization problem. It should be noted, however, that the boundary term could still make a very significant contribution to the dissipation when the duration of the protocol is finite.

In Sec. 4.3 we prove that, in the limit of small ϵ, a natural metric for the dissipation along a fixed protocol $\boldsymbol{\lambda}(t)$ emerges. This metric has the form of a friction tensor [118] and quantifies the energetic cost of driving the system,

$$\zeta(\boldsymbol{\lambda}) = \int_0^\infty d\tau\ \big\langle \delta\boldsymbol{X}\big(\boldsymbol{x}, \boldsymbol{\lambda}\big)\delta\boldsymbol{X}^{\mathrm{T}}\big(\boldsymbol{x}^{\boldsymbol{\lambda}}(\tau), \boldsymbol{\lambda}\big)\big\rangle_{\boldsymbol{\lambda}}, \tag{4.12}$$

where $\boldsymbol{X}(\boldsymbol{x}, \boldsymbol{\lambda}) \equiv -\beta\partial_{\boldsymbol{\lambda}} V(\boldsymbol{x}, \boldsymbol{\lambda})$ and $\boldsymbol{x}^{\boldsymbol{\lambda}}(\tau)$ denotes the solution to (4.9) using the rescaled time $\tau = t/\epsilon$ and keeping the control parameters $\boldsymbol{\lambda}$ fixed. The notation $\langle\cdot\rangle_{\boldsymbol{\lambda}}$ denotes an equilibrium average with the control parameters $\boldsymbol{\lambda}$ fixed, as well. In particular, it should

be noted that $\boldsymbol{x}$ is initially drawn from and samples an equilibrium distribution in the expectation above. The length functional is then given by,

$$\mathcal{L}[\boldsymbol{\lambda}] = \int_0^{t_{\text{obs}}} dt \, \sqrt{\dot{\boldsymbol{\lambda}}^T \zeta\big(\boldsymbol{\lambda}(t)\big)\dot{\boldsymbol{\lambda}}}, \tag{4.13}$$

and this length is independent of the parameterization of $\boldsymbol{\lambda}(t)$ as well as t_{obs}, which we could set to $t_{\text{obs}} = 1$. If we impose the constraint of constant speed along the protocol, then it suffices to perform a minimization over the energy functional $\mathcal{E}[\boldsymbol{\lambda}]$, in which the integrand lacks the square root term, see (4.26).

To perform this minimization, we start with the Euler-Lagrange equation for the geodesic minimizing (4.13). Written componentwise, this equation reads

$$\begin{aligned} & \frac{d}{dt}\left(\zeta_{kj}\dot{\lambda}_j\right) = \frac{1}{2}\dot{\lambda}_i \frac{\partial}{\partial \lambda_k}\zeta_{ij}\dot{\lambda}_j \\ \Leftrightarrow \qquad & \zeta_{kj}\ddot{\lambda}_j + \frac{\partial \zeta_{kj}}{\partial \lambda_i}\dot{\lambda}_j\dot{\lambda}_i - \frac{1}{2}\frac{\partial \zeta_{ij}}{\partial \lambda_k}\dot{\lambda}_i\dot{\lambda}_j = 0. \end{aligned} \tag{4.14}$$

At this point, we can take advantage of the symmetry in the cumulants. When the relaxation time is constant, the metric is proportional to the Fisher information metric. As a result, the derivatives of the metric correspond to time-scaled third cumulants and are invariant under the permutation of indices. Under this assumption, the expression (4.14) simplifies as,

$$\zeta_{kj}\ddot{\lambda}_j + \frac{1}{2}\frac{\partial \zeta_{ij}}{\partial \lambda_k}\dot{\lambda}_i\dot{\lambda}_j = 0. \tag{4.15}$$

which we will write compactly using vectorial notation as $\zeta\ddot{\boldsymbol{\lambda}} + \frac{1}{2}\partial_{\boldsymbol{\lambda}}\zeta : \dot{\boldsymbol{\lambda}}\dot{\boldsymbol{\lambda}} = 0$.

With this geometric perspective, we can efficiently compute minimum dissipation protocols using minimum action methods [121, 122]. In Chapter 5, we detail an algorithm that iteratively updates a trial nonequilibrium protocol and converges to the optimum. The update step depends only on fluctuations in the equilibrium dynamics at points along the protocol. The principal advantage of this method over alternative numerical approaches is its computational power, remaining robust even when the protocol spaces are very high dimensional.

4.3 Derivation of the thermodynamic metric

The expression (4.11) defines the work done on the system for a single realization of its stochastic dynamics. When there are substantial fluctuations in the microscopic variables $\boldsymbol{x}$, the work $\mathcal{W}_\epsilon$ itself is a fluctuating quantity, as Eq. (4.11) depends on the dynamics. To identify an efficient protocol, we want to find a $\boldsymbol{\lambda}(t)$ that minimizes the average dissipation, as opposed to focusing on rare trajectories of the controlled system $\boldsymbol{x}(t)$ that yield anomalously low dissipation. To compute the average over trajectories, we introduce an undetermined

configurational distribution $\rho(\boldsymbol{x}, t)$, which varies with time throughout the duration of the protocol,

$$\langle \mathcal{W}_\epsilon \rangle = \int_0^{t_{\text{obs}}} \int_{\mathbb{R}^d} dt \; d\boldsymbol{x} \; \partial_{\boldsymbol{\lambda}} V\big(\boldsymbol{x}, \boldsymbol{\lambda}(t)\big) \cdot \dot{\boldsymbol{\lambda}} \; \rho(\boldsymbol{x}, t). \tag{4.16}$$

The distribution ρ satisfies a Fokker-Planck equation associated with the Langevin equation (4.9),

$$\partial_t \rho = \epsilon^{-1} \nabla \cdot \big(\nabla V\big(\boldsymbol{x}, \boldsymbol{\lambda}(t)\big) \rho + \beta^{-1} \nabla \rho \big) . \tag{4.17}$$

Because the driving is slow ($\epsilon \ll 1$), we expand ρ around the equilibrium distribution,

$$\rho_0(\boldsymbol{x}, t) = Z\big(\boldsymbol{\lambda}(t)\big)^{-1} \; e^{-\beta V(\boldsymbol{x}, \boldsymbol{\lambda}(t))}, \tag{4.18}$$

at each point along the protocol,

$$\rho(\boldsymbol{x}, t) = \rho_0(\boldsymbol{x}, t) \left(1 + \epsilon \phi(\boldsymbol{x}, t) + \mathcal{O}(\epsilon^2)\right) \tag{4.19}$$

where $Z(\boldsymbol{\lambda}) = \int_{\mathbb{R}^d} d\boldsymbol{x} \; e^{-\beta V(\boldsymbol{x}, \boldsymbol{\lambda})}$ denotes the partition function for a fixed value of the control vector $\boldsymbol{\lambda}$. Using this expansion in the Fokker-Planck equation (4.17) we find that the order ϵ correction ϕ satisfies

$$\partial_t \ln \rho_0(\boldsymbol{x}, t) = \nabla V\big(\boldsymbol{x}, \boldsymbol{\lambda}(t)\big) \cdot \nabla \phi(\boldsymbol{x}, t) + \beta^{-1} \Delta \phi(\boldsymbol{x}, t). \tag{4.20}$$

The left hand side of (4.20) is explicitly

$$\partial_t \ln \rho_0(\boldsymbol{x}, t) = \beta \left(-\partial_{\boldsymbol{\lambda}} V\big(\boldsymbol{x}, \boldsymbol{\lambda}(t)\big) \cdot \dot{\boldsymbol{\lambda}} + \partial_{\boldsymbol{\lambda}} F\big(\boldsymbol{\lambda}(t)\big) \cdot \dot{\boldsymbol{\lambda}} \right) \tag{4.21}$$

$$= \equiv -\beta \delta \boldsymbol{X}(\boldsymbol{x}, \boldsymbol{\lambda}(t)) \cdot \dot{\boldsymbol{\lambda}}, \tag{4.22}$$

where $F(\boldsymbol{\lambda})$ denotes the free energy $-\beta^{-1} \ln Z(\boldsymbol{\lambda})$. The solution to the differential equation (4.20) can be expressed via the Feynman-Kac formula as an average over a virtual fast process $\boldsymbol{x}^{\boldsymbol{\lambda}}(\tau)$ in which the control parameters are kept at a fixed value $\boldsymbol{\lambda}$. The process $\boldsymbol{x}^{\boldsymbol{\lambda}}(\tau)$ satisfies an overdamped equation of motion, with the initial condition $\boldsymbol{x}^{\boldsymbol{\lambda}}(0) = \boldsymbol{x}$,

$$\frac{d}{d\tau} \boldsymbol{x}^{\boldsymbol{\lambda}}(\tau) = -\nabla V(\boldsymbol{x}^{\boldsymbol{\lambda}}(\tau), \boldsymbol{\lambda}) + \sqrt{\frac{2}{\beta}} \; \eta(t). \tag{4.23}$$

Denoting by $\langle \cdot \rangle_{\boldsymbol{\lambda}}$ an expectation taken over this process with the initial condition as above, we have,

$$\phi(\boldsymbol{x}, t) = -\beta \left(\int_0^\infty d\tau \; \left\langle \delta \boldsymbol{X}(\boldsymbol{x}^{\boldsymbol{\lambda}(t)}(\tau), \boldsymbol{\lambda}(t)) \right\rangle_{\boldsymbol{\lambda}(t)} \right) \cdot \dot{\boldsymbol{\lambda}}(t). \tag{4.24}$$

The solution for $\phi(\boldsymbol{x}, t)$ gives us an explicit expression for the configurational density (4.19) up to order ϵ,

$$\rho(\boldsymbol{x}, t) = Z^{-1}(\boldsymbol{\lambda}(t)) e^{-\beta V(\boldsymbol{x}, \boldsymbol{\lambda}(t))} \times \left(1 - \epsilon \beta \left(\int_0^\infty d\tau \; \left\langle \delta \boldsymbol{X}(\boldsymbol{x}^{\boldsymbol{\lambda}(t)}(\tau), \boldsymbol{\lambda}(t)) \right\rangle_{\boldsymbol{\lambda}(t)} \right) \cdot \dot{\boldsymbol{\lambda}}(t) \right) + \mathcal{O}(\epsilon^2). \tag{4.25}$$

With this expression for the configurational distribution ρ we can compute the average excess microscopic work (4.16). The ϵ^0 term is just the free energy difference between the initial and final points of the protocol, and thus has no path dependence. The work performed on the system in excess of that required to overcome the free energy difference at the end points of the protocol is quantified by the higher order terms of $\langle \mathcal{W}_\epsilon \rangle$. The leading order contribution can be expressed as

$$\mathcal{E}[\boldsymbol{\lambda}] = \beta \int_0^{t_{\text{obs}}} dt \; \dot{\boldsymbol{\lambda}}^{\mathrm{T}} \zeta(\boldsymbol{\lambda}(t)) \dot{\boldsymbol{\lambda}}, \tag{4.26}$$

where $\zeta(\boldsymbol{\lambda})$ is the tensor we defined in (4.12). This tensor is a positive semi-definite, symmetric, bilinear form, meaning that it defines a semi-Riemannian metric on the space of nonequilibrium protocols. The metric is related to Kirkwood's linear-response formulation of the friction tensor and can be interpreted as quantifying resistance to changes in the control parameters [63, 118]. Minimizing the length of a protocol with respect to the metric ζ minimizes the excess work, meaning that geodesics in this space are minimum dissipation protocols [117, 118]. One advantage of this "geometric" description is that the time over which the protocol does not alter it, so long as the separation of timescales between the dynamics and control is maintained. However, because the excess work scales inversely with the duration of the protocol [118], smaller values of t_{obs} lead to higher average dissipation.

In what follows, we have made the additional simplification that the Green-Kubo integral in (4.12) can be approximated as a covariance multiplied by an effective timescale τ_{eff}, that is,

$$\begin{aligned} \zeta(\boldsymbol{\lambda}) &= \int_0^\infty \left\langle \delta \boldsymbol{X}(0) \delta \boldsymbol{X}^T(t) \right\rangle_{\boldsymbol{\lambda}} dt \\ &\approx \tau_{\text{eff}} \left\langle \delta \boldsymbol{X}(0) \delta \boldsymbol{X}^T(0) \right\rangle_{\boldsymbol{\lambda}}. \end{aligned} \tag{4.27}$$

We need not make the approximation (4.27), but doing so simplifies the algorithm. We employ this simplified variant throughout Chapter 5, an assumption justified in the examples we consider.

4.4 Connection between thermodynamic metrics and stochastic optimal control

Consider a system with coordinates $\boldsymbol{X} \in \mathbb{R}^n$. An external protocol $\lambda : [0, t_{\text{obs}}] \to \mathbb{R}^d$ is used to control the system as function of time. The dynamics of X under the control of the protocol λ is governed by a stochastic differential equation,

$$dX_t = f(X_t, \lambda_t, t)dt + d\xi_t, \tag{4.28}$$

where f is a drift function and $d\xi$ is a Wiener process with $\langle d\xi_i d\xi_j \rangle = D_{ij} dt$.

We associate a cost function $\mathcal{C}(X_0;\lambda)$ with this control problem and define it as a sum of the cost to reach some target at time t_{obs} and the cost of the dynamics along the trajectory,

$$\mathcal{C}(X_t;\lambda,t,t_{\text{obs}}) = \Big\langle \phi\big(X(t_{\text{obs}})\big) + \int_t^{t_{\text{obs}}} R(X_\tau,\lambda_\tau,\tau)d\tau \Big\rangle_{X_t}, \tag{4.29}$$

where $\langle\cdot\rangle_{X_t}$ denotes an average over trajectories that begin at X_t at time t. The optimal cost associated with some transformation is the cost associated with the optimal protocol,

$$J(X_t,t) = \min_{\lambda} \mathcal{C}(X_t;\lambda,t,t_{\text{obs}}). \tag{4.30}$$

Bellman's principle of dynamic programming [123] says that an optimal protocol starting from any initial condition must remain optimal. This observation allows us to write a recursive equation for the optimal cost function J,

$$J(X_t,t) = \min_u R(X_t,\lambda_t,t) + \langle J(X_{t+dt},t+dt)\rangle_{X_t}. \tag{4.31}$$

Performing a Taylor expansion in X and t yields a stochastic version of the classical Hamilton-Jacobi-Bellman equation,

$$\langle J(X_{t+dt},t+dt)\rangle_{X_t} = J(X_t,t) + \partial_t J(X_t,t) + \nabla J(X_t,t)\langle dX\rangle_{X_t} + \nabla^2 J(X_t,t)\langle dX^2\rangle_{X_t} \tag{4.32}$$

$$= J(X_t,t) + \partial_t J(X_t,t)\ dt + \nabla J(X_t,t)\cdot f(X_t,\lambda_t,t)\ dt + \frac{1}{2}\nabla^2 J(X_t,t) : D\ dt + \mathcal{O}(dt^2), \tag{4.33}$$

noting that $\langle dX^2\rangle$ is a stochastic increment of order dt so must be retained in the expansion (cf. Appendix A). Using this equation in (4.31), we get a differential equation for the optimal cost function,

$$-\partial_t J(X_t,t) = \min_u R(X_t,\lambda_t,t) + \nabla J(X_t,t)\cdot f(X_t,\lambda_t,t)\ dt + \frac{1}{2}\nabla^2 J(X_t,t) : D\ dt, \tag{4.34}$$

which is known as the Hamilton-Jacobi-Bellman equation. The notation $A : B$ denotes the dyadic product of the matrices A and B.

Solving the nonlinear stochastic partial differential equation (4.34) is a formidable task. However, in special cases, some progress can be made. An important special case is that of "path integral" control, in which the cost function has the following form,

$$\mathcal{C}(X_t;\lambda,t,t_{\text{obs}}) = \Big\langle \phi\big(X(t_{\text{obs}})\big) + \int_t^{t_{\text{obs}}} \lambda_\tau^{\mathrm{T}}\zeta\lambda_\tau + V(X_\tau,\tau)d\tau \Big\rangle_{X_t}. \tag{4.35}$$

The control parameters appear as a quadratic cost in the integral. We further assume that the controls couple linearly into the dynamics of the system. This means that we can write (4.28) as,

$$dX_t = \big(f(X_t,t) + \lambda_t\big)dt + d\xi_t, \tag{4.36}$$

and the minimization in (4.34) can be solved explicitly as,

$$\lambda_t^* = -\zeta^{-1}\nabla J(X_t, t). \tag{4.37}$$

Under these assumptions, it is possible to linearize the Hamilton-Jacobi-Bellman equation, using a Cole-Hopf transformation. We let,

$$J(X_t, t) = -\gamma \log \psi(X_t, t), \tag{4.38}$$

so that,

$$\lambda \frac{\partial_t \psi}{\psi} = -\frac{1}{2}\gamma^2 \zeta^{-1} : \left(\frac{\nabla\psi}{\psi}\right)^{\mathrm{T}} \left(\frac{\nabla\psi}{\psi}\right) - \gamma f^{\mathrm{T}} \frac{\nabla\psi}{\psi} + \frac{1}{2}\gamma D : \left[\frac{\nabla^2\psi}{\psi} - \left(\frac{\nabla\psi}{\psi}\right)^{\mathrm{T}} \left(\frac{\nabla\psi}{\psi}\right)\right]. \tag{4.39}$$

As pointed out by Kappen (cf., Refs. [124, 125]), the nonlinearity vanishes if there exists a scalar γ such that,

$$\gamma\zeta^{-1} = D. \tag{4.40}$$

equivalently, when the cost of control is proportional to the inverse diffusion tensor, the optimal cost satisfies a linearized Hamilton-Jacobi-Bellman equation.

4.5 The dynamical Riemannian geometry of the Ising model

While studying model systems helps us glean the general principles of non-equilibrium control, theoretical analysis has thus far been restricted to single-body systems with exactly solvable dynamics [119, 126–129] or in which the dynamics is not incorporated [130, 131]. For most systems of interest we cannot compute the metric (4.12) exactly. Thus, we must develop a general method to predict optimal protocols from incomplete knowledge of the metric tensor, often relying on numerical estimates.

The Ising model is a cornerstone of statistical mechanics that captures the essential physics of a diverse set of systems including ferromagnets, liquid-vapor phase transitions, and lipid membranes [132, 133]. With this example, we gain insight into the unexplored consequences of non-linear dynamics and the presence of a phase transition on optimal protocols. Control can be exercised by applying an external field, but also by varying the spin-spin coupling, as in heat assisted magnetic recording [134]. Applications include magnetic information storage technologies that rely on inverting the net magnetization of microscopic spin domains as well as technologies for ultra low energy computation, such as hybrid spintronics [135]. Low dissipation control of seemingly simple, stochastic systems, such as spins on a magnetic hard drive, has implications for the efficiency of nanodevices already in wide use.

As an illustrative example and proof of practical significance, we will consider the problem of inverting the net magnetization of the 2D Ising model using the spin-spin coupling

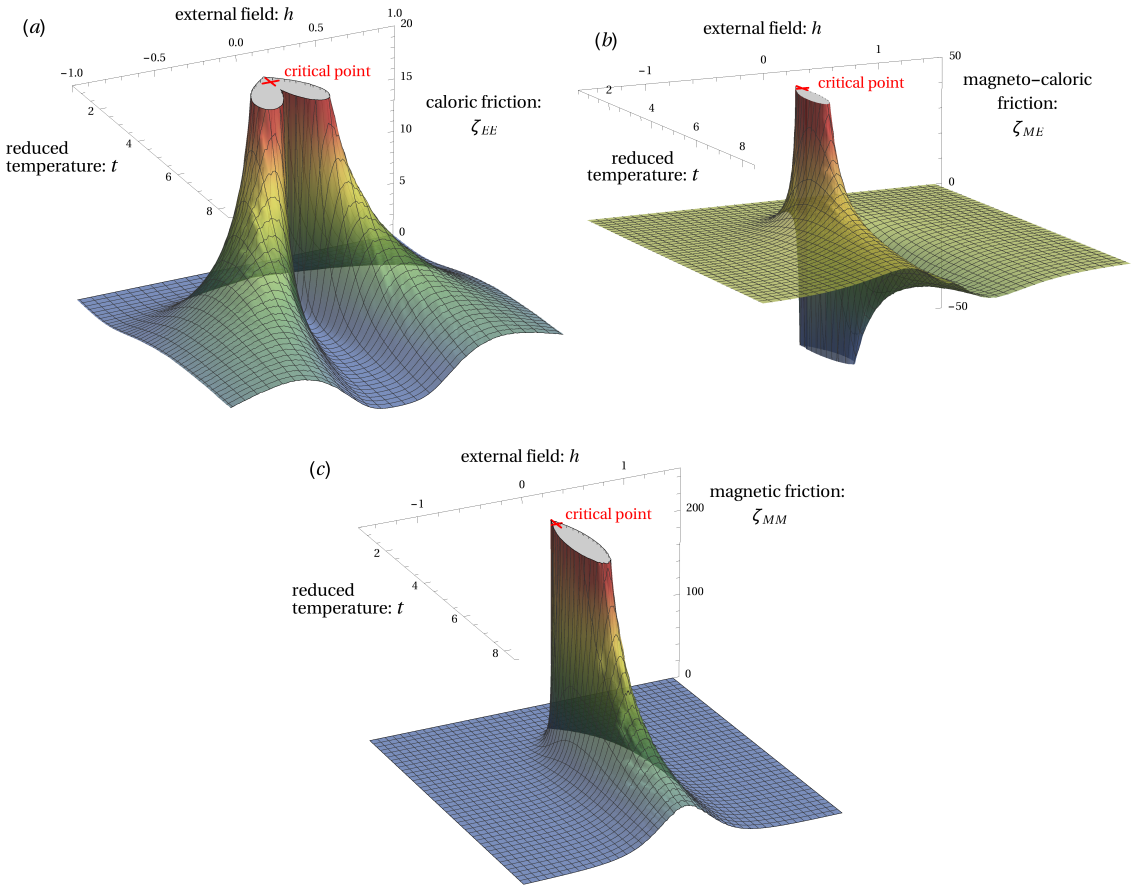

Figure 4.1: The caloric (a), magneto-caloric (b), and magnetic (c) friction coefficients of the 2D Ising model, as defined by Eq. (4.12), plotted in the magnetic field (h), temperature $T = 1/\beta J$ plane. Both relaxation times and static correlations diverge at the critical point which gives rise to the cusp in each of these plots. The friction coefficients are the matrix elements of a Riemannian metric with the property that geodesics minimize the average excess work that a protocol exercises over the system.

and external field as control parameters. The system is governed by the standard Ising Hamiltonian,

$$\mathcal{H} = h(t) \sum_{i=1}^{n} s_i + J(t) \sum_{\langle i,j \rangle} s_i s_j \ , \tag{4.41}$$

where $\langle i, j \rangle$ denotes a sum over all nearest neighbor pairs on the lattice, and the control parameter $\lambda(t) = \big(\beta h(t), \beta J(t)\big)$, varies the coupling, J, and the external magnetic field, h, with time. Controlling the strength of the spin-spin coupling can also be implemented by varying the temperature.

If we drive at a finite rate the system resists the changes in the control parameters. In the linear regime, the friction ζ that the protocol encounters is given by Eq. (4.12). When we control the field and coupling, the conjugate forces are the net magnetization M and internal energy E,

$$X_{\beta h}(t) = \sum_{i=1}^{n} s_i \equiv M, \tag{4.42}$$

$$X_{\beta J}(t) = \sum_{\langle i,j \rangle} s_i s_j \equiv E. \tag{4.43}$$

Similar expressions for the friction (4.12) arise in Kirkwood's linear response formula [63, 118] and also in the study of effective diffusion constants under coarse-graining [136]. The metric defines the distance along a protocol λ,

$$\mathcal{L}[\lambda(t)] = \int_{\lambda} \sqrt{\dot{\lambda}^i \zeta_{ij} \dot{\lambda}^j}, \tag{4.44}$$

and the distance along an optimal protocol sets a lower bound on the excess work exercised by the controller over the system [47, 118],

$$\Delta t \langle W_{\text{ex}} \rangle \geq \mathcal{L}^2. \tag{4.45}$$

For any protocol, equality between the divergence $\Delta t \langle W_{\text{ex}} \rangle$ and the squared thermodynamic length $\mathcal{L}^2$ is achieved when the excess power is constant over the duration of the protocol. As a result, the path of an optimal protocol does not depend its duration [47, 117]. Exact equations for the relaxation of M and E are not known in general, so we must approximate the metric using simulations. We discretize the parameter space and at each point we compute the time correlation matrix for the conjugate forces,

$$\begin{pmatrix} \left\langle \delta X_{\beta h}(0)\ \delta X_{\beta h}(\tau) \right\rangle & \left\langle \delta X_{\beta h}(0)\ \delta X_{\beta J}(\tau) \right\rangle \\ \left\langle \delta X_{\beta J}(0)\ \delta X_{\beta h}(\tau) \right\rangle & \left\langle \delta X_{\beta J}(0)\ \delta X_{\beta J}(\tau) \right\rangle \end{pmatrix}. \tag{4.46}$$

The time correlation functions are estimated with Markov Chain Monte Carlo simulations on a 128 by 128 square lattice of Ising spins with Glauber dynamics [137]. We compared our results to a 256 by 256 system to ensure there were no significant finite size effects, aside from finite size scaling. Integrating the time correlation function (4.46) to infinite time yields the friction coefficient ζ_{ij} (4.12) at each point in the parameter space. In practice, correlations decay exponentially and the friction tensor can be accurately estimated, except very near the critical point.

Once the metric is known on a subspace of the parameter manifold, we recast the problem of approximating geodesic distances in terms of an *eikonal* equation, a partial differential equation commonly used to study wave propagation [138], as discussed further in Chapter 5.

In Fig. 4.1 we plot each of the components of the friction tensor. The caloric friction coefficient ζ_{EE} is the time autocorrelation of the internal energy. At each point in parameter space, this friction can be written as $\zeta_{EE} = \tau_{EE} \langle (\delta E)^2 \rangle = \tau_{EE} k_{\text{B}} T^2 C$, the product of

the heat capacity C and an effective timescale for the relaxation of the energy. Similarly, the cross-correlation of the magnetization and internal energy, the magneto-caloric friction $\zeta_{ME} = \zeta_{EM} = \tau_{EM} k_{\mathrm{B}} T M_t$ is proportional to the magneto-caloric coefficient M_t, and the autocorrelation of the magnetization, the magnetic friction $\zeta_{MM} = \tau_{MM}\chi$ is proportional to the magnetic susceptibility χ.

Both static correlations and relaxation timescales diverge near the critical point of the Ising model. These two effects compound to produce a singularity of the metric where all three components of friction tensor also diverge. The friction coefficients decay according to characteristic power laws in neighborhoods surrounding the critical point [139]. Correlations are also large exactly at the first order phase transition along the line $h = 0$, $T < T_{\mathrm{C}}$. However, spontaneous magnetization reversal is rarely observed in simulations under single spin flip dynamics. Below the critical temperature T_{C} with $h \neq 0$, relaxation times are fast and fluctuations are negligible, which results in small values for each component of the friction tensor.

The geometry of the supercritical region is more intricate. The caloric friction, Fig. 4.1 (a), exhibits symmetric ridges that correspond to maxima in the heat capacity and are reminiscent of "Widom lines" in supercritical fluids [140]. Along these ridges we observe large, slowly relaxing spin domains. The magneto-caloric friction, Fig. 4.1 (b), is antisymmetric about $h = 0$ due to the antisymmetry in the net magnetization. The magnetic friction, Fig. 4.1 (c), is large for an extended region above the critical temperature. At very high temperatures, all the components of the metric are again small due to negligible spin-spin couplings. The large diagonal elements of the metric along these supercritical lines are indicative of a high dissipative cost associated with control.

The form on an optimal protocols depends, of course, on what we can control. For instance, given spatial control of the external field, the minimum dissipation protocol may involve flipping spins at the boundary of a domain. High dimensional parameter spaces will require different approaches to calculating geodesics. Analogous problems in transition state theory have been addressed using the string method [141–143] and path sampling [144]. We adapt some of the cited strategies to the problem of protocol prediction in Chapter 5.

Non-equilibrium nanoscale machines need to be designed for objectives beyond low dissipation. If speed is the objective, the bound in Eq. (4.45) can be used to minimize the total duration of the protocol, while keeping the average dissipation fixed. Supercritical heat engines and magnetic refrigerators [145, 146] could also be studied using the Ising model, but in these cases the objective requires efficient energy transfer around a thermodynamic cycle. In such cases, we must include additional constraints when seeking efficient control. There may also be practical limits on the range of the control parameters. As an illustration, Fig. 5.1 shows minimum dissipation protocols where the maximum temperature is constrained. The optimal protocols we have predicted are weakly constrained where the manifold is flat, affording tremendous flexibility to the controller. Where the metric changes rapidly, protocols are tightly constrained and external control must be precise.

Chapter 5

Computational approaches to protocol prediction

The components of the metric tensor introduced in Chapter 4 are defined as integrals of time correlation functions for the generalized forces that respond to external control. The tensor takes on a different value for each point in the space of control parameters. For nonlinear dynamical systems, analytically evaluating the time correlations for even a single point in the space of all possible couplings may be prohibitively challenging. Thus, in order to make the framework of thermodynamic geometry useful for studying complex, physical systems, we must develop numerical approaches for predicting optimal protocols. This chapter discusses three such schemes, each of which is best suited to a distinct class of problems.

5.1 Exact approach: Fast marching method

The first approach we will describe is appropriate for finding protocols in the case that the metric tensor can be constructed numerically over the portion of parameter space of interest. For example, the metric may be computed from dynamical simulations as was done for the two dimensional Ising model in Chapter 4. Geodesics on two dimensional manifolds embedded in three dimensional space, in the absence of a Riemannian metric, have been computed using a numerical scheme based on finding level sets [147]. Once the metric is known on a subspace of the parameter manifold, we recast the problem of approximating geodesic distances in terms of an *eikonal* equation,

$$|\nabla T(t,h)| = 1/F(t,h), \tag{5.1}$$

a partial differential equation commonly used to study wave propagation [138]. The field F is the instantaneous speed of a wavefront and T represents the arrival time of the wave. In our case, F is the linearized Riemannian distance between neighboring points $\boldsymbol{\lambda}_0$ and $\boldsymbol{\lambda}_1$,

$$d(\boldsymbol{\lambda}_0, \boldsymbol{\lambda}_1) = \sqrt{\frac{1}{2}(\boldsymbol{\lambda}_1 - \boldsymbol{\lambda}_0)^T \big(\zeta(\boldsymbol{\lambda}_0) + \zeta(\boldsymbol{\lambda}_1)\big)(\boldsymbol{\lambda}_1 - \boldsymbol{\lambda}_0)}.$$

We expect the approximate linearized distance to be robust so long as the discretization is sufficiently fine.

The relation in Eq. (5.1) states that the arrival time grows inversely in proportion to the distance that must be traveled on the surface of the manifold. In the case of the Ising model, we computed the friction tensor on a finer mesh near the second order phase transition to account for this approximation. More sensitive systems could employ an adaptive mesh scheme to ensure a high degree of accuracy.

We use the fast marching method [138] to find numerical solutions to the *eikonal* equation. This algorithm approximates continuous geodesic paths, as shown in Fig. 5.1, given discrete knowledge of the distance between neighboring points [147]. A geodesic path in the parameter space travels backwards along the gradient of T. After computing the arrival time field T for geodesics initiated from some initial point $\boldsymbol{\lambda}_0$, we can solve a first order differential equation, to find a geodesic path between $\boldsymbol{\lambda}_0$ and $\boldsymbol{\lambda}_1$. The fast marching algorithm, which will not be described in detail here, is an extremely efficient numerical scheme for finding viscosity solutions to the *eikonal* equation [147]. Thus, once we have sampled the metric tensor, we can rapidly compute optimal protocols between any two points on the manifold.

The technique performs well, even for the complicated manifold shown in Fig. 4.1. The optimal protocols, as shown in Fig. 5.1, clearly avoid the critical point by curving around this feature of the phase diagram due to the high friction in this region. Passing directly through the first order phase transition, even in a finite time, also has a high dissipation cost. Overcoming the broken symmetry requires nucleation of a domain of opposite spin, which can then grow to reverse the net magnetization. Nucleation can be accelerated by applying a large field, but this results in a proportionally higher dissipation when the spins reverse.

At low temperatures, excitations are small and local, which leads to low friction, as shown in the configurations in Fig. 5.1. Because changes in the control parameters elicit little or no change in the system, the protocols are weakly constrained below the critical temperature. Similarly, in the high temperature limit, the vanishing spatial and temporal correlations result in low friction and weakly constrained protocols. Only at intermediate temperatures does higher friction impose tight constraints on the minimum dissipation paths.

Optimal protocols for reversing the magnetization are plotted in Fig. 5.1. Contrary to our expectations, the magnetic field is first applied in the direction of the spontaneous magnetization. Because the generalized friction coefficients are small in the low temperature region, aligning the spins at the outset minimizes the overall contribution to the dissipation by damping fluctuations as the temperature of the system is brought above the critical temperature. The direction of the field is then reversed, but since the value of the magnetic friction coefficient is large along the zero external field line, as shown in Fig. 4.1 (c), crossing between positive and negative field must be performed slowly. The protocol is symmetric about zero field due to the underlying symmetry of the model, hence the remainder of the protocol is just a reflection of the first half: we reduce the temperature and finally turn off the field.

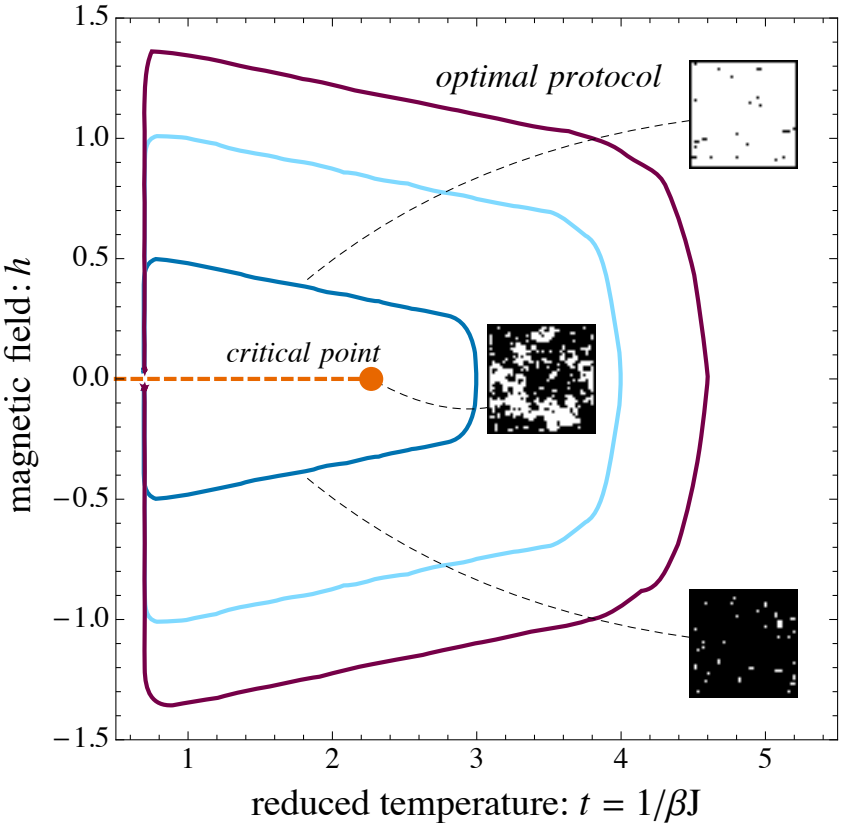

Figure 5.1: Minimum dissipation, finite-time protocols for reversing the magnetization of the two dimensional Ising model with initial and final conditions below the critical temperature, $t_C \approx 2.269$. The outermost protocol is unconstrained, whereas the inner two protocols have a constraint on the maximum temperature. We control the external field h and the spin-spin coupling constant J as a function of time. Initially, the protocols ramp up the external field followed by a temperature increase as the field is turned off. Low dissipation protocols circumscribe the critical region to avoid large spatial and temporal correlations near the second order phase transition. The first order phase transition ($h = 0,\ t < t_C$) is shown as a dashed line ending at the critical point.

5.2 On the fly approach: Geometric minimum action method

Calculating the tensor for use with the fast marching method is computationally demanding, particularly when the space of control parameters is high dimensional. We can make such problems computationally tractable by opting to evaluate the metric locally, rather than globally. We then use the local information about the metric to iteratively relax a trial protocol towards the optimum. In order to calculate optimal nonequilibrium protocols without explicit knowledge of the metric, we employ a geometric minimum action method [121, 122]. These numerical techniques build on the minimum action methods developed to study reaction paths [148]. The advantage of this approach over the fast marching method is that it can be used for arbitrarily high dimensional protocol spaces because it does not require constructing the metric over the entire protocol space.

Our goal is to construct solutions to (4.15). To do so, we follow closely the algorithm proposed in [121, 122]. We first discretize the protocol $\boldsymbol{\lambda}(t)$ on a grid $t_0 = 0, t_1, \ldots, t_k = T$, which $t_j = j\Delta t$, $j = 0, \ldots, k$, $\Delta t = t_{\text{obs}}/k$. Denoting the discretized path by $\boldsymbol{\lambda}_i = \boldsymbol{\lambda}(t_i)$, we also discretize the first and second derivatives along the path, using

$$\dot{\boldsymbol{\lambda}}(t_i) \approx \frac{\boldsymbol{\lambda}_{i+1} - \boldsymbol{\lambda}_{i-1}}{2\Delta t} \equiv D\boldsymbol{\lambda}_i, \tag{5.2}$$

$$\ddot{\boldsymbol{\lambda}}(t_i) \approx \frac{\boldsymbol{\lambda}_{i+1} + \boldsymbol{\lambda}_{i-1} - 2\boldsymbol{\lambda}_i}{\Delta t^2} \equiv D^2\boldsymbol{\lambda}_i. \tag{5.3}$$

We then update the positions of $\boldsymbol{\lambda}_i$ until they approximate the solution to (4.15) as follows: Letting $\boldsymbol{\lambda}_i^{(n)}$ be the $k+1$ positions of the control parameter in the N-dimensional space after n iterations, we get the next update by solving the following linear system of equations,

$$\begin{aligned} &\boldsymbol{\lambda}_i^{(n+1)} - \boldsymbol{\lambda}_i^{(n)} = \\ &\Delta r \left(D^2\boldsymbol{\lambda}_i^{(n+1)} + \frac{1}{2}(\zeta_i^{(n)})^{-1}\partial_{\boldsymbol{\lambda}}\zeta_i^{(n)} : D\boldsymbol{\lambda}_i^{(n)} D\boldsymbol{\lambda}_i^{(n)} \right) \\ &\qquad \text{for } i = 1, \ldots, k \end{aligned} \tag{5.4}$$

with $\boldsymbol{\lambda}_0$ and $\boldsymbol{\lambda}_k$ kept fixed and where $\zeta_i^n = \zeta(\boldsymbol{\lambda}_i^{(n)})$ and Δr is a parameter controlling the size of the update which must be kept small enough for numerical stability. We also ensure constant spacing between points along the protocol, using a reparametrization scheme [142]. This procedure proceeds iteratively until the minimum action path is reached. Note that, letting $\Delta t \to 0$ and $\Delta r \to 0$, (5.4) amounts to solving (4.15) via relaxation using

$$\partial_r \boldsymbol{\lambda} = \partial_t^2 \boldsymbol{\lambda} + \frac{1}{2}\zeta^{-1}\partial_{\boldsymbol{\lambda}}\zeta : \partial_t\boldsymbol{\lambda}\partial_t\boldsymbol{\lambda} + \mu\partial_t\boldsymbol{\lambda} \tag{5.5}$$

in which r plays the role of a relaxation time for the path and $\mu\partial_t\boldsymbol{\lambda}$ is a Lagrange multiplier term that guarantees that $|\partial_t\boldsymbol{\lambda}|$ is a constant. Eq. (5.4) treats the diffusion term $\partial_t^2\boldsymbol{\lambda}$ implicitly to avoid the Courant-Friedrichs-Lewy condition on Δr of an explicit scheme.

The implementation outlined above can easily be made computationally efficient. Very little information is shared between distinct points along the protocol. In fact, only at the final stage of an iteration, when the protocol is updated, is global information about the protocol needed. This means that the metric can be estimated for each point along the protocol in parallel, which dramatically increases performance of the algorithm. Because of the noise when estimating the metric, the algorithm will fluctuate around the minimum action path, which can be addressed by averaging over the sampled paths from the final iterations of the algorithm.

5.3 Applications to minimum dissipation bit-reversal

Modern computers dissipate a vast amount of energy as heat, greatly in excess of the minimum thermodynamic cost of logic operations for classical bits [48, 149]. Recent experiments have demonstrated that magnetic spintronics can be used to implement logic operations on thin nanomagnetic films, providing a route to extremely low dissipation computing [150–157]. However, thermodynamically ideal control cannot be realized in the laboratory, because any finite-time transformation must dissipate heat. The amount of dissipation depends on the protocol used for control: the protocol that dissipates the minimal amount of excess work to the environment is "optimal". Other objectives of nonequilibrium control, such as speed or accuracy [158, 159], could be used to define optimality in more general contexts. Moreover, there has been substantial interest recently in trade-offs among energy, speed, and accuracy [50, 160].

When we control a nanoscale, physical system and drive it away from equilibrium, the character and extent of its fluctuations depend on the history of the perturbation that we apply. Each external protocol used in an irreversible, nonequilibrium transformation has an associated energetic cost: the reversible work plus excess work that is dissipated to a thermal reservoir. At the nanoscale, the cost of control is not a deterministic quantity. Because the fluctuations in the controlled system have a scale comparable to the extent of the system itself, the dissipation fluctuates, varying from one realization of the protocol to another. The inherent noise associated with small systems adds a layer of complexity to the problem of designing protocols that favor low dissipation. Nevertheless, theoretical advances in nonequilibrium statistical mechanics [18, 21, 29] and new experimental tools [87, 88, 157] have inspired a wide range of efforts to find protocols that minimize the dissipated work and achieve efficient control of fluctuating, nanoscale systems.

Here, we compute the optimal protocol for driving a nanomagnetic bit from a state aligned with the "hard" axis to a state aligned with the "easy" access. This process is an important step of experimental bit erasure protocols [157, 161]. The bit is described as the magnetic moment of an anisotropic, nanomagnetic film and we control external fields that couple to the easy and hard axes of the underlying magnet. This model has been widely and successfully used to describe spintronic systems [152, 162, 163].

The idealization of an isolated bit neglects local, ferromagnetic interactions arising from

spin-spin coupling. Such interactions affect the response of the system to external control, so we also study low dissipation bit reversal by computing the optimal protocol to invert the net magnetization of a ferromagnetic Ising model with two energetically degenerate metastable states. In this model, the intricate spectrum of local fluctuations can be overcome by spatially controlling the external field. We control the external magnetic field over small blocks of spins, independently tuning the field strength over domains of a few interacting spins. This set up leads to a very high dimensional space of control parameters, and solving the optimization problem requires the development new computational tools.

The complex interplay between nonlinear dynamics and time-dependent external forces in the systems that we consider puts them outside the reach of analytical treatment. While there is a substantial literature on minimum dissipation control, previous theoretical work on optimal protocols has largely focused on exactly solvable systems [118, 119, 126, 129]. The limited set of systems that can be formally analyzed has inspired recent efforts to compute low-dissipation protocols using numerical techniques [164–167].

The development of numerical strategies to determine optimal protocols has, in part, relied on a geometric interpretation of minimum dissipation protocols. In the linear response limit, an optimal protocol can be characterized as a geodesic on a Riemannian manifold [114, 117, 118, 168]. The equilibrium fluctuations and time correlations at different values of the external control parameters determine a metric tensor, which defines a generalized length proportional to the amount of excess work done along the protocol. If the control parameter space is sufficiently small, this metric can be sampled exhaustively at a discrete set of control parameter values.

As further discussed in Section 5.6, Ref. [165] uses path sampling techniques to harvest nonequilibrium protocols in proportion to their average dissipation. Trajectory space Monte Carlo techniques have also been developed for use in stochastic optimal control theory to iteratively refine importance sampling distributions [166, 167, 169], exploiting the connection between importance sampling and optimal control [170, 171]. With a bias that favors low dissipation, Gingrich et al. [165] explore an ensemble of low dissipation protocols and show that there is a large number of protocols with a dissipation near the minimum achievable value. For a high dimensional protocol, exploring fluctuations in the protocol space remains a significant computational challenge.

Here, we demonstrate that the geometric structure of the protocol space enables the use of a geometric minimum action method [121, 122] to identify the optimal nonequilibrium driving. The geometric minimum action method produces an equation of motion for the protocol, namely Eq. (5.5). Because we update a quasi-one-dimensional "string", the amount of computation need to relax the protocol does not grow exponentially in the number of dimensions. The method relies on only local, equilibrium sampling, meaning that we can productively compute optimal protocols even in high dimensional control spaces.

The goal of reducing the excess work done on the system has many applications outside of low-dissipation computing. Nanoscale engine optimization is one such direction: recent experiments have implemented fluctuating, microscopic variants of the Carnot [100] and Stirling [59] cycles. In these fluctuating engines, the excess power is dissipated to the heat

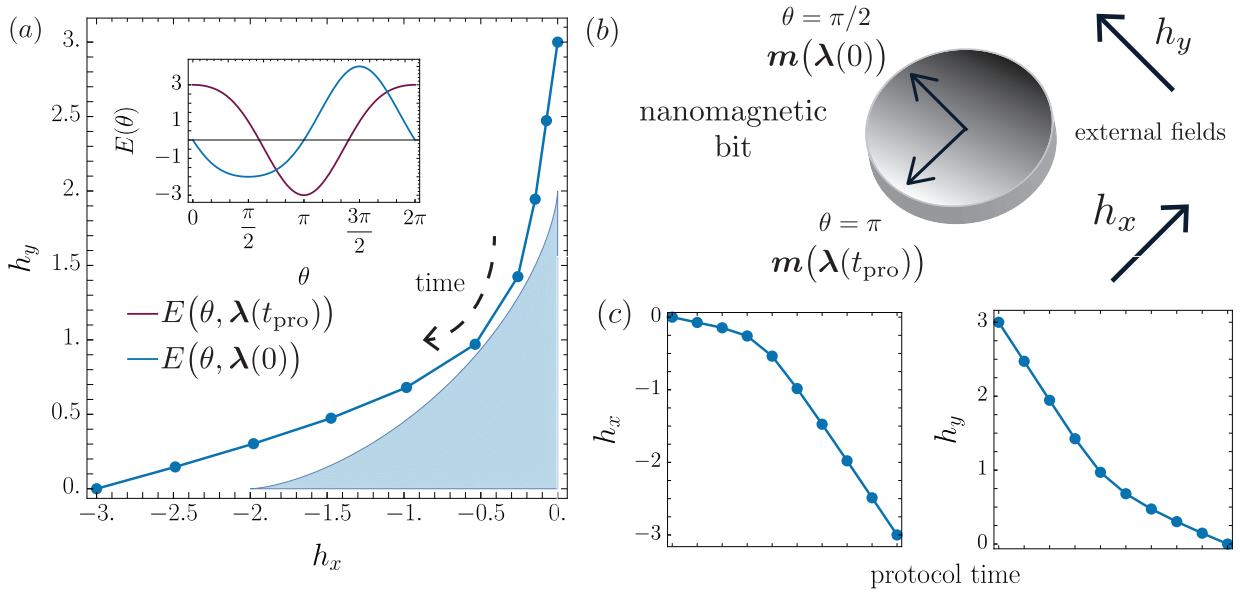

Figure 5.2: Optimal control of the magnetic moment of a thin nanomagnetic film using orthogonal fields h_x and h_y, as described in Sec. 5.4. (*a*) The optimal protocol as determined by the geometric minimum action method. *Inset*: The potential energy of the system at the beginning and end of the protocol. (*b*) A schematic of the control problem: a thin magnetic film is controlled by external fields h_x and h_y. (*c*) The x and y fields as a function of protocol time. Note the significant deviation from the linear ramps commonly used in experiments.

bath, rather than being converted to work. As a result, minimum dissipation protocols maximize the engine's thermodynamic efficiency at finite power. Also, in nonequilibrium experiments that determine free energy differences via the Jarzynski equality [29], minimum dissipation protocols determine the free energy difference with the highest possible accuracy for a fixed, finite number of samples [131, 172].

5.4 Optimal bit control in a thin magnetic film

We represent a bit as the magnetic moment $\boldsymbol{m}$ of a nanoscopic metal film. At this scale, thermal noise leads to spontaneous changes in magnetic moment. The fluctuating magnetic moment satisfies the stochastic Landau-Lifshitz-Gilbert equation,

$$\dot{\boldsymbol{m}} = \boldsymbol{m} \times (\boldsymbol{h}_{\rm ext} + \boldsymbol{h}_{\rm T}) - \alpha \boldsymbol{m} \times \big(\boldsymbol{m} \times (\boldsymbol{h}_{\rm ext} + \boldsymbol{h}_{\rm T})\big), \tag{5.6}$$

where the field $\boldsymbol{h}_{\rm T}$ is random thermal field, α is the Gilbert damping parameter, and $\boldsymbol{h}_{\rm ext}$ is the external field [173]. In the case that the magnet is a thin film, $\boldsymbol{m}$ is confined to the xy-plane and we assume that the magnitude is conserved. The equation of motion for the

angular direction of the moment, θ, is given by a Langevin equation [174],

$$\dot{\theta} = -\alpha E'(\theta) + \sqrt{2\alpha\beta^{-1}}\,\eta(t), \tag{5.7}$$

$$E(\theta) = \beta_2 \sin^2(\theta) - h_x \cos(\theta) - h_y \sin(\theta). \tag{5.8}$$

The noise η has mean zero and is δ-correlated in time. Throughout, we set the anisotropy parameter $\beta_2 = 1$, the Gilbert damping coefficient to $\alpha = 10^{-2}$, and the inverse temperature $\beta = 1$. The value of α is a realistic choice for the materials commonly used in spintronics experiments [152].

We computed the optimal protocol for driving the system from a state in which the magnetic moment is aligned along the easy axis ($\theta = \pi/2$) to a state aligned with the hard axis ($\theta = \pi$). Driving the magnetic moment to the hard axis from the easy axis is the final step in experimental protocols for bit erasure as implemented on thin nanomagnetic films [157]. We took as an initial protocol a line from $(h_x, h_y) = (0, 3)$ to $(h_x, h_y) = (-3, 0)$ discretized into ten equally spaced steps. Using a time step $dt = 10^{-4}$ for the dynamics of the magnetic moment, we estimated the thermodynamic metric and its derivatives with 10000 simulation steps at each point along the protocol. We propagated the protocol according to Eq. (5.4) with a time step of 10^{-6}. The system converged in under 1000 iterations and we ran a total of 10000 iterations to ensure that the protocol was fully relaxed.

The optimal protocol for driving the transition is shown in Fig. 5.2 (a). The blue region in the figure shows the portion of parameter space where there are two minima in the potential. We used boundary conditions outside the region of metastability to ensure unique initial and final equilibrium states. The optimal protocol deviates in a nontrivial way from the protocols used in experiments, in which each field is a linear function of time [152, 157]. Initially, the field in the y-direction is decreased while the field h_x remains small. As the y-field is decreased, the minimum in the potential energy, as shown in the inset of Fig. 5.2, shifts towards the final state at $\theta = \pi$. The orthogonal field increases the curvature around the potential energy minimum as it shifts towards the negative x-axis.

Interestingly, the protocol plotted in Fig. 5.2 has a shape similar to the boundary of the "astroid" regions described in Ref. [174]. In Fig. 5.2, the blue region encloses a parameter regime in which there is low probability of spontaneous bit reversal, i.e., when there is metastability in the potential. Because the protocol avoids this region, at any fixed point along the protocol there is a unique equilibrium state.

5.5 Control of the two-dimensional Ising model with spatially varying field

Magnetic bits are stable on long timescales, due to the large energetic barrier separating the $+1$ and -1 states. The inherent stability of nanomagnets is one of the primary advantages of magnetic spintronics from the engineering perspective because no energy is required to maintain the state of the bit [157]. A nonequilibrium protocol for bit reversal must drive the

system over the large energetic barrier separating two states to switch between the ground states. A naive protocol for this operation will be extraordinarily dissipative [164], but more sophisticated control strategies such as local heating and spatial control may lower the thermodynamic cost of bit reversal in practice [134].

We investigate protocols where the external field is spatially controlled. We consider a ferromagnetic, two dimensional Ising model below the critical temperature, so that the probability of a spontaneous bit reversal is low. We take as our control parameters N independent external magnetic field strengths, $\{h_i\}_{i=1}^{N}$, which couple to non-overlapping blocks of spins as shown in Fig. 5.3 (a). We prepare the system with a fixed boundary condition that creates two metastable states. On the left and right sides, the boundary consists of all up spins. On the top and bottom, the boundary consists of all down spins. We then seek a protocol that drives the system from a configuration where the spin up metastable state is favored ($h_i = 0.05$, for all $i = 1, \ldots, N$) to a region of the protocol space where the spin down configuration is favored, ($h_i = -0.05$, for all $i = 1, \ldots, N$).

In our calculation, the protocol is discretized into sixteen equally spaced points. We initialize the system with a protocol that linearly interpolates the magnetic field between -0.05 and 0.05, so the initial protocol is spatially uniform. The calculations were performed for 40×40 and 100×100 Ising models, controlling 4×4 and 10×10 block magnetic fields, respectively. We carried out the geometric minimum action method with a time step of 10^{-4}. At each iteration, 10000 sweeps of Monte Carlo dynamics with a Glauber acceptance criterion were used to estimate the metric tensor and its derivatives at each point along the discretized protocol. These protocols converge to their final form in roughly 1000 iterations, but we continued to sample for 10000 total iterations. There is no significant dependence on system size.

The optimized protocol for inverting the magnetization is shown in Fig. 5.3 (b). The values of the external field are shown on a gray scale, with spin blocks drawn according to their location. On the top, we show the six snapshots of the spin system near the transition between the metastable states. These configurations of the system are representative of the states seen along the optimized protocol.

The hourglass shapes seen in Fig. 5.3 (b) are characteristic of the spontaneous transition pathways between the metastable states of this model [175]. First, the field reverses along the left and right boundaries of the system. The work associated with flipping these spins is minimal due to the layer of up spins from fixed boundary condition. The protocol proceeds to reverse the magnetization by continuing to grow those domains from the boundary until the bulk domain of up spins can be stabilized. The minimum dissipation protocol drives the magnet from the negative metastable state to the positive metastable state by flipping spins at the boundaries.

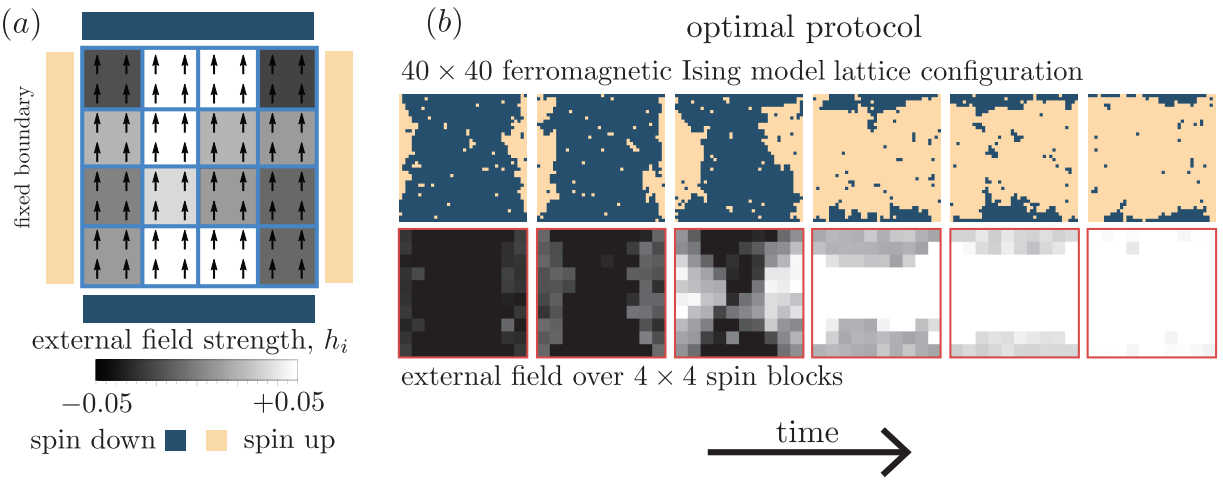

Figure 5.3: Optimal control of the two dimensional Ising model with spatial control of the magnetic field. (a) A schematic of the set-up of the system. Boundary conditions are fixed so that there are two distinct metastable states. The external field is tuned independently for the different spin blocks. (b) An optimized protocol is shown (bottom) driving the transition. Representative structures from the configuration space are shown (top).

5.6 Monte Carlo approach: Protocol sampling

The computational approaches outlined in Sections 5.1 and 5.2 target a single optimal protocol. Identifying the exact optimum, however, may be much more costly than finding a non-optimal protocol that still has a low average dissipation. In many applications, like free energy calculations using Jarzynski's nonequilibrium work relation, the additional gains in performance reaped by using the optimal protocol can be outweighed by the effort required to find the optimum. The advantage of settling for a near-optimal protocol is most pronounced when a plethora of protocols can be found near the optimum.

In Gingrich, Rotskoff, Crooks, and Geissler [165], we introduced an approach for studying the ensemble of low dissipation protocols. The methodology employs Metropolis Monte Carlo dynamics to sample the joint space of protocols and trajectories with a bias towards low average dissipation. Roughly, the algorithm treats the protocol as a polymer subject to a Monte Carlo dynamics. For each protocol visited, the average dissipation associated with that protocol is estimated from a finite sampled mean—this aspect of the procedure relies on path sampling move sets to ensure tolerable acceptance rates [176]. In our formulation, we use a sampling procedure that allows an exact exploration of the following distribution in the linear response limit,

$$P[\Lambda(t)] \propto \exp\left(-\gamma \langle \omega \rangle_\Lambda\right), \tag{5.9}$$

where ω is the dissipation and γ is a statistical bias that favors protocols with low average dissipation. Directly sampling Eq. (5.9) is difficult, because determining $\langle \omega \rangle_\Lambda$ requires an

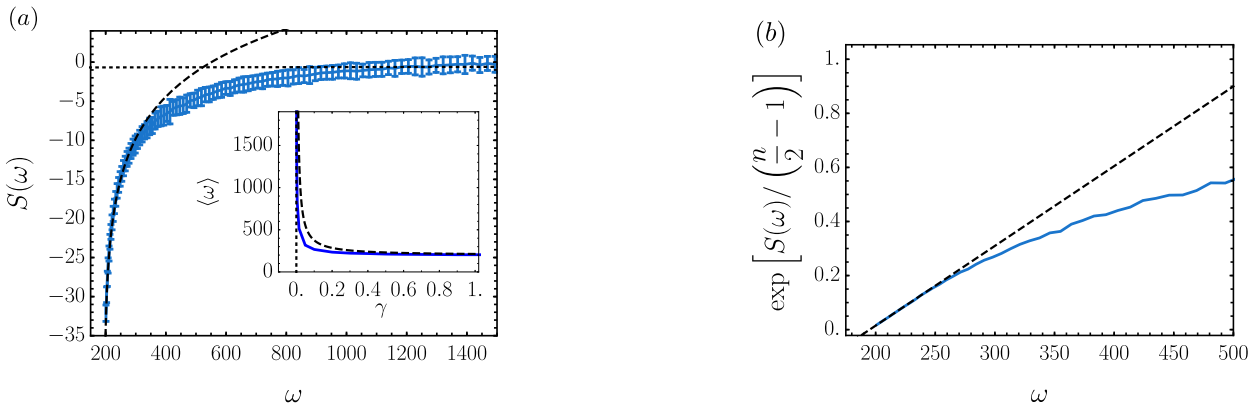

Figure 5.4: (a) The protocol entropy is plotted as a function of average dissipation for the Ising spin inversion process described in Chapter 4. (b) The scaling of the entropy in the large γ limit is consistent with Eq. (5.14).

average over many trajectories for each distinct protcol $\Lambda(t)$ visited in the course of the Monte Carlo dynamics. Our approach allows an exact reweighting of each protocol by instead sampling the joint space of protocols and trajectories. This reweighting uses a correspondence between the mean and variance of a Gaussian dissipation distribution, which is a consequence of the Fluctuation Theorem. The details underlying this method will not be discussed here, but we will instead focus on the generic structure of the low-dissipation protocol ensemble.

Protocol Entropy

The statistical bias γ in Eq. (5.9) can be tuned to target near-optimal protocols. Unfortunately, the fluctuation theorem places an intrinsic limitation on the strength of the bias and large values of γ require that large numbers of trajectories are sampled for each protocol [165]. Nevertheless, we can argue quite generally that we sample protocols with an average dissipation that near the optimal value, even when using a weak statistical bias γ.

We can anticipate the existence of a large class of low dissipation protocols by returning to the thermodynamic metric discussed in Chapter 4. While the metric is sharply peaked near the second order phase transition of the model, in both the high and low temperature regimes correlations decay rapidly (cf. Fig. 4.1). This indicates that, at least within the linear response regime, little dissipation is incurred even when the control parameters are changed by large amounts. The unconstrained path through protocol space in the low temperature regime suggests that many distinct protocols will have comparable dissipation to the optimum.

The density of low dissipation protocols can be analyzed by computing a protocol entropy, which, roughly, measures the population of protocols as a function of average dissipation.

We define the protocol entropy as,

$$S(\omega) = \ln \left[\Omega_0 \int \mathcal{D}[\Lambda(t)] \; \delta(\omega - \langle \omega \rangle_\Lambda) \right], \tag{5.10}$$

the integral is taken over all protocols $\Lambda(t)$ and the parameter Ω_0 is used to set the arbitrary zero of the function. Just as in the microcanonical partition function, Eq. 5.10 gives the volume of protocols that have average dissipation ω. Under very general assumptions, we argue that the protocol entropy grows rapidly near the minimum accessible average dissipation and then crosses over to a regime of slower growth. This generic structure of the protocol entropy suggests that a modestly biased random walk through the space of protocols will target near-optimal average dissipations.

We let $\Lambda^{\rm opt}(t) : \{t_i\}_0^K \to \mathbb{R}^n$ denote the minimum dissipation protocol for a transformation, where t_K is the duration of the protocol and n is the number of control parameters. We assume that the protocol is specified by a discrete set of values $\Lambda^{\rm opt}(t_i)$ which are linearly interpolated to give a piecewise continuous function. The initial configuration is drawn from a time-invariant distribution determined by the initial value of the control parameters $\Lambda^{\rm opt}(0)$. Defined as such, we can view the protocol as a polymer of length K in n-dimensional space.

We can estimate the form of the protocol entropy with an estimate of the partition function for protocols in the strong biasing limit. In the limit of large values of the bias parameter γ, the difference between a sampled protocol $\Lambda(t)$ and the optimal protocol $\Lambda^*(t)$ should be small. If the average dissipation $\langle \omega \rangle_\Lambda$ is a smooth functional of the protocol, we can approximate the deviation from the minimum average dissipation in terms of the small protocol variations $\delta\Lambda = \Lambda - \Lambda^*$,

$$\langle \omega \rangle_\Lambda = \omega^* + \frac{1}{2} \sum_{i,j=1} \delta\Lambda_i \cdot \mathbf{K}_{ij} \cdot \delta\Lambda_j, \tag{5.11}$$

where the indices i and j label times at which the protocol is manipulated. In this limit, we can explicitly calculate the moment generating function for the average dissipation,

$$Z(\gamma) = \left\langle e^{-\gamma \langle \omega \rangle_\Lambda} \right\rangle \underset{\gamma \to \infty}{\sim} \frac{1}{\sqrt{\det(\gamma \mathbf{K})}} e^{-\gamma \omega^*} \propto \gamma^{-n/2} e^{-\gamma \omega^*}. \tag{5.12}$$

and n denotes the total number of degrees of freedom in the protocol.

The Gaussian fluctuations around the minimum dissipation protocol make this partition function tractable. From Eq. (5.12), the unbiased distribution of average dissipations $P(\omega) \propto \exp\big[S(\omega)\big]$ is an inverse Laplace transform of $Z(\gamma)$,

$$P(\omega) \propto (\omega - \omega^*)^{n/2-1}, \tag{5.13}$$

and hence the entropy in the large γ limit is,

$$S(\omega) = \text{const} + \Big(\frac{n}{2} - 1\Big) \ln(\omega - \omega^*). \tag{5.14}$$

Fig. 5.4 compares the asymptotic expression (5.14) with the entropy computed for the two dimensional Ising spin inversion protocols.

From the asymptotic expression for $Z(\gamma)$, we can also compute the average dissipation associated with protocols sampled under the bias γ:

$$\langle \omega \rangle = -\frac{d}{d\gamma} \ln Z(\gamma) \tag{5.15}$$

$$= \omega^* + \frac{n}{2\gamma}. \tag{5.16}$$

In the absence of protocol bias, $\gamma = 0$, typical values of dissipation are quite large. Provided that $\Lambda(t)$ is bounded, the mean dissipation $\langle \omega \rangle_0$ at $\gamma = 0$ should nonetheless be finite, as is the corresponding variance $\langle (\omega - \langle \omega \rangle_0)^2 \rangle_0$. Sufficiently close to the maximum of $P(\omega)$ we therefore have

$$S(\omega) = S(\langle \omega \rangle_0) - \frac{(\omega - \langle \omega \rangle_0)^2}{2\langle (\omega - \langle \omega \rangle_0)^2 \rangle_0} \tag{5.17}$$

The parameters $\langle \omega \rangle_0$ and $\langle (\omega - \langle \omega \rangle_0)^2 \rangle_0$ in this expression are determined numerically by computing average dissipation for protocols generated in the $\gamma = 0$ ensemble. The corresponding curve is plotted in Fig. 5.4 as a dotted line. The moment generating function for the protocols is

$$Z(\gamma) \propto \exp\left[-\gamma \langle \omega \rangle_0 + \frac{1}{2} \langle (\omega - \langle \omega \rangle_0)^2 \rangle_0 \gamma^2 \right], \tag{5.18}$$

giving a γ-biased average dissipation:

$$\langle \omega \rangle = \langle \omega \rangle_0 - \gamma \langle (\omega - \langle \omega \rangle_0)^2 \rangle_0, \tag{5.19}$$

which is shown as a dotted line in the inset of Fig. 5.4.

The protocol entropy directly quantifies the near-optimal protocols that can be used for nonequilibrium control. The arguments above suggest that the structural features we see for the Ising spin inversion protocol entropy are generic. This observation relieves some of the pressure to find an exact optimum. As an illustration of the breadth of near-optimal protocols, we consider those protocols with an average dissipation less than a standard deviation away from the optimal protocol, where the standard deviation measures the fluctuations in dissipation around the optimal protocol. In the linear response regime, we can perform this calculation exactly because the variance of the dissipation is related to the mean as $\langle (\delta\omega)^2 \rangle = 2\langle \omega \rangle$. The optimal protocol has an average near 200, as can be determined from the asymptotic value of the protocol entropy function shown in Fig. 5.4. Within a standard deviation the number of accessible protocols has increased by a factor of roughly e^{10}, a dramatic enhancement.

5.7 Connections between minimum dissipation protocols and spontaneous transition paths

The results throughout this chapter illustrate an appealing potential connection between optimal near-equilibrium control and the intrinsic relaxation of an equilibrium system. Because the dissipation is a measure of irreversibility, it seems natural that an optimal nonequilibrium protocol would drive the system along a spontaneous transition path. There is an extensive theoretical literature in chemical physics dedicated to reaction paths, and many distinct formal tools exist to describe them [142, 144, 177]. Minimum free energy paths are maximum likelihood transition pathways, typically used to describe spontaneous transitions between two metastable states of a physical system in thermal equilibrium [142]. A minimum free energy path minimizes the total variation in the free energy as function of some collective variables $\alpha \in \mathbb{R}^d$. In particular, the minimum $\alpha^*(s)$, $s \in [0,1]$ is a minimizer of the following functional,

$$\mathcal{I}[\alpha(s)] = \int_0^1 |\nabla F(\alpha(s))| \cdot |\alpha'(s)| \, ds. \tag{5.20}$$

Here, we consider nonequilibrium protocols that minimize the average dissipation during some finite time irreversible transformation. Empirically, these optimal protocols give rise to a dynamics of the controlled system that resembles the spontaneous dynamics. An optimal nonequilibrium protocol $\boldsymbol{\lambda}^*$ minimizes the functional introduced in Chapter 4,

$$\mathcal{L}[\boldsymbol{\lambda}(t)] = \int_0^1 \sqrt{\dot{\boldsymbol{\lambda}} \cdot \zeta(\boldsymbol{\lambda}(t)) \cdot \dot{\boldsymbol{\lambda}}} \, dt, \tag{5.21}$$

where,

$$\zeta(\boldsymbol{\lambda}) = \beta \int_0^\infty \langle \delta X^T(\tau) \delta X(0) \rangle_{\boldsymbol{\lambda}} \, d\tau, \tag{5.22}$$

and $X_i = -\frac{\partial \beta V}{\partial \boldsymbol{\lambda}_i}$ is the force conjugate to the control parameter $\boldsymbol{\lambda}_i$.

Thus, one might anticipate that if $\boldsymbol{\lambda}^*$ is an optimal protocol in the sense of Eq. (5.21) then the path $\overline{\alpha}(t) := \langle X \rangle_{\boldsymbol{\lambda}(t)}$ is a minimum free energy path in the sense of Eq. (5.20). We have defined $\overline{\alpha}$ as the minimum action path of the dynamics subject to the optimal protocol. As an attempt to verify this claim, we first compute the derivative of the path with respect to the parametrization,

$$\dot{\overline{\alpha}} = \frac{d}{dt} \langle X \rangle_{\boldsymbol{\lambda}(t)} \tag{5.23}$$

$$= \frac{d}{dt} \int_0^t ds \, \frac{d}{ds} \langle X \rangle_{\boldsymbol{\lambda}(s)} \tag{5.24}$$

$$= \int_0^t \langle \delta X^T(0) \delta X(s) \rangle_{\boldsymbol{\lambda}(s)} \dot{\boldsymbol{\lambda}}. \tag{5.25}$$

The linear response approximation typically integrates this expression with an upper bound of $t \to \infty$, assuming that system relaxes faster than the control parameters change. Next,

we need to calculate the gradient of the free energy along this path. This term is more subtle: we have assumed that the position along the path is given by the equilibrium average position with the current control parameters. This means that $\overline{\alpha}(t) = \langle X \rangle_{\boldsymbol{\lambda}(t)}$ is a free energy minimum and the gradient should vanish. However, during the nonequilibrium driving protocol, we expect that the system will respond to an instantaneous change in control parameters. We thus postulate that we have a correction to the gradient of the free energy that arises from this lag, an approximation reminiscent of the endoreversibility arguments in original approach to thermodynamic geometry [47]. In particular, we evaluate the gradient at $\langle X \rangle_{\boldsymbol{\lambda}(t)}$ in the free energy at time $t + \Delta t$, immediately after changing the control parameters. Let us write out explicitly the expression in terms of the relevant free energy,

$$\nabla F(\langle X \rangle_{\boldsymbol{\lambda}(t)}) = -\beta^{-1} \nabla_X \ln \int_{\mathbb{R}^d} dx \; e^{-\beta V(x, \boldsymbol{\lambda}(t+\Delta t))} \tag{5.26}$$

$$= -\beta^{-1} \nabla_X \ln \int_{\mathbb{R}^d} dx \; e^{-\beta V(x, \boldsymbol{\lambda}(t))} e^{-\beta \dot{\boldsymbol{\lambda}} \Delta t \cdot X} \tag{5.27}$$

$$= \Delta t \; \dot{\boldsymbol{\lambda}}. \tag{5.28}$$

By the argument above, the gradient of the part of the free energy that is associated with the potential $V(x, \boldsymbol{\lambda}(t))$ vanishes when evaluated at $\langle X \rangle_{\boldsymbol{\lambda}(t)}$. One way of seeing this explicitly is to approximate the free energy with a saddle point estimate.

We now combine the terms to write the objective function in terms of this path $\overline{\alpha}(t)$,

$$\mathcal{I}[\overline{\alpha}(t)] = \int_0^1 |\Delta t \; \dot{\boldsymbol{\lambda}}| \cdot |\langle \delta X^T \delta X \rangle_{\boldsymbol{\lambda}(t)} \dot{\boldsymbol{\lambda}}| \; dt. \tag{5.29}$$

A few comments about the terms involved: Δt is a positive constant, and thus can be pulled out of the integral; the tensor is a positive semi-definite one, so we can write,

$$= \Delta t \int_0^1 \dot{\boldsymbol{\lambda}}^T \left(\int_0^\infty \langle \delta X^T(0) \delta X(s) \rangle_{\boldsymbol{\lambda}(t)} ds \right) \dot{\boldsymbol{\lambda}} \; dt. \tag{5.30}$$

We recognize the expression above as, up to a constant, the energy associated with the length functional Eq. (5.21). It should be emphasized that this calculation is heuristic, at best. The precise connection between spontaneous reaction paths and optimal protocols will demand further exploration and more sophisticated arguments.

Establishing a connection would be timely, because determining nonequilibrium driving protocols that minimize dissipation for nanoscale systems has become a significant goal in both the molecular sciences and engineering. Increasing the number of available control parameters leads to more elegant and efficient strategies for control, but the resulting increase in complexity demands new computational tools. Under very general assumptions, the argument given in Sec. 4.3 proves that the notion of thermodynamic geometry emerges only from a time scale separation between the dynamics of the controlled system and the experimental parameters. This derivation encompasses the linear response arguments in Ref. [118] but further elucidates the physical origins of the thermodynamic metric.

The geometry of nonequilibrium control allows us to derive a general, robust numerical method to compute optimal protocols. In analogy to Lagrangian mechanics, the thermodynamic length can be thought of as an action functional in the space of protocols. Optimizing for minimum dissipation is equivalent to minimizing this action. The geometric minimum action method we describe can be used to compute optimal protocols in previously inaccessible, high-dimensional systems.

We applied these general tools to control problems motivated by recent spintronics experiments using nanomagnetic bits. Our calculations reveal protocols that deviate dramatically from those protocols commonly used in experimental settings, suggesting simple strategies for pushing computing closer to the low-power limit.

The nontrivial protocol for changing the orientation of bit described in Sec. 5.4 has an evocative structure. The form of the optimal protocol mimics the astroid shape of the boundary in parameter space between the metastable regime and the stable regime. At this boundary, spontaneous transitions between the initial and final configurations become possible, perhaps indicating that the system is being driven through a set of states followed by an unperturbed transition.

In the interacting example of bit reversal with spatial control over the external fields, Sec. 5.5, the optimal protocol appears to drive the system along a nucleation pathway. This optimal protocol has a striking similarity to spontaneous reaction paths in the absence of nonequilibrium driving (cf., Ref. [175]). Empirically, the optimal protocol appears to drive the system along a minimum free energy path, which is the most likely spontaneous reaction path [142], consistent with the heuristic argument in Section 5.7.

Chapter 6

Nonequilibrium control in self-assembly

When interactions among components can be precisely tuned, it is possible to design materials that spontaneously assemble into microscopic structures [178–182]. These autonomous assembly processes often yield only a fraction of the desired structure. Metastable intermediates can decimate the yield which has led, in some cases, to restricted design spaces or more complex components [183, 184]. For example, DNA origami reduces the prevalence of metastable structures by taking advantage of the large sequence space to ensure that few favorable non-specific interactions are included [179, 185]. Synthesizing molecular components under stringent constraints or with highly specific interactions, however, also restricts the materials we can use and the structures we can productively make.

Strategies that achieve robust assembly without such restrictions are of great practical interest, due to the unique physical properties of microscopic metamaterials [186, 187]. However, kinetic trapping can undermine the goal of high yields when there are strong interactions among the components [188]. When such intermediates are prevalent, the approach to equilibrium is slow, which suggests we may be able to accelerate self-assembly at the nanoscale. Can long-lived, undesired structures be eliminated—or at least significantly mitigated—through the use of nonequilibrium perturbations that accelerate relaxation towards the target structure?

We consider this question through the lens of a colloidal self-assembly process. In particular, we focus on the formation of clusters with a target stoichiometry; given attractive particles of type A and B, we ask how to maximize the yield of AB_n clusters where n is the maximum number of B particles sterically accommodated by the central A particle. The problem of assembling clusters with a high yield of desired structures has previously been studied by varying the size ratio of the constituent colloidal particles [184]. However, the previous work on this issue did not directly consider the kinetics of the assembly process, and doing so offers an alternative route to enhanced yield.

Because experiments typically work with charged colloidal particles in solutions with Debye lengths much smaller than the micron scale particles, the interactions between particles

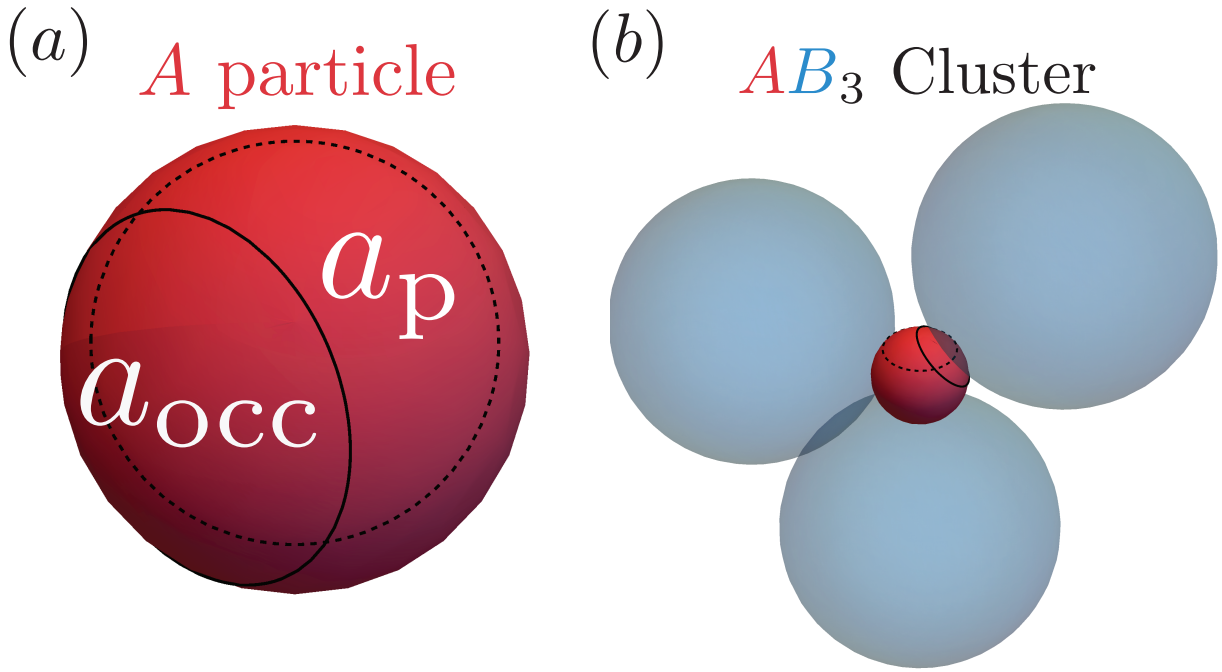

Figure 6.1: (a) In order to bind an additional colloidal particle, a probe area a_{p}, bounded by the dashed line, must be unoccupied. The effective area of occupancy, a_{occ} is determined by the effective hard sphere radius of the B particle and is shown bounded by the solid black line. The probability of density fluctuations in the area a_{p} determines, in part, the probability of each additional binding event. (b) A configuration of an AB_3 cluster with the diameter ratio $r_B/r_A = 3$, which accommodates a maximum stoichiometry of $n = 4$. As in (a), the boundary of probe area is drawn as a dashed line and the boundary of the occupied area overlapping the probe area is drawn as a solid black line.

are extremely short-ranged [184]. What is more, the strength of the surface charges and size of the binding area are such that the particles bind essentially irreversibly. Under these conditions, each binding event $AB_i \rightarrow AB_{i+1}$ is dictated by the probability that a B particle encounters the binding surface of an A particle. Fig. 6.1 illustrates the necessary conditions for binding: a probe area a_{p} must be vacant with sufficient probability that a B particle will encounter an accessible binding surface on the timescale of an experiment. Under equilibrium conditions, if the B particles are "tightly packed" on the surface of the A particle, the relaxation towards the maximally bound AB_n structure becomes extremely slow relative to the timescale of typical particle motions. Though it is difficult to achieve experimentally, if the sizes of the particles can be modulated, the yield of AB_n clusters can be improved by preparing the B particles so that they are as small as possible without allowing an $n+1$st particle to bind [184]. This design principle ensures that the necessary binding surface is available with maximal probability.

Here, we offer a distinct, more easily and sensitively tuned control parameter to improve the yield of the desired assembly. We show that the accessible binding area a_{p} can be exposed with high probability simply by applying an external field that drives the charged colloids in opposite directions. This, together with the streaming motion of the driven particles, leads to a dramatic enhancement of the number of binding events per unit time. Our design principle can be rationalized with a kinetic model of the assembly process based on the density fluctuations within the binding area a_{p}. We verify our results with numerical simulations of the microscopic particle dynamics over a wide range of B/A diameter ratios and electric field strengths.

The increase in yield comes at a dissipative cost. We systematically explore the relation between yield and the dissipated work in both our kinetic model and using particle-based simulations. We suggest a more generic trade-off between the speed of relaxation towards equilibrium and dissipation. Further, we demonstrate the existence of an optimal driving force.

6.1 Kinetic model of equilibrium yield

To explore the relationship between yield and control, we first describe an equilibrium process in which oppositely charged colloidal particles, types A and B, assemble into clusters. At equilibrium, the relationship between the particle diameter ratio $\alpha = D_B/D_A$ and the fraction of maximally bound clusters, i.e., the yield, has been approximated in terms of the free binding area of a central particle [184]. The analysis of Ref [184] does not explicitly model the kinetics, but reveals the importance of the binding dynamics. First, we note that there is a range of diameter ratios that can accommodate a maximum of n bound B particles without the B particles sterically occluding one another. Denote this interval by $[\alpha_n^-, \alpha_n^+]$. As α approaches α_n^+, the packing of B particles around the central A particle becomes tightly constrained. A random encounter from a B particle in the bulk is unlikely to find the binding surface of the A particle, because it requires that the bound particles to occupy a very rare configuration. On the other hand, as the diameter ratio approaches the lower bound α_n^-, the amount of available binding surface increases, nearly to the point that an $n+1$st particle could fit without steric hindrance. The relative rate of binding the nth B particle when the diameter ratio is α_n^- greatly exceeds the rate at $\alpha = \alpha_n^+$.

The kinetics of cluster formation can be modeled explicitly to give a more complete theoretical description of the expected yield. Because the A-B interactions are strong, the equilibrium state of the colloidal system is expected to be entirely populated by maximally bound clusters. The model used to study in Schade et al. [184] focuses on the available area for subsequent binding events, but does not give a physical description of the kinetics of the assembly process. Moreover, their model does not include relaxation of the partially formed clusters.

We model the $A - B$ binding process as a reaction diffusion process. This approach decomposes the process into two contributions to the rate of binding: the fluctuations of

the bound B particles, which determine the probability of finding free binding surface, and the rate of new arrivals. The change in density of the population of AB_i clusters can be expressed solely in terms of the densities $\rho_{AB_{i-1}}$ and ρ_B and a bimolecular rate constant $k_b^{(i)}$,

$$\partial_t \rho_{AB_i} = k_b^{(i)} \rho_{AB_{i-1}} \rho_B. \tag{6.1}$$

Computing $k_b^{(i)}$ involves analyzing the molecular scale fluctuations that give rise to a successful binding event and the diffusive dynamics of the particles themselves.

We first consider the probability that a B particle arrives at a partially formed cluster, assuming that once it has entered the "reaction volume", it joins the cluster with a rate k^*. We assume that the probability that the distance between a B particle and an AB_{i-1} cluster is r, denoted by $c(r)$, relaxes to a steady state, which is constant inside the reaction volume. Outside this region, the probability $c(r)$ satisfies a Smoluchowski equation,

$$D\nabla^2 c(r) = 0, \tag{6.2}$$

with the boundary condition,

$$\lim_{r\to\infty} c(r) = \rho_{AB_{i-1}} \rho_B v_0, \tag{6.3}$$

where v_0 is a constant with units of volume, which we subsequently set to unity. Once a B particle is sufficiently close to the AB_{i-1} cluster, it will bind with a rate k^*. The length scale at which these binding kinetics dominate, which we denote R, defines the microscopic reaction volume for the process. In the steady state, there is a current through the surface of the microscopic reaction volume,

$$J(r) = 4\pi r^2 D \partial_r c(r), \tag{6.4}$$

which can be related to the rate of attachment,

$$J(R) = \frac{4}{3}\pi R^3 k^* c(R). \tag{6.5}$$

Setting $J(r) = J(R)$ yields a first order differential equation which can be solved analytically,

$$c(r) = \rho_{AB_{i-1}} \rho_B - \frac{R^2 k^* p(R)}{Dr}. \tag{6.6}$$

To derive an expression for the reaction rate, we first evaluate $c(r)$ at the boundary of the microscopic volume,

$$c(R) = \frac{\rho_{AB_{i-1}} \rho_B}{1 + \frac{k^* R}{D}}. \tag{6.7}$$

Using this expression in Eq. (6.5) provides an explicit formula for the reaction rate as function of k^*,

$$k_b^{(i)} = \frac{4\pi D R}{1 + \frac{3D}{k^* R^2}}. \tag{6.8}$$

Note that this expression recovers the usual diffusion limited reaction rate in the case that $k^*R \gg D$, i.e., when the reaction is extremely effective.

The microscopic reaction rate k^* is sensitive to the geometry and dynamics of the AB_{i-1} cluster. A robust estimate of k^* can be derived using Kramers' theory [189],

$$k^* = D\left[\int_{x^*}^{R} dx\ p(x) \int_{r_A}^{x} dx'\ \frac{1}{p(x')}\right]^{-1}. \tag{6.9}$$

The Kramers' expression for the rate assumes that the interparticle distance is an appropriate reaction coordinate for the process, i.e., that it captures the slowest degrees of freedom [136, 190, 191]. Evaluating the expression for k^* requires detailed information about the intermolecular interactions during a binding event.

We computed the probability that there are no particles in an area a, given that $n-1$ particles are bound. Note that n, the maximum number of particles that can bind, varies with α. Directly evaluating this probability is challenging due to the correlations among the $n-1$ particles. To estimate the probability at each value of α and $i \geq 2$, which we denote $P_a(i-1)$, we performed hard disk Monte Carlo simulations. In our Monte Carlo dynamics, each of the $i-1$ B particles is represented as a hard disk with area a on the surface of the central, spherical A particle. We then monitor the number of disks $[0, i-1]$ overlapping a probe area of size a. We compute $P_a(i-1)$ over a range of diameter ratios for $i = 3, \ldots, n$. The probability of finding binding space decays faster than exponentially as the maximum diameter ratio α_n^+ is approached.

Using the probabilities $P_a(n-1)$ we can make a simple estimate of the bimolecular rate constant given in Eq. (6.12) without explicitly computing k^*. The integral in Eq. (6.9) can be approximated as,

$$k^* \approx \frac{D}{\sigma_B^2} e^{-\beta \Delta w} \tag{6.10}$$

where Δw is the maximum of the barrier in the potential of mean force for the AB_{i-1} to B distance and σ_B is the diameter of the B particle. Using,

$$e^{-\beta \Delta w} \approx P_a(n-1), \tag{6.11}$$

we get a simple approximate expression for the full bimolecular rate constant,

$$k_b^{(i)} = \frac{4\pi DR}{1 + \frac{D\sigma_B^2}{RP_a(i-1)}}. \tag{6.12}$$

While the approximation we have made is convenient and provides a transparent physical mechanism for the increases we see in the yield, we can also compute the integral in Eq. (6.9) can be computed from simulations of particle dynamics. Doing so entails computing $p(r)$, which could be done using umbrella sampling.

We can compute the yield using the reaction rates derived above. The concentration of sterically accessible clusters is tracked by a vector

$$\boldsymbol{\rho}(t) = \begin{pmatrix} \rho_A \\ \rho_{AB} \\ \vdots \\ \rho_{AB_n} \end{pmatrix}, \tag{6.13}$$

which initially has $\rho_A = \rho_A^{\text{bulk}}$ and all other cluster densities set to zero. The time evolution of the cluster density is described by the master equation,

$$\partial_t \boldsymbol{\rho} = W\boldsymbol{\rho}, \tag{6.14}$$

where W is the rate matrix,

$$W = \begin{pmatrix} -k_b^{(0)}\rho_B & & & & \\ k_b^{(0)}\rho_B & -k_b^{(1)}\rho_B & & & \\ & \ddots & \ddots & & \\ & & k_b^{(n-2)}\rho_B & -k_b^{(n-1)}\rho_B & \\ & & & k_b^{(n-1)}\rho_B & 0 \end{pmatrix} \tag{6.15}$$

which has a solution,

$$\boldsymbol{\rho}(t) = e^{Wt}\boldsymbol{\rho}(0). \tag{6.16}$$

The systems we are considering here have a very high concentration of B particles to A particles, so we assume that depletion effects are negligible and that ρ_B is constant.

Diagonalizing W is straightforward and the eigenvalues of the rate matrix are

$$\boldsymbol{\lambda} = \left(0, -k_b^{(0)}, \ldots, -k_b^{(n-1)}\right). \tag{6.17}$$

The fact that the dominant mode is a zero eigenmode should come as no surprise: the long time limit of the irreversible dynamics of attachment should result in complete saturation of the AB_n stoichiometry. However, the decay towards the final equilibrium state can be extremely slow, depending on α, so it is the transient behavior that determines the yield.

The expected yield can now be expressed as a function of the diameter ratio α at time t

$$\eta(\alpha, t) = \rho_{AB_n}(\alpha, t)/\rho_A(\alpha, 0). \tag{6.18}$$

This quantity is a measure of the efficiency of the equilibrium assembly process; it gives the fraction of A particles that reach the desired, asymptotic equilibrium state, in which the maximum number of B particles are bound. The yield η asymptotically approaches unity in the long time limit, but we are seeking strategies to increase η for finite times.

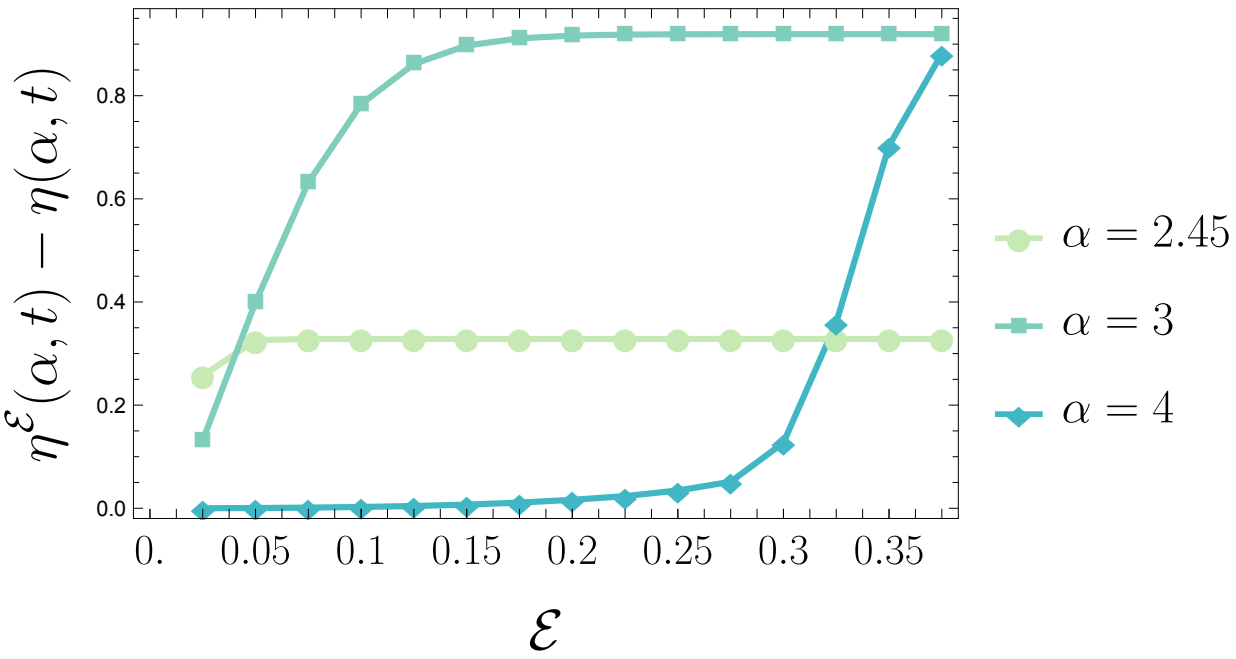

Figure 6.2: The yield as a function of the Monte Carlo bias $\mathcal{E}$. The yield is plotted here for AB_4 clusters in three distinct regimes, $\alpha = 2.45 \approx \alpha_-$, $\alpha = 4 \approx \alpha_+$ and the intermediate value $\alpha = 3$.

6.2 Enhancing the yield

In order to increase the finite time yield, we seek to enhance the rate of $A - B$ binding events using a nonequilibrium drive. The bimolecular reaction rate $k_b^{(i)}$ is highly sensitive to the probability of finding sufficient free binding surface, P_a. This suggests that manipulating P_a by some external means should modulate the yield, $\eta(\alpha, t)$. To test this principle, we sampled the occupancy probability in the presence of an external field $\mathcal{E}$, which we write $P_a^{\mathcal{E}}(i)$. The field $\mathcal{E}$ can be represented in hard disk Monte Carlo simulations as a bias on the displacements of the B particles along the z-axis.

For a broad range of diameter ratios, the kinetic model predicts enhanced yields with the application of an external drive. Fig. 6.2 shows the yield $\eta^{\mathcal{E}}(\alpha, t)$ as a function of field strength $\mathcal{E}$ for $\alpha = 2.45, 3$, and 4. Because the equilibrium yield is high near $\alpha_- \approx 2.45$, as reported in Ref. [184], the enhancement is weak with external driving. This is evident in Fig. 6.2, where the increase in yield saturates under increasing field strength near 0.3. At $\alpha = 4$, which is near α_+, the available area to bind the final particle is sufficiently small that the field cannot perform the reversible work needed to shift the distribution of P_a, except under strong driving. Hence, near α_- and α_+ the expected enhancements are weak for modest driving. However, in the intermediate regime, $\alpha \approx \frac{\alpha_- + \alpha_+}{2}$, the yield is enhanced

dramatically. At $\alpha = 3$, shown in Fig. 6.2, the equilibrium yield is increased nearly 80% with field strengths that produce essentially no change near α_+.

6.3 Simulations of particle dynamics

To test the predictions of the kinetic model more throughly, we conducted Brownian dynamics simulations of the assembly process. While the kinetic model predicts that driving the colloidal particles with an external field should increase the yield, that analysis neglects the detailed dynamics of the particles. To demonstrate that yield is substantially shifted even when such correlations are included, we performed dynamical simulations of the particle assembly process. We treated the A and B particles as charged particles, with charges $\boldsymbol{q}$ subject to an over-damped Langevin equation of motion,

$$\dot{X} = -\big(\nabla V(X) + \mathcal{E}\boldsymbol{q}\big)dt + \sqrt{2\beta^{-1}}\xi. \tag{6.19}$$

The potential energy $V(X)$ is designed to mimic the experimental colloidal systems studied in Schade et al. [184]. First, we set the radius of the A and B particles via a Weeks-Chandler-Anderson potential

$$V_{\mathrm{WCA}}(X) = \begin{cases} \sum_{ij} 4\epsilon \left[\left(\frac{\sigma_{ij}}{|X_i - X_j|} \right)^{12} - \left(\frac{\sigma_{ij}}{|X_i - X_j|} \right)^{6} \right] + \epsilon & |X_i - X_j| \leq 2^{1/6}\sigma_{ij} \\ 0 & |X_i - X_j| > 2^{1/6}\sigma_{ij}. \end{cases} \tag{6.20}$$

The parameter σ determines the hard sphere radius of the A and B particles. For $i, j = A$ it is set to $\sigma_{ij} = \sigma_A$, with the same convention for B particles. The sum of the hard sphere radii is used for $A - B$ interactions. We used $\sigma_A = 1$ and $\sigma_B = \alpha$ throughout. In units of $k_{\mathrm{B}}T$, $\epsilon = 10$ in all simulations.

The attractive interaction between the charged colloids is extremely short-ranged due to short Debye length of the solutions used in experimental settings. Because the attraction is observed to be essentially irreversible, we approximate this interaction using a Morse potential,

$$V_{\mathrm{M}}(X) = \sum_{ij} D_0 \left[\exp(-2\alpha(|X_i - X_j| - r_0) - 2\exp(-\alpha(|X_i - X_j| - r_0)) \right], \tag{6.21}$$

with $D_0 = 15$ and $\alpha = 10$. We truncated the Morse interaction slightly ($0.25\sigma_A$) above the cut-off of the WCA potential, which creates a short-ranged, deep potential. We integrated the equations of motion with a time-step of 5×10^{-5} for 5×10^5 total steps. All simulations were conducted with the HOOMD software package, v1.3.3 [192, 193].

We computed the yield $\eta(\alpha, t_{\mathrm{obs}})$ for a wide range of diameter ratios. The equilibrium results, shown in Fig. 6.3 (a), indicate that the asymptotic state is not accessible on simulation timescales for many values of α. Instead, the yield curves reach a maximal value for each n and then decrease sharply as the maximum diameter ratio is approached.

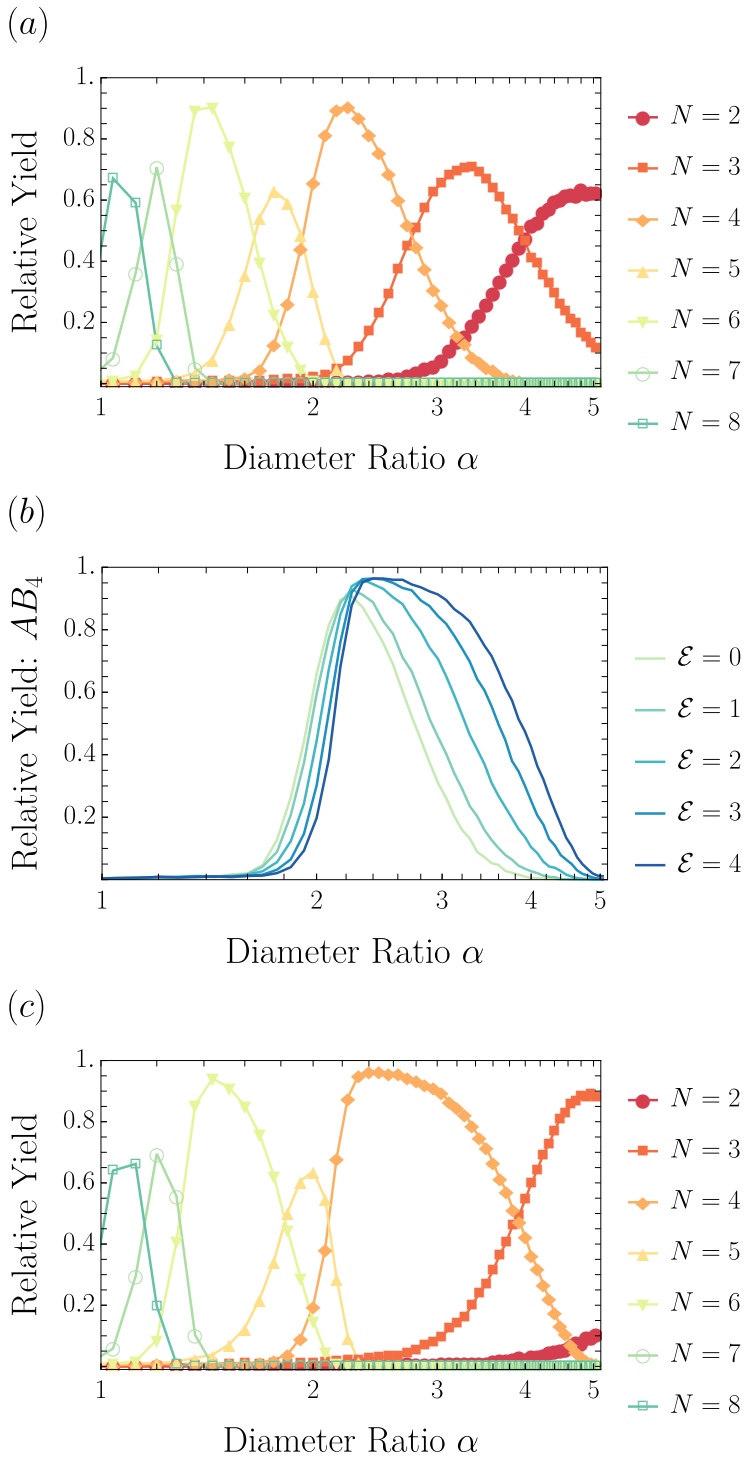

Figure 6.3: The relative yield computed from particle-based Brownian dynamics simulations at $t = 5 \times 10^5 \Delta t$ for various cluster stoichiometries $A : B = 1 : N$ and electric field strengths $\mathcal{E}$. Relative yield is defined as the fraction of clusters of type AB_N over the total number of A particles. (a) The equilibrium yield over a large range of diameter ratios. (b) The AB_4 yield for a range of external electric field strengths. (c) The yield over a large range of diameter ratios with external electric field $\mathcal{E} = 4$.

To test the principle suggested by the kinetic model, we conducted simulations with the same parameters described above with an additional external field $\mathcal{E}$. The external field drives the oppositely charged A and B in opposite directions, modulating both the probability of finding an available binding surface P_a and also the probability of particle encounters. As illustrated by Fig. 6.3 (b), the yield of AB_4 clusters can be enhanced from near zero to near unity under sufficiently strong driving, as is the case for $\alpha = 3.5$. This phenomenon is not unique to the AB_4 clusters: Fig. 6.3 shows that with $\mathcal{E} = 4$ the relative yield of each cluster type AB_n is enhanced.

The full yield curves, Fig. 6.3, are enhanced relative to the equilibrium yield over a broad range of diameter ratios. Note that the fastest relaxing kinetic intermediates, the AB_2 clusters, are almost entirely depleted. Further, AB_3 and AB_4 clusters are formed with high probability even at diameter ratios that very rarely successfully assemble under equilibrium conditions. While the enhanced yields evident in Fig. 6.3 are predicted by considerations of the microscopic binding kinetics, under very strong driving conditions the $A - B$ binding interaction can be disrupted. When the driving is strong enough to break apart existing clusters, the predictions of the kinetic model fail because the assumption of irreversible binding is violated.

6.4 Dissipation-yield trade-off

In order to accelerate the rate of AB_n cluster formation, we have expended energy. When the dynamics are unperturbed, the rate of particle additions to the growing clusters depends on the reversible work required to create a void for the next particle, via Eq. (6.12). As we have shown, the probability of finding such a void can be easily modulated by the applying an external potential. The accelerated assembly, however, requires that we do work on partially formed colloidal clusters. The work in excess of the reversible work must be dissipated to the bath as heat according to the first law of thermodynamics.

We can view the dissipated work as a cost that we pay for higher yields. We can study this trade-off between work and yield by introducing an enhancement factor

$$\Theta(\mathcal{E}, \alpha, t) = \frac{\eta^{\mathcal{E}}(\alpha, t) - \eta(\alpha, t)}{\mathcal{W}}. \tag{6.22}$$

This quantity measures the change in yield from the equilibrium yield due to the work $\mathcal{W}$ performed on the system under the application of the external field $\mathcal{E}$. The dissipative cost of the driving is proportional to $\mathcal{E}$, so in the case of the kinetic model, we simply use $\mathcal{E}$ in the denominator of this generalized efficiency.

The average reversible work done on a cluster by the external field can be expressed as a function of the probabilities P_a. In particular, we note the reversible work required to create a vacancy of area a on a cluster with $n - 1$ particles bound is,

$$\mathcal{W}_{\text{rev}} = -k_{\text{B}}T \ln P_a(n-1). \tag{6.23}$$

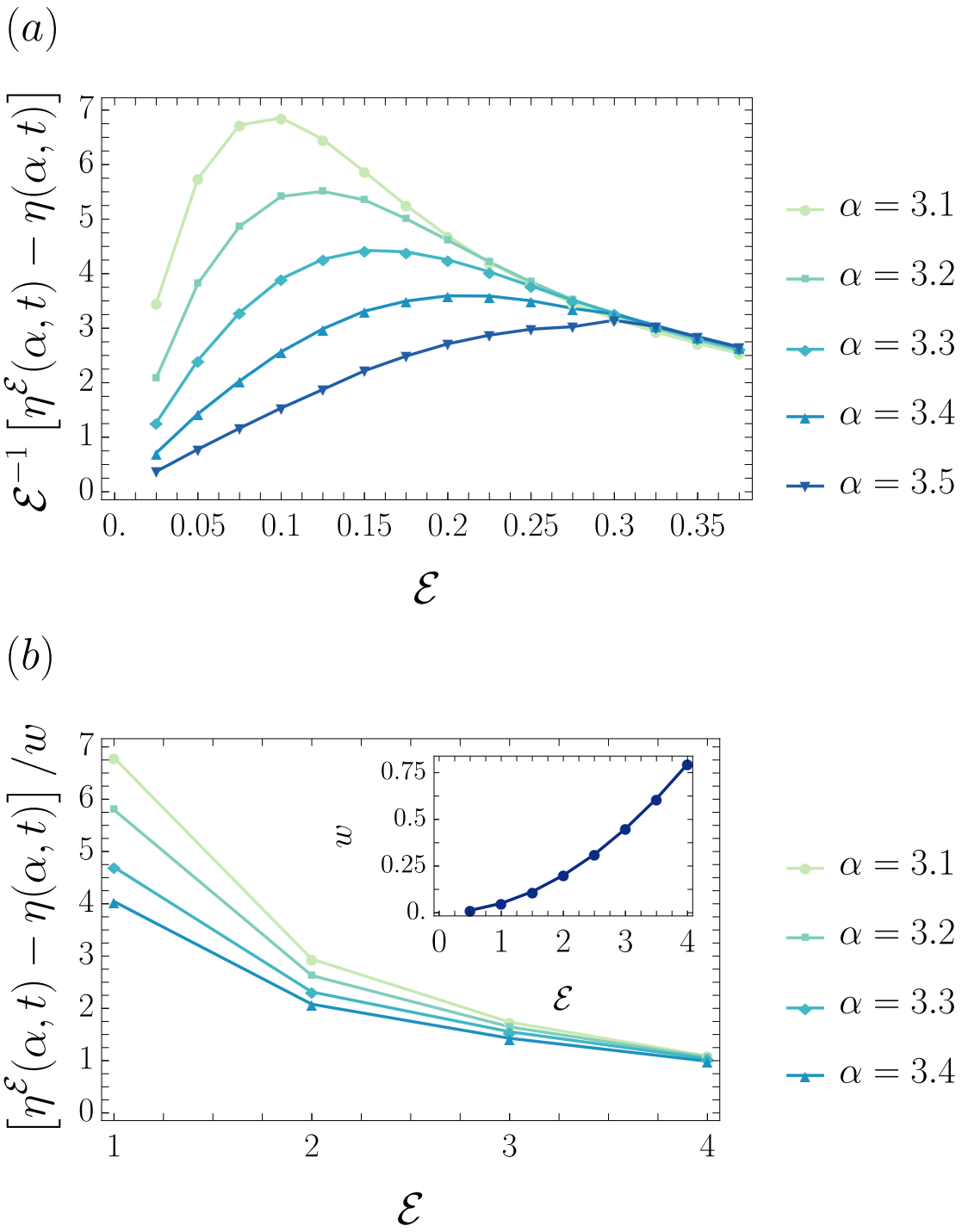

Figure 6.4: The yield enhancement of AB_4 clusters at diameter ratios in the range $\alpha \in [3.1, 3.5]$. (a) The kinetic model. In all cases, the enhancement factor Θ (Eq. (6.22)) shows a maximum. In the kinetic model, we relate the work $\mathcal{W}$ to the field strength $\mathcal{E}$. (b) Particle-based simulations. The strong field values lead to a diminishing enhancement. Inset: the nonlinear scaling of the work per particle w with external field strength should be contrasted with the linear response approximation in (a).

Then we can compute the average work on such clusters associated with the external field $\mathcal{E}$,

$$\Delta\mathcal{W}_{\text{rev}} = \mathcal{W}_{\text{rev}}^{\mathcal{E}} - \mathcal{W}_{\text{rev}} \tag{6.24}$$

$$= -k_{\text{B}}T\left[\ln P_a^{\mathcal{E}}(n-1) - \ln P_a(n-1)\right]. \tag{6.25}$$

The quantity $\Delta\mathcal{W}$ does not account for the total work done on the system by the external field, but rather establishes a direct correspondence between the work and the enhanced yield.

The total work done by the external field on average is,

$$\mathcal{W}_{\text{tot}} = \mathcal{E}\langle \boldsymbol{q} \cdot \Delta X_z \rangle_t, \tag{6.26}$$

where ΔX_z is the total displacement in the z-direction and the average is computed over a trajectory of time t. We computed the work done by the external field on the system for dynamical assembly trajectories to gauge the trade-off between the amount of work being dissipated to the bath and the enhanced yield, as shown in Fig. 6.4. Because the probability $P_a^{\mathcal{E}}(n-1)$ has a maximum value at one, the effect of the field $\mathcal{E}$ becomes saturated as $\mathcal{E}$ increases. That is, when comparing two large values of the external drive, the change in total dissipation will be proportional to $\Delta\mathcal{E}$, but the shift in $\Delta\mathcal{W}_{\text{rev}}$, and hence further yield enhancement, can be small. Interestingly, the diminishing returns of the nonequilibrium work lead to an optimum in the enhancement efficiency. In Fig. 6.4, we plot $\Theta(\mathcal{E}, \alpha, t)$ for $\alpha \in [3.1, 3.5]$ as function of the external field strength, where the target cluster is AB_4. At the smaller values of α, there is appreciable free area on the surface of the A particle, but the equilibrium yield is low because P_a is small. Weak driving can substantially improve the yield, whereas the gains lessen at larger driving strengths. This leads to a maximum in the driving efficiency, which clearly illustrates a trade-off between the dissipative work and the yield.

For larger diameter ratios, the reversible work required to expose the binding surface with sufficient probability to enhance the yield is higher. The additional work required to change the yield shifts the optimal $\mathcal{E}$ to higher driving strengths. Under strong driving conditions, P_a approaches unity, leading all the curves to collapse to a similar value of Θ.

Sophisticated design principles for self-assembly provide us with opportunities to fabricate intricate microscopic structures with exotic material properties. Even so, building the components that realize these design principles can be prohibitively complicated, for example, requiring particles with tightly constrained sizes [184], bond orientations [194, 195], or geometric structures [196]. The intent of these design principles, of course, is to provide us with systems that rapidly and robustly relax to the desired stable state.

Typically, the assembly process is carried out or modeled in equilibrium conditions. If the components can avoid kinetic trapping and rapidly relax to the final equilibrium state, then the yield of the desired structure will be high. Both in experimental and computational settings, this is rarely the case. Long-lived kinetic intermediates prevent the system from accessing the desired state, even if it is lower in free energy, on experimental timescales.

While we have focused on the relationship between the sphere size ratio and cluster yield [184], numerous strategies have been employed in an effort to abet relaxation towards the desired state. Murugan et al. [183], for example, studied the relationship between yield and the stoichiometric make up of the components. Here, we have taken a distinct perspective: we rely on external perturbations to drive the system towards the desired state. In doing so, we transiently drive the system out of equilibrium in such a way that an equilibrium state becomes *more* accessible. While the example that we have studied does not require a complicated protocol for external control, it may be possible to leverage additional freedom in the design space of protocols, e.g., time-dependent external forces or feedback.

The principal advantage of our approach is its flexibility. Regardless of the components used in the assembly, or their propensity to become kinetically trapped, we could design a nonequilibrium protocol to ameliorate slow equilibration. The increased yield comes at a cost—we must do work on the system at a finite rate to enhance the probability of success. The trade-off that we find between the dissipated work and the self-assembly yields instantiates a more general phenomenon, the putative energy-speed-accuracy balance of nonequilibrium dynamics [50, 160]. It remains a significant future challenge to distill generic principles for designing nonequilibrium protocols that will enhance the yield.

Chapter 7

Nonequilibrium Nanoencapsulation

Microorganisms spatially stratify cellular contents to create organization in a chaotic, fluctuating environment. Mesoscopic protein shells called bacterial microcomparments, for example, segregate enzymes to amplify throughput in carbon fixation, a process necessary for the survival of the organism [197–200]. Carbon fixation has become a topic of acute interest and recently the prospect of re-engineering bacterial microcompartments has been explored experimentally [201]. Self-assembling microcompartment structures, which are found in a variety of photosynthetic bacteria, are believed to function as reaction chambers and enhance enzymatic activity by concentrating the reactants in a confined, nanoscopic volume [202–204]. A well-characterized member of this class is the carboxysome, a microcompartment that aids carbon-fixation in cyanobacteria by concentrating the enzyme RuBisCo to near crystalline densities. The proteinaceous shell that coats the enzymatic core passively filters reactants to ensure that concentrations of CO_2 remain high within the compartment [204].

Experimental studies of the carboxysome structure have highlighted a variety of its surprising characteristics. It forms a highly faceted shell [205–207], consisting of protein capsomers with sixfold symmetry. Topological constraints discussed later in this chapter require that any closed shell must also have a combination of five and sevenfold defects. Unlike some highly-symmetrical virus shells, the size of the carboxysome does not appear to be intrinsically templated by its components, which makes its apparent geometry all the more surprising. Further, it has been reported that the observed structures are monodisperse, with a size distribution peaked near 100 nm in diameter.

Both the regular geometric structure and tight size distribution of the shell defy expectations based on previous efforts to model the assembly of viruses and other symmetrical nanoparticles [188, 208]. Generally, kinetic intermediates disrupt an assembly process where features, like the vertices of the icosahedron in the carboxysome, must be globally coordinated throughout the structure. In order to robustly maintain a regular size and shape, the probability of incorporating a fivefold defect during the assembly would need to depend very sensitively on the distance from other defects.

Experiments have attempted to examine the carboxysome assembly process directly, but have been limited in both spatial and temporal resolution [209]. It has been shown,

nevertheless, that the biogenesis of a carboxysome likely relies on nucleation from an existing structure. While the fluorescence microscopy data in Ref. [209] cannot provide detailed dynamical or spatial information about nucleation mechanism or kinetics, the putative model put forth suggests that existing shell material may act as a nucleation site for enzymatic cargo. The fluctuations in the size and shape of the cargo could, in principle, play a significant role in determining the morphology of the protein shell.

The limited understanding that we have of the factors that control encapsulation has hindered efforts to rationally design nanoscale compartments [201]. In fact, even the tightly constrained problem of viral shell self-assembly is not fully understood [178]. In this chapter, we describe a minimal model of nanoshell assembly to help elucidate the conditions necessary for successful encapsulation. We use our model to explore the dynamics of the shell growth process in two distinct scenarios: in the presence of a static, spherical cargo, and in the presence of a fluctuating cargo. Our expressly coarse-grained model is not meant to reveal explicit features of carboxysome assembly that may depend on the detailed features of the proteins involved. Rather, the aim of such a model is to diagnose the necessary physical ingredients for encapsulation to occur. The general nature of our model confers on us the ability to make transferable predictions and uncover design principles for the assembly. Such design principles would facilitate the development of synthetic encapsulations systems with control over the size and contents.

7.1 A minimal model of nanoshell growth

In order to model these phenomena in great generality, we have built a minimal model that represents each protein in the shell as a triangular monomer on a lattice governed by an elastic Hamiltonian, as shown in Fig. 7.1. The assembly dynamics of the nanoshell has two salient physical features: the chemical potential associated with shell's constituent proteins and the elastic energy of the growing shell. This elastic sheet then undergoes a grand canonical Monte Carlo dynamics that inserts and deletes monomers. To study the growth of the shell, we initialize the system with a nonequilibrium concentration of monomers and observe the subsequent relaxation.

Many previous studies have assessed the ways in which elastic properties determine the low energy structures at constant capsule volume, but have not modeled the discrete growth events. Work on virus capsid structures has employed continuum elasticity theory in an effort to predict the shape of the low energy structures in various elastic regimes [210–212]. A discretized representation of the elastic Hamiltonian was used in Ref. [213] to study the diffusion of defects on thin elastic shells.

A separate line of inquiry has analyzed growth using models of discrete triangulated representations of elastic membranes, similar to the one described in this chapter. The Hamiltonians used in those studies impose a global curvature in the elastic energy, and hence template the size of the resulting structure [214–216]. What is more, the dynamics violates microscopic reversibility, complicating any physical interpretation of the trajecto-

ries and resulting structures. Hicks and Henley [216] provide an assessment of the "failure modes" of the structures that arise from the dynamics of their model. They find that a large number of kinetically accessible structures result in configurations that cannot form a closed shell. The physical significance of the observed structures must be disentangled from the nonequilibrium effects of their irreversible growth dynamics, which can present a challenge to interpretation. Some of the kinetic traps have been circumvented with an alternative procedure, described in Ref. [215]. The dynamics remains irreversible and insertion moves are proposed deterministically at the growing edge with minimal angle, which the authors justify as growth along a minimum energy path. This procedure mitigates the irregular defect incorporation observed in previous models [216]. Our approach, on the other hand, focuses on equilibrium dynamics, so the model we have designed takes care to ensure that detailed balance is preserved. The local dynamics in our model ensures microscopic reversibility, an indispensable feature of Markovian dynamics in and out of equilibrium.

7.2 Elastic description of thin shell mechanics

The continuum elasticity theory of a thin plate has been applied to biological protein assemblies, such as virus capsids, with surprising fidelity to the experimentally observed structures [210]. A thin elastic membrane resists both stretching and out-of-plane fluctuations. We can represent the total energy of a thin elastic membrane as a sum of these two contributions,

$$\mathcal{H} = \mathcal{H}_{\text{stretch}} + \mathcal{H}_{\text{bend}}, \tag{7.1}$$

where the terms denote the in-plane stretching energy and the out-of-plane bending energies, respectively. We define a coordinate system that measures relative displacements,

$$u(x) = (x_1 + u_1, x_2 + u_2, z) \tag{7.2}$$

which in turn defines a strain tensor,

$$u_{ij} = \frac{1}{2}\left(\partial_i u_j + \partial_j u_i + \partial_i z \partial_j z\right), \tag{7.3}$$

neglecting the terms quadratic in u_k. The stretching energy can be expressed in terms of Lamé coefficients μ_{L} and λ_{L} and the strain tensor u_{ij},

$$\mathcal{H}_{\text{stretch}} = \frac{1}{2}\int dS \ \left(2\mu_{\text{L}} u_{ij}^2 + \lambda_{\text{L}} u_{kk}^2\right). \tag{7.4}$$

The integral is taken over some parametrization of the surface of the membrane. Familiar material properties can be expressed in terms of the Lamé coefficients. For example, the Young's modulus, which measures the tensile stress relative to the extensional strain of an in-plane deformation is

$$Y = \frac{4\mu_{\text{L}}\left(\mu_{\text{L}} + \lambda_{\text{L}}\right)}{2\mu_{\text{L}} + \lambda_{\text{L}}}. \tag{7.5}$$

The Poisson ratio, a material characteristic that relates the compression along the direction of an applied strain to the expansion along a transverse direction, is

$$\nu = \frac{\lambda_{\mathrm{L}}}{2\mu_{\mathrm{L}} + \lambda_{\mathrm{L}}}. \tag{7.6}$$

These material properties determine, in part, the propensity of a thin elastic shell to "buckle", developing visible, aspherical facets.

The bending contribution to the Hamiltonian enforces a quadratic cost to changes in the local curvature. Written in terms of the mean curvature H, the Gaussian curvature K, and the rigidities $\tilde{\kappa}, \tilde{\kappa}_{\mathrm{G}}$, the bending energy is

$$\mathcal{H}_{\mathrm{bend}} = \frac{1}{2} \int dS \ \left(\tilde{\kappa} H^2 + 2\tilde{\kappa}_{\mathrm{G}} K \right). \tag{7.7}$$

The Gaussian curvature is a topological invariant, due to the Gauss-Bonnet theorem; it is ignored as an arbitrary constant in most analytical calculations [217]. The continuum description of the membrane is particularly convenient for highly symmetrical objects, for which some calculations can be performed analytically. However, this formulation of the energy is not immediately applicable to the discrete growth of a nanoscale shell.

The elastic energy associated with the thin sheet can be discretized and expressed as a sum over the edges ij of a triangulated surface [218]. The total Hamiltonian is,

$$\mathcal{H} = \frac{\epsilon}{2} \sum_{ij} (l_{ij} - l_0)^2 + \frac{\kappa}{2} \sum_{ij} 1 - \cos(\theta_{ij}), \tag{7.8}$$

where ϵ is the renormalized stretching constant and $\kappa = \frac{2\sqrt{3}}{3}\tilde{\kappa}$ is the renormalized bending rigidity. The scalars relating the continuous and discrete Hamiltonians arise from a detailed analysis that involves taking the limit of infinitesimally small edge lengths so that the discretized Hamiltonian coincides exactly with continuum description of a thin elastic sheet. Such a calculation [218] exposes the precise relationship between ϵ and κ and the parameters of the continuum model. In particular, it has been shown,

$$Y = \frac{2\sqrt{3}}{3}\epsilon \qquad \nu = \frac{1}{3} \tag{7.9}$$

$$\tilde{\kappa} = \frac{\sqrt{3}}{2}\kappa \qquad \tilde{\kappa}_{\mathrm{G}} = -\frac{4}{3}\kappa. \tag{7.10}$$

It is then possible to choose parameters in the discretized model that are consistent with the experimentally predicted material properties of protein shells [210].

We associate each triangular face in our discrete lattice with a protein monomer. Because the deformations of the protein are expected to be small relative to the bending fluctuations between adjacent monomers, we work in a limit where $\epsilon l_0^2 \gg \kappa$. Previous calculations give

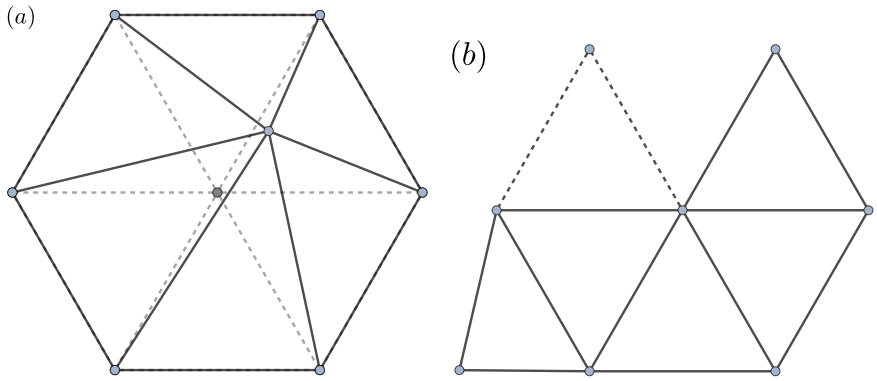

Figure 7.1: A schematic depiction of the vertex move proposal. (a) A vertex is randomly displaced and the energy difference between the initial and final states is computed. The move is then accepted or rejected so as to guarantee detailed balance. A schematic depiction of the monomer insertion move. (b) A new monomer is added to the surface of the growing structure by randomly selecting a surface edge. The configuration of the inserted monomer is chosen randomly and the insertion is accepted in a detailed balance fashion.

further guidance on the elastic parameters: if we anticipate well-defined icosahedral structures, then we need to be in a regime where the ratio of the Young's modulus to the bending rigidity, scaled by the radius of the shell, is

$$YR^2/\kappa > 154, \tag{7.11}$$

the point at which a flat disk will be noticeably buckled around a defect. This ratio (7.11) is called the Föppl-von Kármán number and is a measure of the relative cost of a stretching deformation to a bending deformation.

7.3 Monte Carlo dynamics of the model

The thermal relaxation of the shell subject to the Hamiltonian discussed in the previous section can be simulated with a straightforward Monte Carlo dynamics. In each move, a single vertex v of the lattice is chosen at random. As depicted in Fig. 7.1, a random perturbation in three-dimensional space is made to the selected vertex, giving a new coordinate v'. The energy difference, ΔE, which can be computed locally, is used in a standard Metropolis acceptance criterion,

$$\mathrm{acc}(v \to v') = \min\left[1, \exp\left(-\beta \Delta E\right)\right]. \tag{7.12}$$

Following this procedure ensures that the configurations sampled in the Markov chain are consistent with a Boltzmann distribution.

The Monte Carlo move set for growth of the shell is significantly more involved. In order that the dynamics satisfy detailed balance, care must be taken throughout to ensure that every step is reversible and the generation probabilities are properly incorporated into the acceptance criterion. Our goal is to draw samples from a grand canonical distribution,

$$p(X,N) = \Xi^{-1} \left(\frac{e^{\beta\mu}}{\lambda^9} \right)^N e^{-\beta E(X)}, \tag{7.13}$$

where Ξ denotes the grand canonical partition function, X gives the coordinates the configuration, N is the number of monomers, and the length scale λ establishes a standard state for the monomer chemical potential μ. The factor λ^9 appears in the denominator because each monomer consists of three connected three-dimensional degrees of freedom. We define the "activity",

$$z = \frac{e^{\beta\mu}}{\lambda^9}. \tag{7.14}$$

Developing a Monte Carlo move set that leaves this distribution invariant requires extreme care.

The simplest type of growth move that we implement is depicted in Fig. 7.1. First, an edge at the surface of the growing sheet is selected at random; there are N_{surf} possible choices. A new monomer is proposed by initially generating an "ideal" insertion, i.e., a configuration that is aligned perfectly with the monomer to which it is attaching with edge lengths close to the equilibrium values. The newly inserted vertex is subsequently randomly selected within a volume v_b. This final step ensures that the procedure is statistically reversible.

In our model the vertices of bound monomers coincide exactly along the shared edge. The physical situation we have in mind is that separating such vertices by even a small distance r_a costs a large energy. For a favorable binding energy ϵ that is constant within the small volume $v_a = 4\pi r_a^3/3$, the affinity $K = v_a e^{-\beta\epsilon}$ completely characterizes the equilibrium statistics of binding. In setting $r_a = 0$, we implicitly take $\epsilon \to -\infty$ such that K remains nonzero and finite. We thus impose a well-defined binding equilibrium without resolving fluctuations of bound monomers on irrelevantly small scales. The constant K must be appropriately incorporated into the acceptance criterion for the growth move to ensure detailed balance.

We are working a regime in which the Young's modulus (7.5) of the thin sheet is large. A typical displacement of a vertex governed by the Hamiltonian $\mathcal{H}$ of Eq. (7.8) is of order,

$$\delta \approx (\beta\epsilon)^{-1/2} \equiv \xi, \tag{7.15}$$

so we take ξ as a natural unit of length. We can then define a dimensionless affinity,

$$K^* = K\xi^{-3}. \tag{7.16}$$

The fusion moves discussed in detail later depend on K^* and v^*_{fuse}. We want to be working in a limit where the probability of fission is small. Accordingly, we choose another parameter,

$$f^* \equiv K^*/v_{\text{fuse}}* \gg 1. \tag{7.17}$$

When f^* is large, the probability of accepting a proposed fusion move is large and, inversely, the probability of accepting a fission move is small.

The move that we have described generates a configuration X' from another configuration X which possesses one fewer monomer. In order to leave the distribution sampled by the Markov chain invariant, we must have

$$p(X \to X')p(X) = p(X' \to X)p(X') \tag{7.18}$$

The acceptance probability $\mathrm{acc}(X \to X')$ for the insertion of a monomer must incorporate the generation probabilities $\mathrm{gen}(X \to X')$ and $\mathrm{gen}(X' \to X)$, corresponding to insertion and deletion, respectively, so that

$$p(X \to X') = \mathrm{gen}(X \to X')\mathrm{acc}(X \to X') \tag{7.19}$$

satisfies Eq. (7.18). We employ the Metropolis criterion for this purpose

$$\mathrm{acc}(X \to X') = \min\left[1, \frac{\mathrm{gen}(X' \to X)P(X')}{\mathrm{gen}(X \to X')P(X)}\right]. \tag{7.20}$$

In some cases, it may be possible to insert a monomer by adding a single edge to the system, e.g., adding an edge between the top two vertices in Fig. 7.1. When a surface edge is selected at random, if a new monomer can be created by adding an edge of length $l < l_{\max}$ to another surface vertex, we propose the edge insertion move. The number of edges for which such a move is possible is N_{wedge}.

The generation probability for a monomer insertion is thus determined by the probability of selecting an eligible edge for monomer insertion that is not among the N_{wedge} edges

$$\mathrm{gen}(X \to X') = \frac{1}{v_b(N_{\mathrm{surf}}(X) - N_{\mathrm{wedge}}(X))}. \tag{7.21}$$

This expression tabulates the probability of choosing a particular surface edge and point within the volume v_b. The reverse move depends on selecting one of the two surface edges of a monomer with a single bound edge. The generation probability in the reverse direction is

$$\mathrm{gen}(X' \to X) = \frac{1}{N_{\mathrm{mono,1}}(X')}. \tag{7.22}$$

where $N_{\mathrm{mono,1}}$ is the number of surface edges with a single bound edge in configuration X', i.e., the set of monomers that could have been generated by a monomer insertion move.

The grand canonical probability of the configuration X' relative to the initial configuration is

$$\frac{p(X')}{p(X)} = zK^2 \exp\left(-\beta\Delta E\right), \tag{7.23}$$

where we have effectively integrated over binding fluctuations on the unresolved scale r_a. The factors of K resulting from this integration account for the binding energy and entropy of the vertices.

The energy difference ΔE can be computed locally and involves the elastic energy of the edges of the new monomer and the interaction with the cargo. The acceptance criterion for the insertion move is thus

$$\mathrm{acc}(X \to X') = \min\left[1, zK^2 \frac{v_b(N_{\mathrm{surf}}(X) - N_{\mathrm{wedge}}(X))}{N_{\mathrm{mono},1}(X')} e^{-\beta\Delta E}\right], \tag{7.24}$$

which acts as a guarantee of detailed balance.

The probability of proposing the monomer insertion moves with two bound edges depends only on N_{wedge}

$$\mathrm{gen}(X \to X') = \frac{1}{N_{\mathrm{wedge}}(X)} \tag{7.25}$$

The reverse move has a generation probability that is determined by the number of surface edges that have two bound edges, $N_{\mathrm{mono},2}$, the set of monomer edges that could have been generated by the forward move. The acceptance criterion for insertion moves of this type is

$$\mathrm{acc}(X \to X') = \min\left[1, zK^3 \frac{N_{\mathrm{wedge}}(X)}{N_{\mathrm{mono},2}(X')} e^{-\beta\Delta E}\right]. \tag{7.26}$$

Two nearby edges must also be able to join, to represent the binding of nearby edges. The likelihood of an edge fusion event is determined by the favorable binding interaction described above. If there is a pair of unconnected surface vertices v and v' within a distance l_{fuse}, we attempt to join the vertices at their midpoint. In order for a fusion move to be proposed, the two vertices must either share a mutual neighbor or the involved surface edges must share a surface edge. Using the convention that K estimates the partition function of the bound vertices, we can specify an acceptance criterion that will lead to a detailed balance dynamics. The acceptance criterion depends on K and a volume v_b of a sphere with radius $l_{\mathrm{fuse}}/2$,

$$\mathrm{acc}(\mathrm{fusion}) = \min\left[1, \frac{K}{v_b} e^{-\beta\Delta E}\right]. \tag{7.27}$$

The energy difference ΔE is due to bond stretching and bending resulting from the vertex fusion.

The simulations described here are in a regime where the reverse move, which we call vertex fission, is extremely unlikely. Nevertheless, we specify and implement it to ensure that we are sampling a well-defined distribution. The fission move is illustrated in Fig. 7.2. We select an eligible vertex v and choose a random point in the sphere of radius $l_{\mathrm{fuse}}/2$, with a volume we call v_{fuse}. Not all surface vertices are eligible for fission. However, the set of fissionable vertices is precisely the set of fusible vertices, so there is no additional contribution to the acceptance criterion. This constraint only amounts to a detail in the implementation.

The random point we choose within the volume v_b designated as the location of one of the two new vertices. The other is uniquely determined, as well, because the midpoint of the vector connecting the new vertices is v. A randomly selected non-surface edge connected

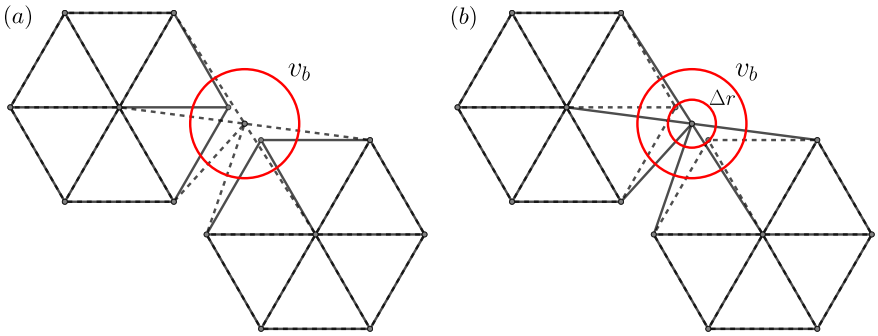

Figure 7.2: Vertices of the lattice must be able to combine and split. In order to do this in a way that preserves detailed balance, we choose the fusion point randomly within some volume v_b, depicted in (a). The reverse operation, splitting vertices, requires choosing a pair of points along a symmetric axis within the volume v_b, as shown in (b).

to v is duplicated and the neighbors of v become are split between the two new vertices. The acceptance criterion for the fission move can be constructed so that the dynamics obeys detailed balance. Its exact expression is

$$\text{acc(fission)} = \min\left[1, \frac{v_b}{K} e^{-\beta(\Delta E)}\right] \tag{7.28}$$

In practice, no fission moves are observed in the course of a simulation owing to the relatively large values of K that, in the case of fission, are not counteracted by the activity z.

Our model of nanoencapsulation also includes the molecules within the shell. We model the cargo as an Ising lattice gas on a simple cubic lattice, allowing it fluctuate as a liquid-like droplet. Nearest neighbors are coupled through an interaction energy ϵ_{C} and chemical potential μ_{C},

$$\mathcal{H} = -\mu_{\mathrm{C}} \sum_{i=1}^{N_c} \sigma_i - \epsilon_{\mathrm{C}} \sum_{\langle ij \rangle} \sigma_i \sigma_j, \tag{7.29}$$

where $\sigma_i = \{0, 1\}$ for occupied and unoccupied lattice sites, respectively, and $\langle ij \rangle$ indicates that the sum is taken over nearest neighbors. The typical size and shape of a cargo domain is determined by the parameters of the lattice model. Fluctuations in the cargo and physically motivated choices of μ and ϵ_{C} are described in Section 7.5.

We must further specify the nature of the cargo-shell interaction. In biological systems, any attraction between the proteins in the cargo and the proteinaceous shell will likely be short-ranged protein-protein interactions. Additionally, steric repulsions ensure that the shell proteins cannot overlap with the cargo. To capture both features, we specify the cargo-shell

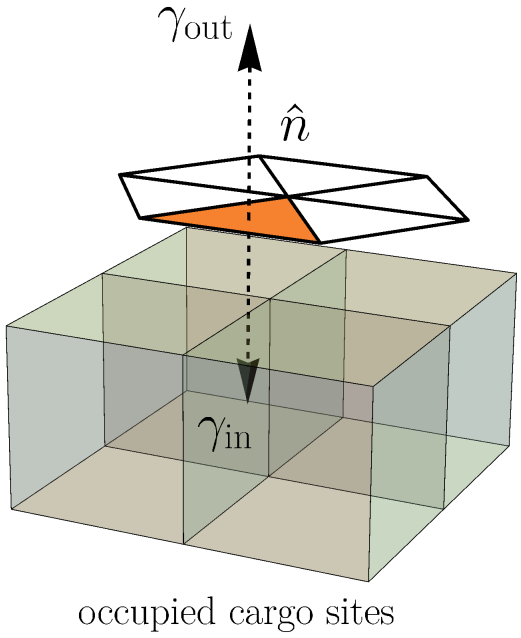

Figure 7.3: The cargo molecules, represented as a lattice gas, interact with the shell through an interaction that energetically favors the inward facing normal vectors with a factor $-\gamma_{\text{in}}$. The lattice site occupied by the outward facing normal vector of a monomer in the shell carries and energy γ_{out}, which serves to bias the shell fluctuations so that it is sterically occluded by the cargo.

interaction as follows,

$$\mathcal{H}_{\text{int}}(\sigma_i, v_j) = \begin{cases} -\gamma_{\text{in}} & \text{if } -\hat{n}_j \in i \text{ and } \sigma_i = 1 \\ +\gamma_{\text{out}} & \text{if } \hat{n}_j \in i \text{ or } v_j \in i \text{ and } \sigma_i = 1 \\ 0 & \text{otherwise.} \end{cases} \tag{7.30}$$

The full interaction is the sum over all i and j of H_{int}. This interaction is depicted schematically in Fig. 7.3. There is an energetic benefit if the inward facing normal vector to the shell is within a cargo-occupied lattice site. Conversely, there is an energetic penalty if the outward facing normal vector lies within an occupied site.

7.4 Dynamics of defect formation in nonequilibrium nanoshell growth

Many viruses and bacterial microcompartments are known to consist primarily, if not entirely, of protein capsomers with sixfold symmetry. These roughly hexagonal tiles cannot coat the surface of a sphere. The mathematical reason for this is the famous Gauss-Bonnet theorem, which relates the Euler characteristic of a manifold to its topological genus, i.e., the number of handles in the structure. Given a triangular tessellation of a topological surface, the number of vertices V, edges E, and faces F of the tessellation specify its Euler characteristic χ,

$$\chi = V - E + F. \tag{7.31}$$

According to the Gauss-Bonnet theorem, the Euler characteristic is a topological invariant, $\chi = 2 - 2g$, where g denotes the number of handles in a topological surface [217]. For a sphere, which has no handles, the Euler characteristic is $\chi_{\text{sphere}} = 2$. Thus, any triangulation of a sphere has the constraint that $V - E + F = 2$, a remarkable fact considering that this must also be true for any topological surface diffeomorphic to a sphere—arbitrary polyhedra, for example.

In the context that we are considering, we can use this very general result to determine the possible defect densities. First, we write the total number of faces in terms of the edges $F = \frac{2}{3}E$, since each edge is shared by two faces. We next note that the total number of edges can be written in terms of the coordination number of each vertex, $E = \frac{1}{2}\sum_{j=1}^{N} Z_j$, where Z_j denotes the number of bonds of vertex j. The requirement is thus [219],

$$\sum_{j=1}^{V} 6 - Z_j = 12. \tag{7.32}$$

This equation places a precise constraint on the total number of deviations away from sixfold bond order. In the context of the model we are considering, the energetic penalty for $Z_j < 5$

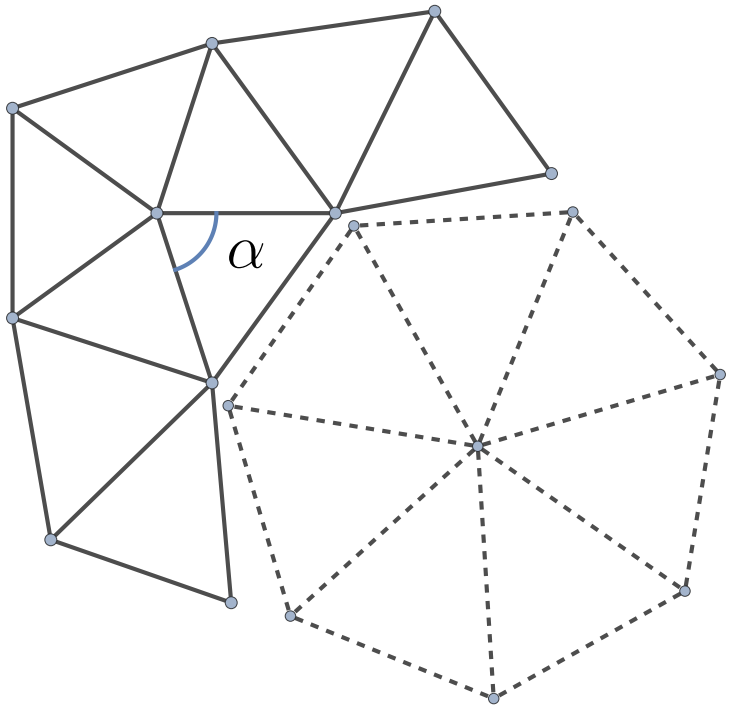

Figure 7.4: A fivefold defect distorts the hexagonal lattice by a factor of $+2\pi/6$, the angle associated with the missing monomer. This effect can be compensated by a sevenfold disclination that offsets the distortion with an additional monomer, which inserts material subtending an angle α. We associate with the sevenfold defect a topological charge $-2\pi/6$.

or $Z_j > 7$ is too extreme for these configurations to be observed. As a result, it must be the case that the difference between the number of fivefold and sevenfold defects is,

$$N_5 - N_7 = 12. \tag{7.33}$$

Determining the configuration of defects on a spherical surface that minimizes the energy is equivalent to a classic problem in quantum mechanics, the Thomson problem, which asks for minimum energy electron configurations on the surface of a sphere. This problem has proved difficult to analyze due to the large number of metastable configurations when the number of defects is large [220].

The fivefold and sevenfold topological defects balance one another when adjacent, as shown in Fig. 7.4. A fivefold defect can be introduced into a flat hexagonal lattice by removing a triangular face subtending an angle $s = +2\pi/6$. The angle s is often called the charge of the defect, because it attracts oppositely charged defects. Removing the wedge of material introduces strain into the lattice proportional to s, leading, in turn, to a divergence in the energy that grows as R^2, where R measures the radial extent of the sheet [218]. The

energetic penalty associated with the material removed can be compensated by adding a triangular face and creating a sevenfold defect. The addition of a sevenfold defect adjacent to the fivefold defect "screens" the charge by lowering the strain in the lattice. Bowick et al. [221] argues that, on a spherical surface, the expected number of dislocations pairs that screen an isolated defect in the thermodynamic ground state should grow as,

$$N_{\mathrm{d}} = \frac{\pi}{3}\left(\sqrt{11} - 5\arccos(5/6)\right) R_{\mathrm{c}}/l_0, \tag{7.34}$$

a prediction which is universal in the sense that it is independent of the microscopic interaction energy. The Coulomb like interaction between defects makes clear the connection to the Thomson problem: The zero temperature configurations of the defects on a spherical surface are the solutions of the minimum energy configurations of charges on a spherical shell [221].

Much of the continuum theory describing the shapes of low energy thin elastic sheets has focused on the case that there are exactly 12 five-coordinated topological defects, but some studies have looked at higher defect concentrations [213]. Proposed models of the carboxysome structure have emphasized potential protein capsomers with fivefold symmetry as the vertices of the icosahedral shell [198]. A few numerical and experimental studies have, on the other hand, shown the existence of linear strings of defects, called grain boundary scars that exist even for equilibrium configurations and in thermodynamic ground states [222–225]. This work characterizing grain boundary scars, suggests, a much richer interplay between the defects and geometry of the resulting shell. In the stochastic assembly process of bacterial microcompartment formation, this type of defect structure offers an alternative to the stringent requirement that exactly 12 pentameric capsomers are incorporated into the shell. Balancing these defects with oppositely charged disclinations offers a route to compensating additional fivefold defects via Eq. (7.33) and can lead to lower energy structures [221].

We simulated the model described in Section 7.1 to study the dynamics of defect formation in the assembly process. For these simulations, we represented the cargo of the microcompartment as a static, roughly spherical droplet of radius R_{c}. We parameterized the system in a regime where the activity z associated with the shell monomers was sufficiently low that insertions in the absence of a favorable cargo interaction were rare relative to the rate of removal. In order to ensure that the dynamics sampled distinct configurations for each inserted monomer, the dimensionless effective binding affinity K^* was chosen to be small enough relative to the activity z so that unbinding could occur with reasonable probability. Snapshots from a trajectory with $R_{\mathrm{c}} = 15l_0$ are shown in Fig. 7.5. All trajectories that we collected, $R_{\mathrm{c}} = 10l_0, 15l_0, 20l_0$ with ten independent runs at each radius, successfully encapsulated the cargo.

Fivefold defects spontaneously form due to the curvature of the cargo. The system we are studying is thermal, so fluctuations could significantly alter the defect configurations predicted from purely energetic considerations. The configurations we observe, nevertheless, are qualitatively in agreement with the structures observed in theoretical studies of the ground state. The defect structures, shown in Fig. 7.6, form local, grain boundary like chains of dislocation pair in regions of high curvature. The incorporation of additional topological

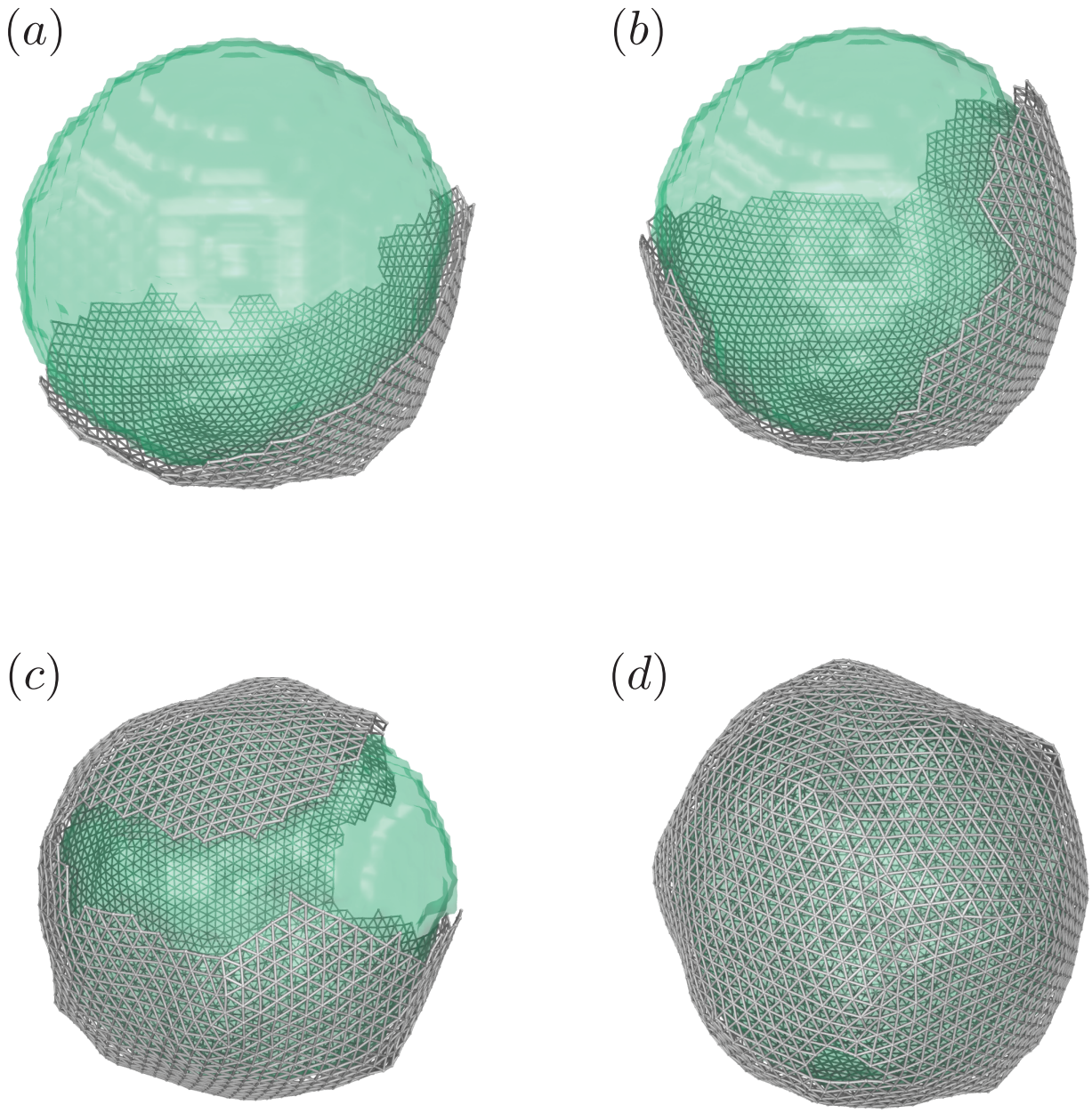

Figure 7.5: Cargo surface shown in transparent green. (a) $t = 250 \times 10^4$ MC sweeps. (b) $t = 500 \times 10^4$ MC sweeps. (c) $t = 750 \times 10^4$ MC sweeps. (d) $t = 1000 \times 10^4$ MC sweeps.

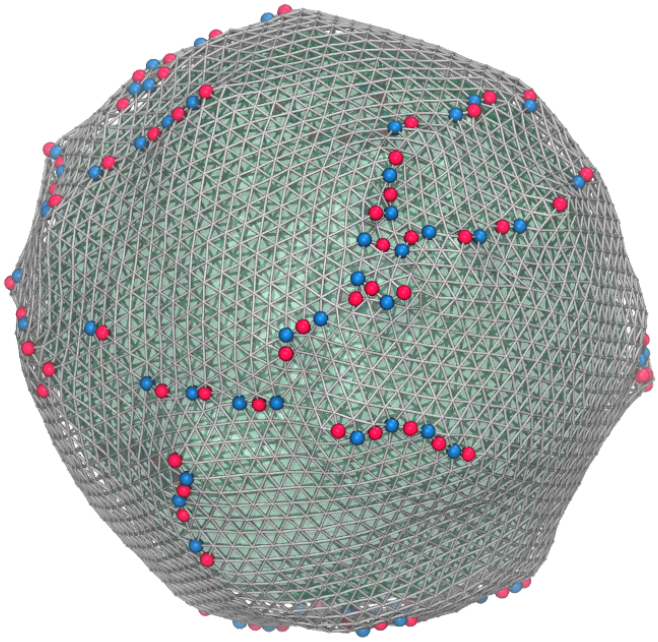

Figure 7.6: A configuration from an encapsulation trajectory around a static cargo with $R_c = 15$. Fivefold defects are with topological charge $+2\pi/6$ are shown in red. Sevenfold defects with topological charge $-2\pi/6$ are shown in blue.

defects suppresses the excess strain associated with deforming the thin elastic sheet around a disclination.

The dynamics of defect formation are shown in Fig. 7.7. In the conditions we considered, defect formation events appear to be uncorrelated at long times. This fact is apparent from the clear exponential time lag in Fig.7.7 (b). At short times, the dynamics may be more involved. The defect trajectories, Fig. 7.7 (a) show that many defects are removed shortly after being introduced into the system. A detailed description of the dynamics of defect formation will involve local geometric and energetic considerations.

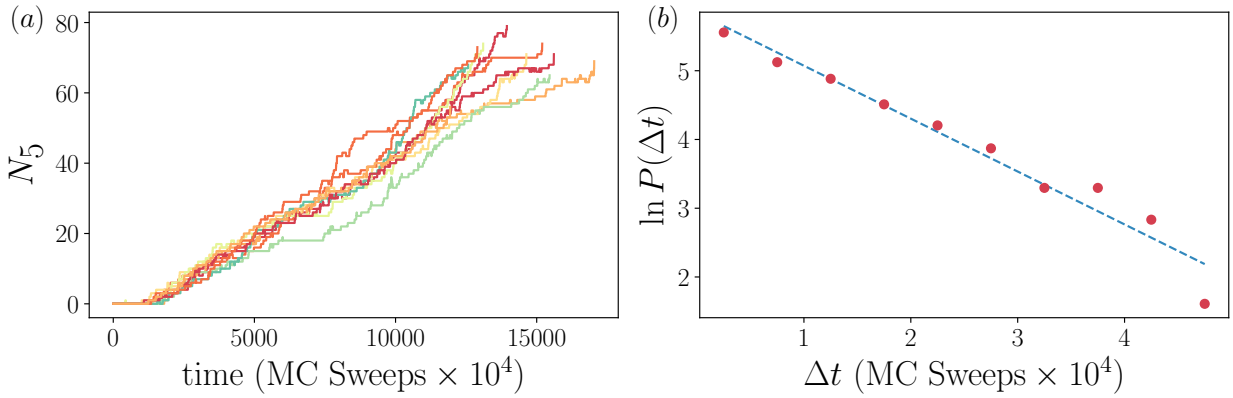

Figure 7.7: (*a*) Trajectories of the number of fivefold defects N_5 during the encapsulation of a static cargo with $R_c = 15$. (*b*) A histogram of waiting times between the formation of fivefold defects, the blue dashed line is a linear fit that clearly shows the exponential decay of waiting times consistent with a Poisson distribution.

7.5 Assembly around a fluctuating cargo

The spontaneous assembly of the carboxysome structure features dynamics of the cargo in addition to the shell. As mentioned in the introduction to this chapter, little is currently understood about the carboxysome self-assembly process from the biological point of view. The model we have developed suggests several simple physical principles that might govern the biogenesis of regularly sized nanoscale capsules.

In order for the size control to be robust, it should be the case that spontaneous nucleation of the cargo is rare. If the cargo can spontaneously nucleate and grow to large sizes, then it may deplete the local environment entirely of cargo material and the size will only be constrained by the stochastic density distribution of material in environment. Alternatively, spontaneous cargo nucleation could lead to many small structures with uncontrolled sizes. In the context of our model, depletion is not modeled, so the cargo would continue to grow without bound until entirely encapsulated. Because the bending energy of an icosahedral shell scales as $\ln(R_c^2)$, the free energetic contributions of shell-shell, cargo-cargo, and cargo-shell interactions dominate when the number of monomers is finite. As a result, we expect that microcompartments exist as nonequilibrium structures. The kinetics of encapsulation may compete with the kinetics of spontaneous cargo growth to achieve structures of a characteristic size. Here, we explore the conditions in which the resulting nonequilibrium structures have the faceted shape and narrow distribution of sizes observed in images of the carboxysome.

We also expect the growth of the shell to have slow nucleation dynamics in the absence of an interaction with the cargo. The cost of creating curvature in the assembling sheet

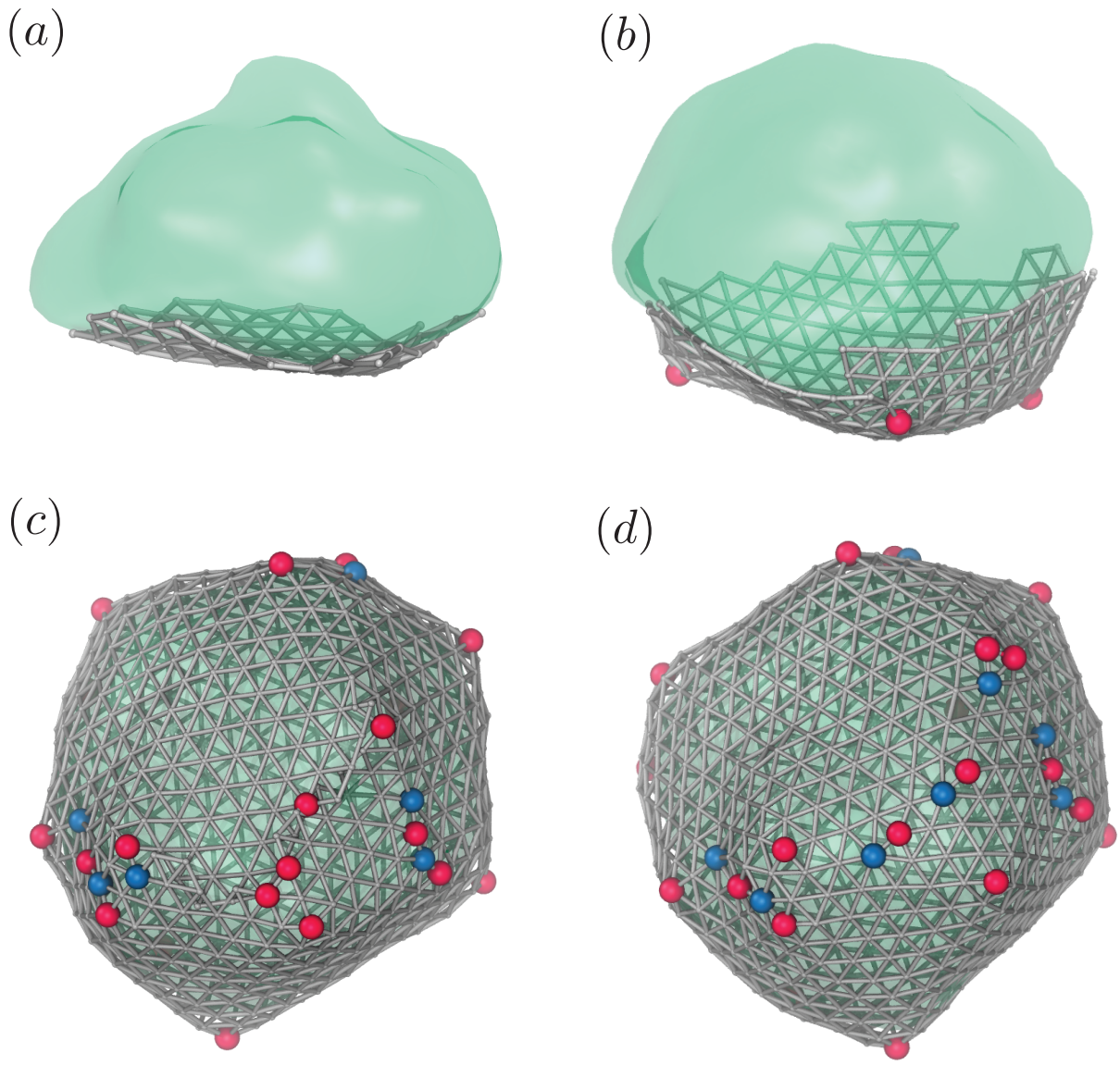

Figure 7.8: A trajectory demonstrating the cooperative nucleation of the cargo and the shell. Curvature in the shell is spontaneously generated by the surface tension of the cargo droplet on the initially flat shell. As defects are incorporated, the shell grows to entirely envelop the cargo. (*a*) t=50000 MC Sweeps (*b*) t = 150000 MC Sweeps (*c*) t=350000 MC Sweeps (*d*) t=400000 MC Sweeps.

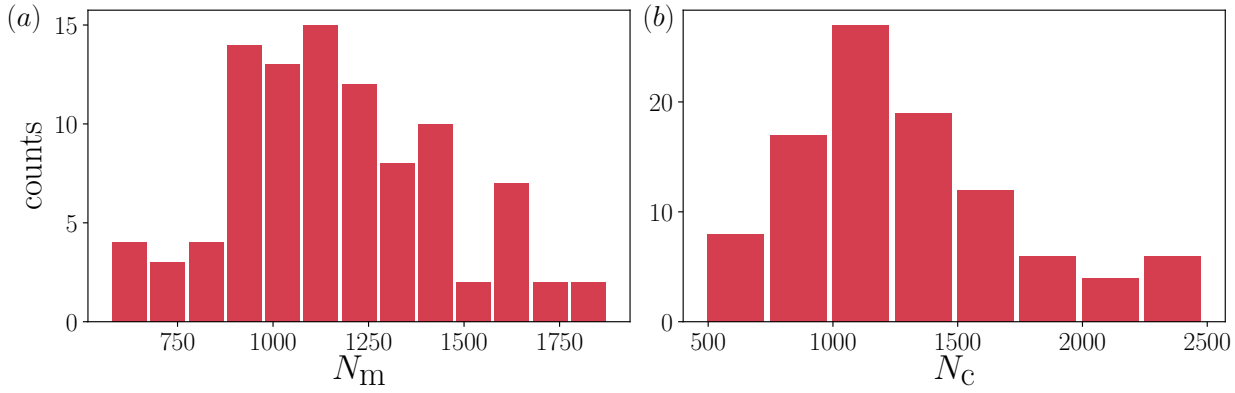

Figure 7.9: Histogram of (a) the number of monomers in the closed shell at the end of an assembly trajectory, and (b) the number of cargo molecules. These simulations were run with $\ln K = 10, \ln z = -50, \gamma = 40, \epsilon = 1200, \kappa = 40, \mu_{\mathrm{C}} = -16, \epsilon_{\mathrm{C}} = 4$.

is prohibitively high for spontaneous shell formation. This emphasizes that some structure must exist to template the curvature that leads to spontaneous defect formation. Empirical evidence that we have gathered suggests that cargo is a necessary condition for large, closed shells to form. Even in the case that we incorporate defects at a tunable rate, it is unlikely that the defects spontaneously organize so that the structure can meet itself to seal the shell, an observation emphasized by the model proposed in Ref. [216].

Cooperation between the shell and cargo may provide a strategy that achieves both control of the size and shape of the carboxysome structure as well as robust, high yield encapsulation. In the case that growth of the isolated components is rare, the nature of the interaction between the carboxysome and its cargo could dictate the growth dynamics. What is more, experimental evidence suggests that existing carboxysome structures may acts as the nucleation site for subsequent carboxysome formation [209].

We explored this putative mechanism for size and shape control by parameterizing our model in a regime where spontaneous growth of the isolated cargo or isolate shell was rare. In this regime, we found that the system relaxes to an enclosed structure with high probability. As shown in Fig. 7.8, the encapsulation proceeds by initially forming a droplet of cargo material on the shell proteins. The favorable interaction between the shell and cargo subsequently induce curvature in shell as it grows around the hemispherical droplet. Throughout the trajectory, defects form and anneal away as the shell relaxes. At the end of the trajectory, grain boundaries in the shell heal to close off the shell, cf. Fig 7.8 (c) and (d).

In addition to a high success rate, this mechanism of encapsulation yields structures that, like bacterial microcompartments, are relatively monodisperse. In Fig. 7.9 (a), we show a histogram of the number of monomers incorporated into the shell of the assembled structure for 100 independent assembly trajectories. The distribution is peaked near the average (1200 monomers) and the number of cargo, shown in Fig. 7.9 (b), has a similarly shaped

distribution. A tail of the distribution, indicative of rare, larger structures, is discernible from the statistics that we have gathered. However, there is not sufficient experimental data to confirm whether or not such a feature should be expected in the size distribution of carboxysome structures.

The potential applications of synthetic microcompartments are widespread and may have a significant technological impact, e.g., as a scalable carbon-fixation scheme. However, a sophisticated understanding of the necessary conditions for successful assembly of these nanoscale capsules could greatly enhance efforts to repurpose these biological entities. Our results provide a simple framework for characterizing the assembly process. We have demonstrated that encapsulation can be achieved when the propensity for nucleation of both the cargo and shell proteins is low. Our results also elucidate strategies for controlling the size of fully formed capsule by modulating the interaction strength.

Appendix A

Stochastic Calculus

Models of nonequilibrium dynamics often abstract the environment into a random spectrum of perturbations, represented as a stochastic process. Each realization of the dynamics samples a noise history from a probability distribution which, in general, depends on the length of the trajectory and the model used to represent the noise. In many applications, we want to collect trajectories from this probability distribution to compute distributions of dynamical observables. Mathematical ambiguities arise when computing expectations over the probability distribution of stochastic trajectories, so we must employ the tools of stochastic calculus to make such calculations rigorous.

We take, as a motivating example, the computation of the heat that flows from the system into the environment over the course of a trajectory. Let X_t denote the trajectory where $X_t : [0, T] \to \mathbb{R}^n$ defines the map from a time interval into the full phase space of the system. A generic equation of motion for the trajectory is of the Langevin form,

$$dX_t = (-\nabla V(X_t, t) + f_{\text{ext}}(X_t, t))\, dt + \sqrt{2\beta^{-1}} b(X_t, t) dW_t. \tag{A.1}$$

In this equation $V(x, t)$ is a time-dependent, conservative potential energy function, $f_{\text{ext}}(t)$ is a time-dependent, non-conservative external force, β is the inverse temperature, and $\frac{1}{2} bb^{\mathrm{T}} \equiv D$ is a space- and time-dependent diffusion tensor. The precise mathematical meaning of dW is discussed in the next section. The material presented here draws on a number of sources, primarily Refs. [226–229].

A.1 The Wiener process

For the noise in a Langevin equation, the random process $W(t)$ should satisfy a number of properties in order to be consistent with our physical expectations. The random forces from the environment, if they are frequent and of a magnitude smaller than the direct forces in the Langevin equation, will manifest as a Gaussian process due to the central limit theorem. We would therefore like $W(t)$ to be a Gaussian process with the following properties,

- $\langle W_t \rangle = 0, \forall t \geq 0$

- $W(t_1) - W(t_0) \sim \mathcal{N}(0, t_1 - t_0)$, for $t_1 > t_0 \geq 0$
- Increments $W(t_k) - W(t_{k-1})$ are statistically independent of distinct, disjoint increments.

These three properties characterize standard Brownian motion, also known as the Wiener Process. As should be expected from our knowledge of diffusion processes, the variance grows linearly in time, $\langle W_t^2 \rangle = t$.

The random forces from the fast degrees of freedom in fluctuating environment appear in the stochastic differential equation as a "derivative" of W_t with respect to time. We call the process dW_t a white noise process because it adds uncorrelated noise to the dynamics. Unfortunately, the derivative along a sample path of the Wiener process is pathological. Using the notion of Hölder continuity, it can be shown that paths drawn from W_t are nowhere differentiable [228]. This pathology leads to a number of important formal considerations when defining the solution to Eq. (A.1)

$$X_t = X_0 + \int -\nabla V(X_s, s) + f_{\text{ext}}(X_s, s)\ ds + \int b(X_s, s) dW_s. \tag{A.2}$$

The interpretation of the integral with respect to the white noise process has consequences for the solution X_t, i.e., distinct choices of the stochastic calculus employed lead to distinct physical systems. We can write the stochastic integral

$$\mathcal{I} = \int b(X_s, s) dW_s, \tag{A.3}$$

in terms of a Riemann sum to illustrate the choice that we have to make,

$$\mathcal{I} \approx \sum_{i=0}^{n-1} b(X_i, t_i)(W_{i+1} - W_i). \tag{A.4}$$

In the expression above, the function $b(X, t)$ is evaluated at time t_i. Taking the mean-squared limit $n \to \infty$ defines the Itō integral. Mathematically, the Itō conventions afford a number of advantageous properties because $b(X, t)$ is "non-anticipating" in the expression for $\mathcal{I}$.

An alternative commonly encountered in the physics literature is the following discretizaiton,

$$\mathcal{I}_{\text{S}} \approx \sum_{i=0}^{n-1} \frac{b(X_{i+1}, t_{i+1}) + b(X_i, t_i)}{2}(W_{i+1} - W_i). \tag{A.5}$$

In this expression, the integral defined as a Riemann sum that evaluates the function at the midpoint of the random displacement $W_{i+1} - W_i$. The subscript S is used to indicate that we are using the Stratonovich conventions for stochastic integration. There are a number of justifications for giving preference to the Stratonovich calculus for physical problems. In particular, it has been suggested that the Stratonovich approach more effectively models

the fact that random environmental effects have short, but non-zero, correlation times [230]. However, a pragmatic reason for the preference is illustrated in the following section—the Stratonovich conventions obey the same basic rules encountered in ordinary calculus.

For our purposes, the Stratonovich integral with respect to a stochastic process has the advantage of being time-symmetric. Physically measurable quantities like work and heat possess this inherent microscopic symmetry, so, for diffusions, we define the heat flow in terms of a Stratonovich product,

$$\mathcal{Q} = -\int_0^T \nabla V(X_t, t) \circ dX_t, \tag{A.6}$$

where we use the notation $\circ$ to emphasize that the integral must be interpreted using the Stratonovich calculus.

A.2 Itō and Stratonovich conventions

On average, the Itō integral vanishes for a non-anticipating function g,

$$\left\langle \int_0^t g(t')dW_{t'} \right\rangle = 0. \tag{A.7}$$

This is not the case for the Stratonovich calculus. The two conventions also differ when it comes to changes of variables. Let

$$dX_t = b(X_t, t)dW_t. \tag{A.8}$$

In the Stratonovich convention differentials transform according to the standard rules of calculus

$$dg = \frac{dg}{dx}\big(g(X_t, t)dt + b(X_t, t)dW_t\big). \tag{A.9}$$

Hence, there are no special considerations upon change of variables. On the other hand, Itō rules lead to a distinct expression for the differential,

$$dg = \frac{\partial g}{\partial t}dt + \frac{1}{2}\frac{\partial^2 g}{\partial t^2}dt + \cdots + \frac{\partial g}{\partial W}dW + \frac{\partial^2 g}{\partial W^2}dW^2 + \ldots \tag{A.10}$$

We retain only terms of order dt, which includes those proportional to dW and dW^2, with all other terms having higher order, and,

$$dg = \left(\frac{\partial g}{\partial t} + \frac{1}{2}\frac{\partial^2 g}{\partial W^2}\right)dt + \frac{\partial g}{\partial W}dW. \tag{A.11}$$

The additional term $\frac{1}{2}\frac{\partial^2 g}{\partial W^2}$ is sometimes called the Itō drift.

The additional "drift" appears in the Fokker-Planck equation associated with an Itō Langevin equation. The Fokker-Planck equation describes the evolution of the probability distribution $p(X_t, t|X_0, t_0)$. For an arbitrary function g,

$$\frac{d\langle g\rangle}{dt} = \int dX\ g(X)\frac{\partial p(X,t|X_0,t_0)}{\partial t}, \tag{A.12}$$

integrating by parts twice gives an expression for the Fokker-Planck equation,

$$\partial_t P(X,t|X_0,t_0) = -\frac{\partial}{\partial X}\left(g(X,t)\right)p(X,t|X_0,t_0) + \frac{1}{2}\frac{\partial^2}{\partial X^2}b(X,t)b(X,t)^{\mathrm{T}}p(X,t|X_0,t_0). \tag{A.13}$$

Note that this expression indicates that the statistics of the time evolution of an Itō Langevin equation are distinct from those of a Stratonovich equation.

A.3 Converting between Stratonovich and Itō

In many cases, it is mathematically preferable to work with the Itō calculus. However, the stochastic differential equations encountered in physics often require the Stratonovich interpretation. As discussed in Section A.1, the definitions of work and heat flow in a nonequilibrium diffusion process rely on the time-symmetric Stratonovich product. In order to ensure that the Itō process studied is statistically identical to the Stratonovich stochastic differential equation, we can simply add an additional drift term to the original Stratonovich stochastic differential equation in order to guarantee the same Fokker-Planck equation,

$$dX_t = \left(-\nabla V(X_t,t) + f_{\mathrm{ext}}(X_t,t) + \frac{1}{2}b(X_t,t)\partial_X b(X_t,t)\right) + b(X_t,t)dW_t. \tag{A.14}$$

It is easy to check that Itō equation above recovers the same Fokker-Planck equation as Eq. (A.1) under the Stratonovich interpretation. This conversion can be reversed symmetrically if one wants to convert from Itō to Stratonovich, as well.

A.4 Girsanov formula and change of measure

The Girsanov formula establishes the conditions under which two stochastic differential equations are equivalent, in the sense that the probability measures associated with paths arising from these stochastic processes have equivalent measures. A formal definition of this concept is given in the next section. Informally, we call two probability measures equivalent if any event with non-vanishing probability under one measure also has non-vanishing probability under the second measure. For the Wiener process, the probability of a trajectory can be written as a product over the Gaussian-distributed increments,

$$P(W_t) = \prod_{i=1}^{N}(\sqrt{2\pi\Delta t_i})^{-1/2}\exp\left(-\frac{\Delta W_i^2}{2\Delta t_i}\right). \tag{A.15}$$

The continuous representation of process gives a path integral formulation of the weight. Girsanov's Theorem says that, given two stochastic differential equations,

$$dX_t = f(t)dt + g(t)dW_t \tag{A.16}$$
$$dY_t = h(t)dt + k(t)dW_t \tag{A.17}$$

they are equivalent if $g(t) = k(t)$, that is, if they have the same diffusion tensors. From this perspective, we can view the latter equation as a stochastic differential equation with the drift of dX_t, but a different measure,

$$dY_t = f(t)dt + g(t)dV_t, \tag{A.18}$$

where the measure of the new process is,

$$P(V_t) = \exp\left(\int_0^T \frac{f(t) - a(t)}{b(t)} dV_t - \frac{1}{2}\left[\frac{f(t) - a(t)}{b(t)}\right]^2 dt\right) P(W_t). \tag{A.19}$$

The exponential factor can be viewed as a weight that expresses the likelihood of the noise history required to achieve the drift $f(t)$ under the dynamics of the process defined by dY_t. The change of measure is related to the Radon-Nikodym derivative, discussed in the next section.

A.5 Radon-Nikodym derivative

A formal treatment of the Girsanov theorem makes clear its relation to the Radon-Nikodym theorem. In order to state the Radon-Nikodym theorem, we require some language from probability theory that is not used elsewhere in this thesis, but can be found in, e.g., Øksendal [226]. We include these formal definitions to make explicit the meaning of Eq. (3.11). We let $(\Omega, \mathcal{F}, P)$ be a probability space, and let $\{\mathcal{F}\}_{t\geq 0}$ be a filtration on $(\Omega, \mathcal{F})$. We formalize the notion of equivalent measures introduced in the previous section by defining "absolute continuity". Let $T > 0$, then we say that a measure Q is absolutely continuous with respect to P if

$$P(F) = 0 \implies Q(F) = 0, \ \forall F \in \mathcal{F}_T. \tag{A.20}$$

The Radon-Nikodym theorem states that Q is absolutely continuous with respect P if and only if there exist an $\mathcal{F}_T$ measurable random variable $Z_T(\omega)$ such that,

$$dQ(\omega) = Z_T(\omega)dP(\omega). \tag{A.21}$$

This expression defines the Radon-Nikodym derivative, typically written,

$$\frac{dQ}{dP} = Z_T. \tag{A.22}$$

The re-weighting factor, Eq. (A.19) in the previous section is one example of a more general formulation of the Girsanov theorem in terms of Radon-Nikodym derivatives. These measures are particularly useful in the context of statistical physics because the appear in computations of the entropy production and of large deviation rate functions for nonequilibrium diffusions [27, 42, 75].

Bibliography

[1] F. Jülicher, A. Ajdari, and J. Prost, "Modeling molecular motors", Rev. Mod. Phys. **69**, 1269–1282 (1997).

[2] W. R. Browne and B. L. Feringa, "Making molecular machines work.", Nat. Nanotechnol. **1**, 25–35 (2006).

[3] S. Liu, G. Chistol, and C. Bustamante, "Mechanical Operation and Intersubunit Coordination of Ring-Shaped Molecular Motors: Insights from Single-Molecule Studies", Biophys. J. **106**, 1844–1858 (2014).

[4] T. D. Pollard and J. A. Cooper, "Actin, a central player in cell shape and movement.", Science **326**, 1208–1212 (2009).

[5] V. I. Risca, E. B. Wang, O. Chaudhuri, J. J. Chia, P. L. Geissler, and D. A. Fletcher, "Actin filament curvature biases branching direction.", Proc. Natl. Acad. Sci. U.S.A. **109**, 2913–2918 (2012).

[6] A. Mogilner and G. Oster, "Force generation by actin polymerization II: the elastic ratchet and tethered filaments.", Biophys. J **84**, 1591–1605 (2003).

[7] T. E. Schaus, E. W. Taylor, and G. G. Borisy, "Self-organization of actin filament orientation in the dendritic-nucleation/array-treadmilling model.", Proc. Natl. Acad. Sci. U.S.A. **104**, 7086–7091 (2007).

[8] P. A. Janmey, "Phosphoinositides and calcium as regulators of cellular actin assembly and disassembly.", Annu. Rev. Physiol. **56**, 169–191 (1994).

[9] A. Gopinathan, K.-C. Lee, J. M. Schwarz, and A. J. Liu, "Branching, Capping, and Severing in Dynamic Actin Structures", Phys. Rev. Lett. **99**, 058103–4 (2007).

[10] A. Mogilner and G. Oster, "Cell motility driven by actin polymerization", Biophys. J **71**, 3030–3045 (1996).

[11] N. Fakhri, A. D. Wessel, C. Willms, M. Pasquali, D. R. Klopfenstein, F. C. MacKintosh, and C. F. Schmidt, "High-resolution mapping of intracellular fluctuations using carbon nanotubes", Science **344**, 1031–1035 (2014).

[12] A. Hibbel, A. Bogdanova, M. Mahamdeh, A. Jannasch, M. Storch, E. Schäffer, D. Liakopoulos, and J. Howard, "Kinesin Kip2 enhances microtubule growth in vitrothrough length-dependent feedback on polymerization and catastrophe", eLife **4**, e10542 (2015).

[13] D. E. Smith, S. J. Tans, S. B. Smith, S. Grimes, D. L. Anderson, and C. Bustamante, "The bacteriophage [phis]29 portal motor can package DNA against a large internal force", Nature **413**, 748–752 (2001).

[14] A. Einstein, "On the movement of small particles suspended in stationary liquids required by the molecular-kinetic theory of heat", Ann. Phys. (Berlin) **17**, 549–560 (1905).

[15] N. G. Van Kampen, "Markov Processes", in *Stochastic processes in physics and chemistry* (2007), pp. 73–95.

[16] N. G. Van Kampen, "The Diffusion Type", in *Stochastic processes in physics and chemistry* (Elsevier, 2007), pp. 273–291.

[17] G. E. Crooks, "Nonequilibrium Measurements of Free Energy Differences for Microscopically Reversible Markovian Systems", J. Stat. Phys. **90**, 1481–1487 (1998).

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

[19] G. E. Crooks, "Path-ensemble averages in systems driven far from equilibrium", Phys. Rev. E **61**, 2361–2366 (2000).

[20] D. J. Evans and D. J. Searles, "Equilibrium microstates which generate second law violating steady states", Phys. Rev. E **50**, 1645 (1994).

[21] J. L. Lebowitz and H. Spohn, "A Gallavotti–Cohen-type symmetry in the large deviation functional for stochastic dynamics", J. Stat. Phys. **95**, 333–365 (1999).

[22] U. Seifert, "Entropy production along a stochastic trajectory and an integral fluctuation theorem", Phys. Rev. Lett. **95**, 040602 (2005).

[23] J. M. Horowitz and S. Vaikuntanathan, "Nonequilibrium detailed fluctuation theorem for repeated discrete feedback", Phys. Rev. E **82**, 061120–6 (2010).

[24] J. M. Horowitz, T. Sagawa, and J. M. Parrondo, "Imitating chemical motors with optimal information motors", Phys. Rev. Lett. **111**, 010602 (2013).

[25] T. Hatano and S.-i. Sasa, "Steady-State Thermodynamics of Langevin Systems", Phys. Rev. Lett. **86**, 3463–3466 (2001).

[26] T. Speck and U. Seifert, "Integral fluctuation theorem for the housekeeping heat", J. Phys. A: Math. Gen. **38**, L581–L588 (2005).

[27] R. Chetrite and K. Gawedzki, "Fluctuation relations for diffusion processes", Comm. Math. Phys. **282**, 469–518 (2008).

[28] R. M. Neal, "Annealed importance sampling", Stat. Comput. **11**, 125–139 (2001).

[29] C. Jarzynski, "Nonequilibrium Equality for Free Energy Differences", Phys. Rev. Lett. **78**, 2690–2693 (1997).

[30] D. Ruelle, *Thermodynamic Formalism*, 2nd ed., The Mathematical Structures of Equilibrium Statistical Mechanics (Cambridge University Press, Cambridge, 2009).

[31] V. Lecomte, C. Appert-Rolland, and F. van Wijland, "Thermodynamic Formalism for Systems with Markov Dynamics", J. Stat. Phys. **127**, 51–106 (2007).

[32] H. Touchette, "The large deviation approach to statistical mechanics", Phys. Rep. **478**, 1–69 (2009).

[33] J. Schnakenberg, "Network theory of microscopic and macroscopic behavior of master equation systems", Rev. Mod. Phys. **48**, 571–585 (1976).

[34] J. P. Garrahan, R. L. Jack, V. Lecomte, E. Pitard, K. van Duijvendijk, and F. van Wijland, "Dynamical first-order phase transition in kinetically constrained models of glasses", Phys. Rev. Lett. **98**, 195702 (2007).

[35] J. P. Garrahan, R. L. Jack, V. Lecomte, E. Pitard, K. van Duijvendijk, and F. van Wijland, "First-order dynamical phase transition in models of glasses: an approach based on ensembles of histories", J. Phys. A **42**, 075007 (2009).

[36] L. O. Hedges, R. L. Jack, J. P. Garrahan, and D. Chandler, "Dynamic Order-Disorder in Atomistic Models of Structural Glass Formers", Science **323**, 1309–1313 (2009).

[37] V. Lecomte, J. P. Garrahan, and F. van Wijland, "Inactive dynamical phase of a symmetric exclusion process on a ring", J. Phys. A: Math. Theor. **45**, 175001 (2012).

[38] A. Pal and D. A. Huse, "Many-body localization phase transition", Phys. Rev. B **82**, 174411 (2010).

[39] R. Nandkishore and D. A. Huse, "Many-Body Localization and Thermalization in Quantum Statistical Mechanics", Annu. Rev. Cond. Matt. Phys. **6**, 15–38 (2015).

[40] J. H. Bardarson, F. Pollmann, and J. E. Moore, "Unbounded Growth of Entanglement in Models of Many-Body Localization", Phys. Rev. Lett. **109**, 017202 (2012).

[41] D. Frenkel and B. Smit, "Rare Events", in *Understanding molecular simulation* (Elsevier, 2002), pp. 431–464.

[42] R. Chetrite and H. Touchette, "Nonequilibrium Markov Processes Conditioned on Large Deviations", Ann. Henri Poincaré **16**, 2005–2057 (2014).

[43] L. Onsager and S. Machlup, "Fluctuations and Irreversible Processes", Phys. Rev. **91**, 1505–1512 (1953).

[44] S. Machlup and L. Onsager, "Fluctuations and Irreversible Process. 2. Systems with Kinetic Energy", Phys. Rev. **91**, 1512–1515 (1953).

[45] R. Kubo, "The fluctuation-dissipation theorem", Rep. Prog. Phys. **29**, 255–284 (1966).

[46] R. Kubo, M. Toda, and N. Hashitsume, *Statistical Physics II: Nonequilibrium Statistical Mechanics*, Vol. 31 (Springer Science & Business Media, 2012).

[47] P. Salamon and R. S. Berry, "Thermodynamic length and dissipated availability", Phys. Rev. Lett. **51**, 1127–1130 (1983).

[48] C. H. Bennett, "The Thermodynamics of Computation—A Review", Int. J. Theor. Phys. **21**, 905–940 (1982).

[49] T. D. Schindler, L. Chen, P. Lebel, M. Nakamura, and Z. Bryant, "Engineering myosins for long-range transport on actin filaments.", Nat. Nanotechnol. **9**, 33–38 (2014).

[50] G. Lan, P. Sartori, S. Neumann, V. Sourjik, and Y. Tu, "The energy-speed-accuracy trade-off in sensory adaptation", Nat. Phys. **8**, 422–428 (2012).

[51] P. T. X. Li, C. Bustamante, and I. Tinoco, "Real-time control of the energy landscape by force directs the folding of RNA molecules.", Proc. Natl. Acad. Sci. U.S.A. **104**, 7039–7044 (2007).

[52] C. Jarzynski, "Equalities and inequalities: irreversibility and the second law of thermodynamics at the nanoscale", Annu. Rev. Condens. Matter Phys. **2**, 329–351 (2011).

[53] M. Esposito and C. Van den Broeck, "Three faces of the second law. I. Master equation formulation", Phys. Rev. E **82**, 011143 (2010).

[54] C. Van den Broeck and M. Esposito, "Three faces of the second law. II. Fokker-Planck formulation", Phys. Rev. E **82**, 011144 (2010).

[55] T. R. Gingrich, G. M. Rotskoff, S. Vaikuntanathan, and P. L. Geissler, "Efficiency and large deviations in time-asymmetric stochastic heat engines", New J. Phys. **16**, 102003 (2014).

[56] S. Carnot, *Reflections on the Motive Power of Fire* (Chez Bachelier, Paris, 1824).

[57] G. Verley, M. Esposito, T. Willaert, and C. Van den Broeck, "The unlikely Carnot efficiency", Nat. Commun. **5**, 4721 (2014).

[58] H. B. Callen, *Thermodynamics and an Introduction to Thermostatistics* (John Wiley & Sons, 1960).

[59] V. Blickle and C. Bechinger, "Realization of a micrometre-sized stochastic heat engine", Nat. Phys. **8**, 143–146 (2012).

[60] H. Touchette and R. J. Harris, *Large Deviation Approach to Nonequilibrium Systems*, Fluctuation Relations and Beyond (Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, 2013).

[61] U. Seifert, "Stochastic thermodynamics, fluctuation theorems and molecular machines", Rep. Prog. Phys. **75**, 126001 (2012).

[62] T. R. Gingrich, G. M. Rotskoff, and J. M. Horowitz, "Inferring dissipation from current fluctuations", J. Phys. A: Math. Theor. **50**, 184004 (2017).

[63] J. G. Kirkwood, "The statistical mechanics theory of transport processes I. General Theory", J. Chem. Phys. **14**, 180–201 (1946).

[64] J. Weber, "Fluctuation Dissipation Theorem", Phys. Rev. **101**, 1620–1626 (1956).

[65] T. Speck and U. Seifert, "Restoring a fluctuation-dissipation theorem in a nonequilibrium steady state", Euro. Phys. Lett. **74**, 391–396 (2006).

[66] B. Cleuren, C. Van den Broeck, and R. Kawai, "Fluctuation and dissipation", C. R. Physique **8**, 567–578 (2007).

[67] C. Maes and K. Netočny, "Canonical structure of dynamical fluctuations in mesoscopic nonequilibrium steady states", Euro. Phys. Lett. **82**, 30003 (2008).

[68] M. Baiesi, C. Maes, and B. Wynants, "Fluctuations and Response of Nonequilibrium States", Phys. Rev. Lett. **103**, 010602 (2009).

[69] T. Speck and U. Seifert, "Extended fluctuation-dissipation theorem for soft matter in stationary flow", Phys. Rev. E **79**, 040102 (2009).

[70] C. Maes, "Fluctuations and response out-of-equilibrium", Prog. Theor. Phys. Supplement **184**, 318–28 (2010).

[71] U. Seifert and T. Speck, "Fluctuation-dissipation theorem in nonequilibrium steady states", Euro. Phys. Lett. **89**, 10007 (2010).

[72] G. M. Rotskoff, "Mapping current fluctuations of stochastic pumps to nonequilibrium steady states.", Phys. Rev. E **95**, 030101 (2017).

[73] L. Bertini, A. Faggionato, and D. Gabrielli, "Flows, currents, and cycles for Markov chains: Large deviation asymptotics", Stoch. Proc. Appl. **125**, 2786–2819 (2015).

[74] L. Bertini, A. Faggionato, and D. Gabrielli, "Large deviations of the empirical flow for continuous time Markov chains", Ann. Inst. H. Poincaré Probab. Statist. **51**, 867–900 (2015).

[75] A. C. Barato and R. Chetrite, "A Formal View on Level 2.5 Large Deviations and Fluctuation Relations", J. Stat. Phys. **160**, 1154–1172 (2015).

[76] M. D. Donsker and S. R. S. Varadhan, "Asymptotic evaluation of certain markov process expectations for large time. IV", Comm. Pure Appl. Math. **36**, 183–212 (1983).

[77] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, Vol. 38, Stochastic Modelling and Applied Probability (Springer Berlin Heidelberg, Berlin, Heidelberg, 2010).

[78] R. S. Ellis, *Entropy, Large Deviations, and Statistical Mechanics* (Springer Berlin Heidelberg, 2006).

[79] L. Bertini, A. De Sole, D. Gabrielli, G. Jona-Lasinio, and C. Landim, "Macroscopic fluctuation theory", Rev. Mod. Phys. **87**, 593–636 (2015).

[80] A. C. Barato and U. Seifert, "Thermodynamic Uncertainty Relation for Biomolecular Processes", Phys. Rev. Lett. **114**, 158101 (2015).

[81] A. C. Barato and U. Seifert, "Universal Bound on the Fano Factor in Enzyme Kinetics", J Phys. Chem. B **119**, 6555–6561 (2015).

[82] P. Pietzonka, A. C. Barato, and U. Seifert, "Universal bounds on current fluctuations", Phys. Rev. E **93**, 052145 (2016).

[83] P. Pietzonka, A. C. Barato, and U. Seifert, "Affinity- and topology-dependent bound on current fluctuations", J. Phys. A **49**, 34LT01 (2016).

[84] M. Polettini, A. Lazarescu, and M. Esposito, "Tightening the uncertainty principle for stochastic currents.", Phys. Rev. E **94**, 052104 (2016).

[85] T. R. Gingrich, J. M. Horowitz, N. Perunov, and J. L. England, "Dissipation Bounds All Steady-State Current Fluctuations", Phys. Rev. Lett. **116**, 120601 (2016).

[86] P. Pietzonka, F. Ritort, and U. Seifert, "Finite-time generalization of the thermodynamic uncertainty relation", arXiv:1702.07699v1 (2017).

[87] S. B. Smith, Y. Cui, and C. Bustamante, "Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules.", Science **271**, 795–799 (1996).

[88] D. Collin, F. Ritort, C. Jarzynski, S. B. Smith, I. Tinoco Jr, and C. Bustamante, "Verification of the Crooks fluctuation theorem and recovery of RNA folding free energies", Nature **437**, 231–234 (2005).

[89] G. Hummer and A. Szabo, "Free energy reconstruction from nonequilibrium single-molecule pulling experiments", Proc. Natl. Acad. Sci. U.S.A. **98**, 3658–3661 (2001).

[90] P. Talkner, "Stochastic resonance in the semiadiabatic limit", New J. Phys. **1**, 4.1–4.25 (1999).

[91] D. Mandal and C. Jarzynski, "A proof by graphical construction of the no-pumping theorem of stochastic pumps", J. Stat. Mech. Theor. Exp. **2011**, P10006 (2011).

[92] S. Rahav, J. Horowitz, and C. Jarzynski, "Directed Flow in Nonadiabatic Stochastic Pumps", Phys. Rev. Lett. **101**, 140602 (2008).

[93] V. Y. Chernyak and N. A. Sinitsyn, "Pumping restriction theorem for stochastic networks.", Phys. Rev. Lett. **101**, 160601 (2008).

[94] N. A. Sinitsyn and I. Nemenman, "Universal Geometric Theory of Mesoscopic Stochastic Pumps and Reversible Ratchets", Phys. Rev. Lett. **99**, 220408 (2007).

[95] M. Yoshida, E. Muneyuki, and T. Hisabori, "ATP synthase—a marvellous rotary engine of the cell", Nat. Rev. Mol. Cell Biol. **2**, 669–677 (2001).

[96] M. R. Wilson, J. Solà, A. Carlone, S. M. Goldup, N. Lebrasseur, and D. A. Leigh, "An autonomous chemically fuelled small-molecule motor", Nature **534**, 235–240 (2016).

[97] W. R. Browne and B. L. Feringa, "Making molecular machines work", Nat. Nanotechnology **1**, 25–35 (2006).

[98] J. M. Horowitz and C. Jarzynski, "Exact Formula for Currents in Strongly Pumped Diffusive Systems", J. Stat. Phys. **136**, 917–925 (2009).

[99] N. A. Sinitsyn, "The stochastic pump effect and geometric phases in dissipative and stochastic systems", J. Phys. A **42**, 193001 (2009).

[100] I. A. Martínez, É. Roldán, L. Dinis, D. Petrov, J. M. R. Parrondo, and R. A. Rica, "Brownian Carnot engine", Nat. Phys. **12**, 67–70 (2015).

[101] O. Raz, Y. Subaşı, and C. Jarzynski, "Mimicking Nonequilibrium Steady States with Time-Periodic Driving", Phys. Rev. X **6**, 021022 (2016).

[102] M. Polettini, G. Verley, and M. Esposito, "Efficiency Statistics at All Times: Carnot Limit at Finite Power", Phys. Rev. Lett. **114**, 050601 (2015).

[103] K. Proesmans, Y. Dreher, M. Gavrilov, J. Bechhoefer, and C. Van den Broeck, "Brownian Duet: A Novel Tale of Thermodynamic Efficiency", Phys. Rev. X **6**, 041010 (2016).

[104] R. Zia and B. Schmittmann, "Probability currents as principal characteristics in the statistical mechanics of non-equilibrium steady states", J. Stat. Mech. Theor. Exp. **2007**, P07012 (2007).

[105] Y. Oono and M. Paniconi, "Steady state thermodynamics", Prog. Theor. Phys. Supplement **130**, 29–44 (1998).

[106] V. Lecomte and J. Tailleur, "A numerical approach to large deviations in continuous time", J. Stat. Mech. Theor. Exp. **2007**, P03004–P03004 (2007).

[107] S. Erbas-Cakmak, D. A. Leigh, C. T. McTernan, and A. L. Nussbaumer, "Artificial Molecular Machines", Chem. Rev. **115**, 10081–10206 (2015).

[108] A. B. Bortz, M. H. Kalos, and J. L. Lebowitz, "A new algorithm for Monte Carlo simulation of Ising spin systems", J. Comput. Phys. **17**, 10–18 (1975).

[109] A. F. Voter, "Introduction to the kinetic Monte Carlo method", in *Radiation effects in solids* (Springer, Netherlands, 2007), pp. 1–23.

[110] A. Prados, J. J. Brey, and B. Sánchez-Rey, "A dynamical monte carlo algorithm for master equations with time-dependent transition rates", J. Stat. Phys. **89**, 709–734 (1997).

[111] T. M. Cover and J. A. Thomas, *Elements of information theory*, Second (A. John Wiley & Sons, Hoboken, NJ, 2006).

[112] N. A. Sinitsyn and I. Nemenman, "The Berry phase and the pump flux in stochastic chemical kinetics", Euro. Phys. Lett. **77**, 58001 (2007).

[113] A. C. Barato and U. Seifert, "Cost and Precision of Brownian Clocks", Phys. Rev. X **6**, 041053 (2016).

[114] F. Weinhold, "Metric geometry of equilibrium thermodynamics", J. Chem. Phys. **63**, 2479–2483 (1975).

[115] G. Ruppeiner, "Thermodynamics: A Riemannian geometric model", Phys. Rev. A **20**, 1608–1613 (1979).

[116] F. Schlögl, "Thermodynamic metric and stochastic measures", Z Phys. B – Condensed Matter **59**, 449–454 (1985).

[117] G. E. Crooks, "Measuring Thermodynamic Length", Phys. Rev. Lett. **99**, 100602 (2007).

[118] D. A. Sivak and G. E. Crooks, "Thermodynamic metrics and optimal paths", Phys. Rev. Lett. **108**, 190602 (2012).

[119] P. R. Zulkowski, D. A. Sivak, G. E. Crooks, and M. R. DeWeese, "Geometry of thermodynamic control", Phys. Rev. E **86**, 041148 (2012).

[120] K. Sekimoto, "Langevin Equation and Thermodynamics", Prog. Theor. Phys. Supplement **130**, 17–27 (1998).

[121] M. Heymann and E. Vanden-Eijnden, "The geometric minimum action method: A least action principle on the space of curves", Comm. Pure Appl. Math. **61**, 1052–1117 (2008).

[122] M. Heymann and E. Vanden-Eijnden, "Pathways of Maximum Likelihood for Rare Events in Nonequilibrium Systems: Application to Nucleation in the Presence of Shear", Phys. Rev. Lett. **100**, 140601 (2008).

[123] H. J. Kappen, "Optimal control theory and the linear Bellman equation", in *Bayesian time series models*, edited by D. Barber, A. T. Cemgil, and S. Chiappa (Cambridge University Press, Cambridge, 2009), pp. 363–387.

[124] H. J. Kappen, "Path integrals and symmetry breaking for optimal control theory", J. Stat. Mech. Theor. Exp. **2005**, P11011–P11011 (2005).

[125] H. J. Kappen, "Linear Theory for Control of Nonlinear Stochastic Systems", Phys. Rev. Lett. **95**, 200201 (2005).

[126] T. Schmiedl and U. Seifert, "Stochastic thermodynamics of chemical reaction networks", J. Chem. Phys. **126**, 044101 (2007).

[127] M. V. S. Bonança and S. Deffner, "Optimal driving of isothermal processes close to equilibrium", J. Chem. Phys. **140**, 244119 (2014).

[128] P. Muratore-Ginanneschi, "On the use of stochastic differential geometry for non-equilibrium thermodynamic modeling and control", J. Phys. A: Math. Theor. **46**, 275002 (2013).

[129] P. R. Zulkowski and M. R. DeWeese, "Optimal control of overdamped systems", Phys. Rev. E **92**, 032117 (2015).

[130] J. E. Hunter III, W. P. Reinhardt, and T. F. Davis, "A finite-time variational method for determining optimal paths and obtaining bounds on free energy changes from computer simulations", J. Chem. Phys. **99**, 6856–6864 (1993).

[131] D. K. Shenfeld, H. Xu, M. P. Eastwood, R. O. Dror, and D. E. Shaw, "Minimizing thermodynamic length to select intermediate states for free-energy calculations and replica-exchange simulations", Phys. Rev. E **80**, 046705 (2009).

[132] D. Chandler, *Introduction to Modern Statistical Mechanics* (Oxford University Press, Oxford, 1987).

[133] A. R. Honerkamp-Smith, B. B. Machta, and S. L. Keller, "Experimental Observations of Dynamic Critical Phenomena in a Lipid Membrane", Phys. Rev. Lett. **108**, 265702 (2012).

[134] M. H. Kryder, E. C. Gage, T. W. McDaniel, W. A. Challener, R. E. Rottmayer, G. Ju, Y.-T. Hsia, and M. F. Erden, "Heat Assisted Magnetic Recording", Proc. IEEE **96**, 1810–1835 (2008).

[135] K. Roy, S. Bandyopadhyay, and J. Atulasimha, "Hybrid spintronics and straintronics: A magnetic technology for ultra low energy computing and signal processing", Appl. Phys. Lett. **99**, 063108 (2011).

[136] A. Berezhkovskii and A. Szabo, "Time scale separation leads to position-dependent diffusion along a slow coordinate", J. Chem. Phys. **135**, 074108–6 (2011).

[137] R. J. Glauber, "Time-Dependent Statistics of the Ising Model", J. Math. Phys. **4**, 294–307 (2004).

[138] J. A. Sethian, "Fast Marching Methods", SIAM Rev. **41**, 199–235 (2006).

[139] J. Cardy, *Scaling and Renormalization in Statistical Physics* (Cambridge University Press, Cambridge, 2009).

[140] L. Xu, P. Kumar, S. V. Buldyrev, S. H. Chen, P. H. Poole, F. Sciortino, and H. E. Stanley, "Relation between the Widom line and the dynamic crossover in systems with a liquid-liquid phase transition.", Proc. Natl. Acad. Sci. U.S.A. **102**, 16558–16562 (2005).

[141] W. E, W. Ren, and E. Vanden-Eijnden, "String method for the study of rare events", Phys. Rev. B **66**, 052301 (2002).

[142] L. Maragliano, A. Fischer, E. Vanden-Eijnden, and G. Ciccotti, "String method in collective variables: Minimum free energy paths and isocommittor surfaces", J. Chem. Phys. **125**, 024106 (2006).

[143] L. Maragliano, B. Roux, and E. Vanden-Eijnden, "Comparison between Mean Forces and Swarms-of-Trajectories String Methods", J. Chem. Theory Comput. **10**, 524–533 (2014).

[144] P. G. Bolhuis, D. Chandler, C. Dellago, and P. L. Geissler, "Transition path sampling: throwing ropes over rough mountain passes, in the dark.", Annu. Rev. Phys. Chem. **53**, 291–318 (2002).

[145] X. Bohigas, E. Molins, A. Roig, J. Tejada, and X. X. Zhang, "Room-temperature magnetic refrigerator using permanent magnets", IEEE Trans. Magn. **36**, 538–544 (2000).

[146] K. H. Kim and H. Qian, "Fluctuation theorems for a molecular refrigerator", Phys. Rev. E **75**, 022102 (2007).

[147] R. Kimmel and J. A. Sethian, "Computing geodesic paths on manifolds.", Proc. Natl. Acad. Sci. U.S.A. **95**, 8431–8435 (1998).

[148] W. E, W. Ren, and E. Vanden-Eijnden, "Minimum action method for the study of rare events", Comm. Pure Appl. Math. **57**, 637–656 (2004).

[149] R. Landauer, "Irreversibility and Heat Generation in the Computing Process", IBM J. Res. Dev. **5**, 183–191 (1961).

[150] H. Ohno, D. Chiba, F. Matsukura, T. Omiya, E. Abe, T. Dietl, Y. Ohno, and K. Ohtani, "Electric-field control of ferromagnetism.", Nature **408**, 944–946 (2000).

[151] T. Lottermoser, T. Lonkai, U. Amann, D. Hohlwein, J. Ihringer, and M. Fiebig, "Magnetic phase control by an electric field", Nature **430**, 541–544 (2004).

[152] B. Lambson, D. Carlton, and J. Bokor, "Exploring the Thermodynamic Limits of Computation in Integrated Systems: Magnetic Memory, Nanomagnetic Logic, and the Landauer Limit", Phys. Rev. Lett. **107**, 010604 (2011).

[153] J. T. Heron, M. Trassin, K. Ashraf, M. Gajek, Q. He, S. Y. Yang, D. E. Nikonov, Y.-H. Chu, S. Salahuddin, and R. Ramesh, "Electric-Field-Induced Magnetization Reversal in a Ferromagnet-Multiferroic Heterostructure", Phys. Rev. Lett. **107**, 217202–217205 (2011).

[154] S. Joo, T. Kim, S. H. Shin, J. Y. Lim, J. Hong, J. D. Song, J. Chang, H.-W. Lee, K. Rhie, S. H. Han, K.-H. Shin, and M. Johnson, "Magnetic-field-controlled reconfigurable semiconductor logic", Nature **494**, 72–76 (2013).

[155] J. T. Heron, J. L. Bosse, Q. He, Y. Gao, M. Trassin, L. Ye, J. D. Clarkson, C. Wang, J. Liu, S. Salahuddin, D. C. Ralph, D. G. Schlom, J. Íñiguez, B. D. Huey, and R. Ramesh, "Deterministic switching of ferromagnetism at room temperature using an electric field", Nature **516**, 370–373 (2014).

[156] L. You, O. Lee, D. Bhowmik, D. Labanowski, J. Hong, J. Bokor, and S. Salahuddin, "Switching of perpendicularly polarized nanomagnets with spin orbit torque without an external magnetic field by engineering a tilted anisotropy", Proc. Natl. Acad. Sci. U.S.A. **112**, 10310–10315 (2015).

[157] J. Hong, B. Lambson, S. Dhuey, and J. Bokor, "Experimental test of Landauer's principle in single-bit operations on nanomagnetic memory bits", Sci. Adv. **2**, e1501492–e1501492 (2016).

[158] A. L. Pankratov, S. N. Vdovichev, and I. M. Nefedov, "Effect of noise on the high-speed reversal of single-domain uniaxial magnetic nanoparticles", Phys. Rev. B **78**, 052401 (2008).

[159] A. A. Smirnov and A. L. Pankratov, "Influence of the size of uniaxial magnetic nanoparticle on the reliability of high-speed switching", Phys. Rev. B **82**, 132405 (2010).

[160] N. Shiraishi, K. Saito, and H. Tasaki, "Universal Trade-Off Relation between Power and Efficiency for Heat Engines", Phys. Rev. Lett. **117**, 190601 (2016).

[161] D. E. Nikonov, G. I. Bourianoff, and P. A. Gargini, "Power Dissipation in Spintronic Devices Out of Thermodynamic Equilibrium", J. Supercond. Nov. Magn. **19**, 497–513 (2006).

[162] S. Manipatruni, D. E. Nikonov, and I. A. Young, "All-spin nanomagnetic state elements", Appl. Phys. Lett. **103**, 063503 (2013).

[163] D. E. Nikonov, S. Manipatruni, and I. A. Young, "Automotion of domain walls for spintronic interconnects", J. Appl. Phys. **115**, 213902 (2014).

[164] G. M. Rotskoff and G. E. Crooks, "Optimal control in nonequilibrium systems: Dynamic Riemannian geometry of the Ising model", Phys. Rev. E **92**, 060102 (2015).

[165] T. R. Gingrich, G. M. Rotskoff, G. E. Crooks, and P. L. Geissler, "Near-optimal protocols in complex nonequilibrium transformations.", Proc. Natl. Acad. Sci. U.S.A. **113**, 10263–10268 (2016).

[166] W. Zhang, H. Wang, C. Hartmann, M. Weber, and C. Schütte, "Applications of the Cross-Entropy Method to Importance Sampling and Optimal Control of Diffusions", SIAM J. Sci. Comput. **36**, A2654–A2672 (2014).

[167] H. J. Kappen and H. C. Ruiz, "Adaptive Importance Sampling for Control and Inference", J. Stat. Phys. **162**, 1244–1266 (2016).

[168] P. Salamon, A. Nitzan, B. Andresen, and R. S. Berry, "Minimum entropy production and the optimization of heat engines", Phys. Rev. A **21**, 2115–2129 (1980).

[169] S. Thijssen and H. J. Kappen, "Path integral control and state-dependent feedback", Phys. Rev. E **91**, 032104 (2015).

[170] P. Dupuis and H. Wang, "Importance Sampling, Large Deviations, and Differential Games", Stochastics **76**, 481–508 (2004).

[171] R. Chetrite and H. Touchette, "Variational and optimal control representations of conditioned and driven processes", J. Stat. Mech. Theor. Exp. **2015**, P12001 (2015).

[172] P. Maragakis, F. Ritort, C. Bustamante, M. Karplus, and G. E. Crooks, "Bayesian estimates of free energies from nonequilibrium work data in the presence of instrument noise", J. Chem. Phys. **129**, 024102 (2008).

[173] T. L. Gilbert, "A phenomenological theory of damping in ferromagnetic materials", IEEE Trans. Magn. **40**, 3443–3449 (2004).

[174] R. V. Kohn, M. G. Reznikoff, and E. Vanden-Eijnden, "Magnetic Elements at Finite Temperature and Large Deviation Theory", J. Nonlinear Sci. **15**, 223–253 (2005).

[175] M. Venturoli, E. Vanden-Eijnden, and G. Ciccotti, "Kinetics of phase transitions in two dimensional Ising models studied with the string method", J. Math. Chem. **45**, 188–222 (2009).

[176] T. R. Gingrich and P. L. Geissler, "Preserving correlations between trajectories for efficient path sampling.", J. Chem. Phys. **142**, 234104 (2015).

[177] W. E, W. Ren, and E. Vanden-Eijnden, "Transition pathways in complex systems: Reaction coordinates, isocommittor surfaces, and transition tubes", Chem. Phys. Lett. **413**, 242–247 (2005).

[178] M. F. H. Jason D Perlmutter, "Mechanisms of Virus Assembly", Annu. Rev. Phys. Chem. **66**, 217–239 (2015).

[179] W. M. Jacobs and D. Frenkel, "Self-Assembly of Structures with Addressable Complexity", J. Am. Chem. Soc., 2457–2467 (2016).

[180] S. Whitelam and R. L. Jack, "The Statistical Mechanics of Dynamic Pathways to Self-Assembly", Annu. Rev. Phys. Chem. **66**, 143–163 (2015).

[181] B. Wei, M. Dai, and P. Yin, "Complex shapes self-assembled from single-stranded DNA tiles", Nature **485**, 623–626 (2012).

[182] Y. Ke, L. L. Ong, W. M. Shih, and P. Yin, "Three-Dimensional Structures Self-Assembled from DNA Bricks", Science **338**, 1177–1183 (2012).

[183] A. Murugan, J. Zou, and M. P. Brenner, "Undesired usage and the robust self-assembly of heterogeneous structures", Nat. Commun. **6**, 6203–10 (2015).

[184] N. B. Schade, M. C. Holmes-Cerfon, E. R. Chen, D. Aronzon, J. W. Collins, J. A. Fan, F. Capasso, and V. N. Manoharan, "Tetrahedral Colloidal Clusters from Random Parking of Bidisperse Spheres", Phys. Rev. Lett. **110**, 148303 (2013).

[185] L. O. Hedges, R. V. Mannige, and S. Whitelam, "Growth of equilibrium structures built from a large number of distinct component types", Soft Matter **10**, 6404–6416 (2014).

[186] M. Bowick and P. Chaikin, "Colloidal crystals: Stresses come to light", Nat. Mater. **15**, 1151–1152 (2016).

[187] Y. Wang, Y. Wang, X. Zheng, É. Ducrot, J. S. Yodh, M. Weck, and D. J. Pine, "Crystallization of DNA-coated colloids", Nat. Commun. **6**, 7253 (2015).

[188] M. F. Hagan, "Modeling Viral Capsid Assembly", Adv. Chem. Phys. **155**, 1–68 (2014).

[189] P. Hänggi, P. Talkner, and M. Borkovec, "Reaction-rate theory: fifty years after Kramers", Rev. Mod. Phys. **62**, 251–341 (1990).

[190] A. Berezhkovskii and A. Szabo, "One-dimensional reaction coordinates for diffusive activated rate processes in many dimensions", J. Chem. Phys. **122**, 014503 (2004).

[191] B. Peters, P. G. Bolhuis, R. G. Mullen, and J.-E. Shea, "Reaction coordinates, one-dimensional Smoluchowski equations, and a test for dynamical self-consistency", J. Chem. Phys. **138**, 054106 (2013).

[192] J. Glaser, T. D. Nguyen, J. A. Anderson, P. Liu, F. Spiga, J. A. Millan, D. C. Morse, and S. C. Glotzer, "Strong scaling of general-purpose molecular dynamics simulations on gpus", Comput. Phys. Commun. **192**, 97–107 (2015).

[193] J. A. Anderson, C. D. Lorenz, and A. Travesset, "General purpose molecular dynamics simulations fully implemented on graphics processing units", J. Comput. Phys. **227**, 5342–5359 (2008).

[194] G.-R. Yi, D. J. Pine, and S. Sacanna, "Recent progress on patchy colloids and their self-assembly", J. Phys. Condens. Matter **25**, 193101 (2013).

[195] C. Zeng, Y. Chen, K. Kirschbaum, K. J. Lambright, and R. Jin, "Emergence of hierarchical structural complexities in nanoparticles and their assembly", Science **354**, 1580–1584 (2016).

[196] D. C. Rapaport, "Role of Reversibility in Viral Capsid Growth: A Paradigm for Self-Assembly", Phys. Rev. Lett. **101**, 186101 (2008).

[197] C. A. Kerfeld, S. Heinhorst, and G. C. Cannon, "Bacterial microcompartments.", Annu. Rev. Microbiol. **64**, 391–408 (2010).

[198] T. O. Yeates, C. S. Crowley, and S. Tanaka, "Bacterial Microcompartment Organelles: Protein Shell Structure and Evolution", Annu. Rev. Biophys. **39**, 185–205 (2010).

[199] T. A. Bobik, B. P. Lehman, and T. O. Yeates, "Bacterial microcompartments: widespread prokaryotic organelles for isolation and optimization of metabolic pathways", Mol. Microbiol. **98**, 193–207 (2015).

[200] C. Chowdhury, S. Sinha, S. Chun, T. O. Yeates, and T. A. Bobik, "Diverse bacterial microcompartment organelles.", Microbiol. Mol. Biol. Rev. **78**, 438–468 (2014).

[201] R. Frey, S. Mantri, M. Rocca, and D. Hilvert, "Bottom-up Construction of a Primordial Carboxysome Mimic", J. Am. Chem. Soc. **138**, 10072–10075 (2016).

[202] G. D. Price and M. R. Badger, "Evidence for the role of carboxysomes in the cyanobacterial CO 2-concentrating mechanism", Can. J. Bot. **69**, 963–973 (1991).

[203] D. F. Savage, B. Afonso, A. H. Chen, and P. A. Silver, "Spatially Ordered Dynamics of the Bacterial Carbon Fixation Machinery", Science **327**, 1258–1261 (2010).

[204] N. M. Mangan, A. Flamholz, R. D. Hood, R. Milo, and D. F. Savage, "pH determines the energetic efficiency of the cyanobacterial CO2 concentrating mechanism.", Proc. Natl. Acad. Sci. U.S.A. **113**, E5354–62 (2016).

[205] S. Tanaka, C. A. Kerfeld, M. R. Sawaya, F. Cai, S. Heinhorst, G. C. Cannon, and T. O. Yeates, "Atomic-Level Models of the Bacterial Carboxysome Shell", Science **319**, 1083–1086 (2008).

[206] B. D. Rae, B. M. Long, M. R. Badger, and G. D. Price, "Structural Determinants of the Outer Shell of β-Carboxysomes in Synechococcus elongatus PCC 7942: Roles for CcmK2, K3-K4, CcmO, and CcmL", PLOS ONE **7**, e43871 (2012).

[207] M. T. Lin, A. Occhialini, P. J. Andralojc, J. Devonshire, K. M. Hines, M. A. J. Parry, and M. R. Hanson, "β-Carboxysomal proteins assemble into highly organized structures in Nicotiana chloroplasts", Plant J. **79**, 1–12 (2014).

[208] M. F. Hagan and R. Zandi, "Recent advances in coarse-grained modeling of virus assembly", Curr. Opin. Virol. **18**, 36–43 (2016).

[209] J. C. Cameron, S. C. Wilson, S. L. Bernstein, and C. A. Kerfeld, "Biogenesis of a Bacterial Organelle: The Carboxysome Assembly Pathway", Cell **155**, 1131–1140 (2013).

[210] J. Lidmar, L. Mirny, and D. R. Nelson, "Virus shapes and buckling transitions in spherical shells", Phys. Rev. E **68**, 051910 (2003).

[211] M. Widom, J. Lidmar, and D. R. Nelson, "Soft modes near the buckling transition of icosahedral shells", Phys. Rev. E **76**, 031911 (2007).

[212] E. H. Yong, D. R. Nelson, and L. Mahadevan, "Elastic Platonic Shells", Phys. Rev. Lett. **111**, 177801 (2013).

[213] C. M. Funkhouser, R. Sknepnek, and M. O. de la Cruz, "Topological defects in the buckling of elastic membranes", Soft Matter **9**, 60–68 (2012).

[214] A. Levandovsky and R. Zandi, "Nonequilibirum Assembly, Retroviruses, and Conical Structures", Phys. Rev. Lett. **102**, 198102 (2009).

[215] J. Wagner and R. Zandi, "The Robust Assembly of Small Symmetric Nanoshells", Biophys. J **109**, 956–965 (2015).

[216] S. D. Hicks and C. L. Henley, "Irreversible growth model for virus capsid assembly", Phys. Rev. E **74**, 031912 (2006).

[217] M. P. do Carmo, "Differential Geometry of Surfaces", in *Differential forms and applications* (Springer Berlin Heidelberg, 1994), pp. 77–98.

[218] H. S. Seung and D. R. Nelson, "Defects in flexible membranes with crystalline order", Phys. Rev. A **38**, 1005–1018 (1988).

[219] D. R. Nelson, "Spherical Crystallography: Virus Buckling and Grain Boundary Scars", arXiv:cond-mat/0311413 (2003).

[220] M. Bowick, A. Cacciuto, D. R. Nelson, and A. Travesset, "Crystalline Order on a Sphere and the Generalized Thomson Problem", Phys. Rev. Lett. **89**, 185502 (2002).

[221] M. J. Bowick, D. R. Nelson, and A. Travesset, "Interacting topological defects on frozen topographies", Phys. Rev. B **62**, 8738 (2000).

[222] A. R. Bausch, M. J. Bowick, A. Cacciuto, A. D. Dinsmore, M. F. Hsu, D. R. Nelson, M. G. Nikolaides, A. Travesset, and D. A. Weitz, "Grain Boundary Scars and Spherical Crystallography", Science **299**, 1716–1718 (2003).

[223] T. Einert, P. Lipowsky, J. Schilling, M. J. Bowick, and A. R. Bausch, "Grain Boundary Scars on Spherical Crystals", Langmuir **21**, 12076–12079 (2005).

[224] X. S. Ling, "Dislocation dynamics: Scars on a colloidal crystal ball", Nat. Mater. **4**, 360–361 (2005).

[225] T. Kohyama and G. Gompper, "Defect Scars on Flexible Surfaces with Crystalline Order", Phys. Rev. Lett. **98**, 198101 (2007).

[226] B. Øksendal, "Stochastic Differential Equations", in *Stochastic differential equations* (Springer Berlin Heidelberg, 2003), pp. 65–84.

[227] N. G. Van Kampen, "Stochastic Differential Equations", in *Stochastic processes in physics and chemistry* (Elsevier, 2007), pp. 396–421.

[228] L. Evans, *An Introduction to Stochastic Differential Equations* (American Mathematical Society, Providence, Rhode Island, Dec. 2013).

[229] C. W. Gardiner, *Handbook of Stochastic Methods*, Vol. 13, Springer Series in Synergetics (Springer Berlin Heidelberg, Berlin, Heidelberg, 2004).

[230] A. W. C. Lau and T. C. Lubensky, "State-dependent diffusion: Thermodynamic consistency and its path integral formulation", Phys. Rev. E **76**, 011123 (2007).