Efficient control and spontaneous transitions

by

Miranda D. Louwerse

B.Sc., University of the Fraser Valley, 2017

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

in the
Department of Chemistry
Faculty of Science

© Miranda D. Louwerse 2022
SIMON FRASER UNIVERSITY
Summer 2022

Declaration of Committee

Name: **Miranda D. Louwerse**

Degree: **Doctor of Philosophy**

Thesis title: **Efficient control and spontaneous transitions**

Committee: **Chair:**   Paul Li
Professor, Chemistry

David Sivak
Supervisor
Associate Professor, Physics

Paul Percival
Committee Member
Professor Emeritus, Chemistry

Krzysztof Starosta
Committee Member
Professor, Chemistry

John Bechhoefer
Examiner
Professor, Physics

Christoph Dellago
External Examiner
Professor
Department of Physics
University of Vienna

Abstract

Quantifying the dynamics and energetics of a system as it undergoes a transition between stable conformations is central to the study of reaction mechanisms and the derivation of reaction rates. For conformational changes in biomolecules, the space the system navigates is high dimensional, presenting challenges to the observation of reactive events in simulation or experiment and obscuring dynamical details that are relevant to the reaction mechanism. Developments in recent decades in transition-path theory, transition-path sampling, and single-molecule experimental techniques have transformed our ability to observe biomolecular reactions in microscopic detail and extract general features of the mechanism. Nearly simultaneously, rapid developments in the field of stochastic thermodynamics extend familiar notions of work, heat, and entropy to nonequilibrium contexts. One focus of these efforts is the design of protocols that dynamically manipulate the system's conformation with minimal energetic cost. In this thesis, I investigate slow, energetically efficient driving protocols that drive a system between conformations corresponding to endpoints of a reaction, aiming to find connections between principles of efficient driving and the spontaneous transition mechanism in the absence of driving. First, I develop an alternative perspective of transition-path theory (which describes reactive events) that unites it with stochastic thermodynamics to describe flows of entropy, energy, and information during the reaction. This also provides a thermodynamic measure of the relevance of a particular degree of freedom to the reaction, providing an optimization criterion for selecting collective variables. Next, I design protocols that invert the magnetization of a 3×3 Ising model with minimal energetic cost, determining that using multiple control parameters, which provide additional flexibility in manipulating the system conformation, allows it to be driven along a fast-relaxing pathway between reaction endpoints. Finally, I directly compare these designed protocols with the spontaneous transition mechanism for magnetization inversion in the same Ising model, finding that designed protocols capture general features of the spontaneous mechanism and energetics given the constraints on the control parameters. This work provides a basis for investigating the connection between efficient protocols and spontaneous transition mechanism which can be further probed in a wider variety of systems.

Keywords: Reaction mechanisms, transition-path theory, minimum-work protocols, nonequilibrium thermodynamics

Acknowledgements

Over the years, I have come to a fuller appreciation of the undergraduate education I received at UFV. I particularly thank my undergraduate advisor Noham Weinberg for sparking my interest in chemical physics, for teaching me about statistical mechanics, Brownian motion, and Kramers' theory, and for probing for the first time my (lack of) understanding of reaction coordinates at the microscopic level. I can now confirm that a one-dimensional reaction coordinate does exist and, in theory at least, I know how to find it.

My family and friends have been an endless support. My husband Mitch has been a steady support through the ups and downs of life and research and celebrated victories with me when they do occur. Thank you for taking me outside away from the computer screen to recharge, catch fish, and let my subconscious work while staring at rivers, lakes, and mountains. I am excited to see what the future holds for our life together. Thank you to my parents, siblings, and extended family (both biological and through marriage) for encouraging me throughout this process and keeping me grounded in what truly matters. To my nieces Tessa and Hallie, you are amazing and inspire me every day.

Table of Contents

List of Figures

Chapter 1

Introduction

The reaction mechanism describing how reactants transform into products is of central interest to chemistry. Complex reactions can be deconstructed into a series of elementary steps whose interdependence leads to the overall reaction kinetics. Additionally, rate-limiting steps that are critical to reaction progress can be identified. Theoretical models describing reaction mechanisms significantly improve understanding of these complex processes, providing language to qualitatively describe the transient behaviour of the system as it undergoes a reaction and quantifying its speed (kinetics) and spontaneous direction (thermodynamics), allowing predictions of how perturbations to the system or environment might change the reaction outcomes.

In biochemical contexts, reactions often consist of conformational changes in large biomolecules [1], where the system must navigate its high-dimensional conformation space between one region characterizing the reactants to another characterizing the products. For example, proteins fold from a disordered peptide chain into an energetically stable conformation whose structure and dynamics serve a specific biological function. Molecular machines catalyze unfavorable reactions, pump ions or molecules across cell membranes, and transport cargo around the cell, undergoing significant conformational changes while performing these tasks. These molecules are typically flexible, soft-matter materials in aqueous solution, constantly bombarded with surrounding water molecules that make their motion stochastic. Additionally, biomolecules operate in nonequilibrium environments, consuming chemical energy to fuel a myriad of biological processes and dissipating heat along the way as required by the second law of thermodynamics [1].

Progress in single-molecule experimental techniques, such as atomic force microscopy and optical tweezers [2], has greatly improved the ability to observe biological processes and determine quantitative aspects of their thermodynamics and kinetics [3]. To complement this, computer simulations are routinely performed on large biomolecules, providing atomistic detail of their mechanisms [4, 5].

In this thesis, nonequilibrium driving protocols are investigated [6, 7], which can be used to study conformational changes in biomolecules. If a set of externally controllable

parameters are chosen that restrain the system conformation to the reactant region or to the product region, changing these parameters drives the system between conformations of interest. These protocols can be implemented in experiment, for example by holding the two ends of a biomolecule in optical tweezers and pulling them apart or pushing them together to unfold and fold the molecule [8, 9, 10], and in simulation, by providing an energetic bias that restrains the system's conformation [11]. Driving protocols provide work to the system and, when performed in finite duration, necessarily dissipate some excess work as heat.

This thesis examines whether there is a correspondence between protocols designed to be energetically efficient when driving the system in finite time between reactant and product, and the reaction mechanism for the spontaneous transition in the absence of driving. Throughout, a classical description of the system that incorporates stochastic dynamics is used, which is well-suited to describing conformational changes in biomolecules in noisy environments [1, 2]. Additionally, a reaction is described generally as a transition from a reactant molecule or conformation A to product B in order to represent the concept of a reaction without details of the system involved.

1.1 Spontaneous transitions

How do reactions occur? The reactant and product are usually metastable molecular species or conformations, meaning their free energy is low relative to the free-energetic barrier separating them (Fig. 1.1). In order for the reaction to occur, the system must accumulate sufficient free energy to overcome the barrier; reaction theories attempt to quantify the frequency of this event by considering the underlying dynamics and energetics.

The speed and spontaneous direction of a reaction from reactant A to product B are parameterized by phenomenological rate constants $k_{A\to B}$ and $k_{B\to A}$ and equilibrium free-energy difference ΔF_{AB} [12]

$$A \underset{k_{B\to A}}{\overset{k_{A\to B}}{\rightleftharpoons}} B \,. \tag{1.1}$$

Finding values for these reaction parameters either experimentally, theoretically, or computationally, is a major focus of research in quantitative sciences. Importantly, the rate constants and free-energy difference are related by the principle of detailed balance [13]

$$\frac{k_{A\to B}}{k_{B\to A}} = e^{-\beta\Delta F_{AB}} \,, \tag{1.2}$$

providing a constraint on the equilibrium kinetic and thermodynamic parameters of the process.

1.1.1 Reaction theories

One of the first conceptual insights into chemical reactions is the existence of an energetic barrier separating metastable states A and B; a successful reaction requires the system to
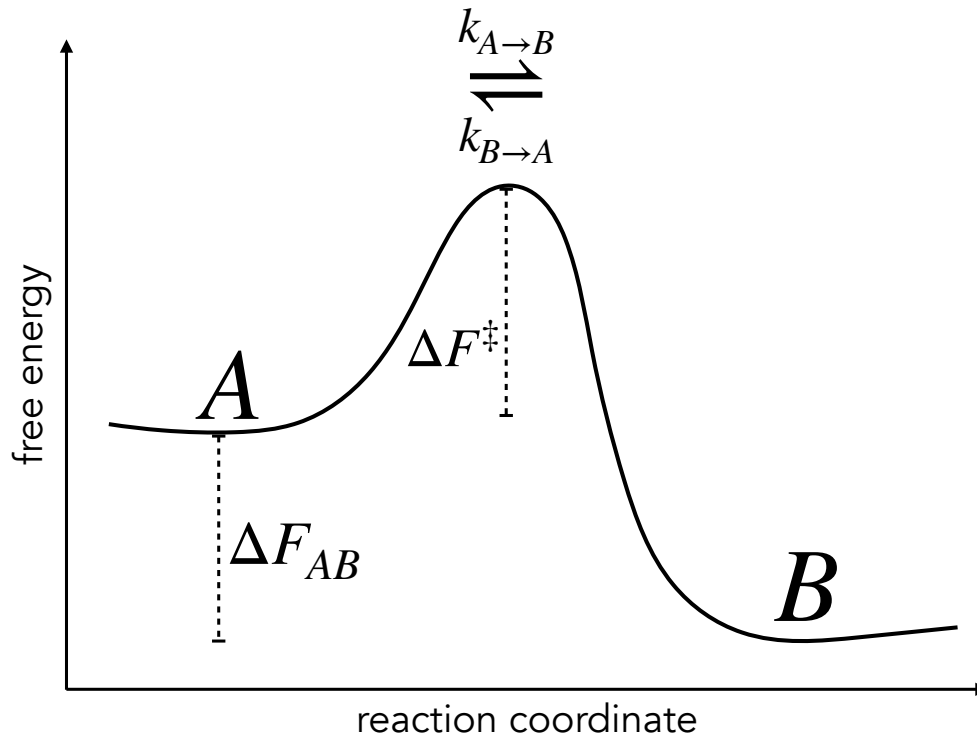
Figure 1.1: Schematic for an $A \rightarrow B$ reaction. The system free energy is plotted as a function of the reaction coordinate, with metastable states A and B separated by a free-energy barrier. The reaction free energy is the difference between A and B, ΔF_{AB}. The frequency with which the system crosses the energy barrier from $A \rightarrow B$ ($B \rightarrow A$) is given by rate constants $k_{A \rightarrow B}$ and $k_{B \rightarrow A}$.

gain sufficient *activation energy* to form the *activated complex* or *transition-state species* at the peak of the energy barrier, then release this energy as it forms the stable product. The activation energy relates the rate of a reaction to the temperature of the environment through the Arrhenius equation [12]

$$k_{A \rightarrow B} = \Gamma e^{-\beta \Delta E^{\ddagger}} \, , \tag{1.3}$$

for inverse temperature $\beta = 1/k_{B}T$ in units of Boltzmann's constant, and energy difference $\Delta E^{\ddagger} \equiv E_{TS} - E_A$ between the activated complex (transition state) and the reactant. This explains a qualitative observation about reactions: as the temperature is increased, the speed of the reaction increases. Measurements of the rate constant at a range of temperatures allow calculation of the activation energy. The prefactor Γ is referred to as a frequency factor, accounting for the frequency of collisions that allow reactant A to form the activated complex. For reactions in the gas phase, this frequency factor can be accounted for via collision theory [12] which estimates the frequency of collisions with sufficient energy and

3

correct orientation for the system to overcome the energy barrier at a particular temperature. For conformational changes in solution, collisions come from the surrounding solvent molecules and the prefactor can be estimated using transition-state theory and a diffusion equation for the dynamics [14].

Transition-state theory derives the rate constant using statistical mechanics [15, 16, 17], providing a connection between the rate constant and thermodynamic parameters of the activated complex. The transition state is imagined as a hypersurface of the configuration space parameterized by a reaction coordinate that the system must pass through as it reacts from A to B. Assuming that the reactant region is in thermal equilibrium and forbidding any recrossing of the transition-state surface, statistical mechanics can be used to derive the rate constant in terms of the *free-energy* difference between the reactant A and the transition state,

$$k_{A\to B}^{\text{TST}} = \frac{k_{\text{B}}T}{h}e^{-\beta\Delta F^{\ddagger}} = \frac{k_{\text{B}}T}{h}e^{\Delta H^{\ddagger}/k_{\text{B}}}e^{-\beta\Delta E^{\ddagger}} . \tag{1.4}$$

where $\Delta H^{\ddagger}$ is the activation entropy of the reaction, accounting for the difference in the entropy of the accessible conformational space in the transition-state region and the reactant region, and h is Planck's constant. Transition-state theory has been wildly successful at describing a wide variety of reactions and extended to include quantum effects such as tunnelling [18]. Its language and concepts form the basis of a chemist's understanding of activated processes.

A drawback of transition-state theory is the absence of recrossing; once the system has passed through the transition-state surface from A it is assumed to never return. Thus the unidirectional flux across the barrier used to derive the rate constant in transition-state theory overestimates the true rate constant,

$$k_{A\to B} = \kappa k_{A\to B}^{\text{TST}} \tag{1.5}$$

where $\kappa \leq 1$ is a transmission coefficient correcting for recrossing events [19]. The transmission coefficient and the activation free energy each depend on the choice of dividing surface (although their product is always the "true" rate constant). By varying the position of the dividing surface away from the top of the energy barrier, different estimates of the rate constant are obtained via transition-state theory (which excludes calculation of the transmission coefficient). For example, variational transition-state theory [18, 19] finds a position for the transition-state surface that minimizes the rate constant estimated using transition-state theory.

However, consideration of recrossing events is necessary to estimate the transmission coefficient and the "true" rate constant. Ideally, its calculation should consider both the immediate dynamics at the transition state and information about the trajectory at later times so that trajectories that cross the transition-state surface multiple times are counted a single time toward the reactive flux. An improvement on the transition-state theory as-

4

sumptions was made by Kramers [14, 20] to allow calculation of the transmission coefficient for dynamics described by diffusion on a one-dimensional reaction coordinate, considering both high- and low-friction regimes. Assuming that the system is in local equilibrium in the energetic well corresponding to mesostate A (meaning the initial probability distribution in the state space on A's side of the transition barrier is proportional to the Boltzmann distribution), Kramers assumes a steady state where once the system reaches the product it is reinitialized in the reactant state, generating a flux of trajectories from A to B. The rate at which system probability leaks over the barrier into mesostate B can be estimated from the diffusion equation by approximating the free-energy surface near A and the transition-state region as quadratic functions with curvatures ω_A and $\omega^{\ddagger}$ respectively. In the high-friction limit, this yields the transmission coefficient

$$\kappa_{\text{Kramers}} = \frac{\omega_A \omega^{\ddagger}}{2\pi\gamma} \tag{1.6}$$

as a correction to the transition-state theory rate constant that accounts for recrossing events, where γ is the friction coefficient for the diffusive system.

Since Kramers' reaction theory models the noisy environment encountered by biomolecules, it is useful in predicting rate constants for biomolecular transformations so long as the dynamics are accurately described by diffusion on a one-dimensional reaction coordinate. For example, it has been useful in interpreting information from atomic force spectroscopy experiments [21, 22], where, e.g., protein folding can be probed by pulling the ends of the molecule apart. However, the *a priori* assumption of a one-dimensional reaction coordinate that accurately captures the reaction dynamics is challenging since criteria for identifying such a coordinate are unclear [23]. Kramers' description of diffusion-dominated activated processes has been extended to multidimensional diffusion [24], easing the assumption of a one-dimensional reaction coordinate, but still requiring knowledge of a small set of relevant coordinates to make the calculation tractable. However, the development of transition-path theory provides a quantitative definition of the "true" one-dimensional reaction coordinate required for Kramers' description, as discussed in Section 1.1.2.

1.1.2 Transition-path theory

While the concept of a transition state is useful in many contexts, there are also many reactions that cannot be described using transition-state theory. These include conformational changes where the energy landscape is rugged or lacks an identifiable transition state that is a bottleneck for the reaction. Additionally, biomolecules have very high-dimensional conformational spaces (in principle, $3N$ coordinates are required to describe the three-dimensional position of an N-atom molecule), resulting in massive amounts of data that require dimensionality reduction to extract useful information, such as a small set of coordinates that capture important aspects of the reaction mechanism. It is usually not clear *a priori* what

system coordinates might serve as good reaction coordinates, despite the fact that reaction theories tend to rely on their existence and appropriate identification. It is therefore important to derive a reaction theory that does not depend on already knowing a physical one-dimensional coordinate or any specific details about the reaction mechanism, including the presence of a single transition state. This allows the theory to describe many types of reactions, either as an activated process with a single energy barrier or as diffusion through a network of mesostates as in Markov-state models [25]. Seminal work in the last few decades in transition-path sampling [26, 27, 28] and transition-path theory [29, 30] have come far in quantitatively describing reactions using statistical tools, also yielding a one-dimensional reaction coordinate between any two mesostates A and B.

It is first important to understand some of the challenges associated with describing reactions in atomistic detail, rather than integrating out degrees of freedom deemed irrelevant to the reaction. Of particular relevance to this discussion is classical molecular dynamics simulations, which describe a molecule's three-dimensional position and velocity as the system is propagated forward in time using Newton's equations of motion, with the system's potential energy described by an empirical force field which sums energetic contributions from internal degrees of freedom (bonds, angles, and torsions) as well as Coulombic and inter-atomic repulsive forces between pairs of atoms. Molecular dynamics simulations are widely used throughout the natural sciences to study a variety of problems in biology, chemistry, and materials, elucidating relationships between the structure, function, and dynamics of molecules.

To ensure accuracy in simulating the dynamics of many interacting atoms, a relatively small integration time step must be used, typically on the order of femtoseconds; in contrast, timescales for biologically relevant reactions are typically on the order of micro-seconds to seconds [25]. This timescale separation means that gathering samples of the rare *reactive trajectories* of the system that travel between metastable states in order to determine their statistics cannot be done with "brute force" by simply simulating the dynamics until reactions occur; instead, techniques must be developed to directly sample these trajectories. The same timescale problem will appear again when discussing collective variables and free-energy calculations in Section 2.3.1.

The seminal method for directly sampling reactive trajectories was developed in 1998 by Dellago, *et al.* [26], known as *transition-path sampling.* Since then, many more methods to sample and analyze transition paths have been developed, such as transition-interface sampling [31], milestoning [28], and forward-flux sampling [32]. The original transition-path sampling technique enacts a Markov process in the space of trajectories, using random shooting from a trial trajectory to generate new samples. From this, a sample of trajectories from A to B is obtained, which can be used to obtain rate constants using Chandler's reactive-flux formalism [26, 33]. In principle, information about the reaction mechanism is

also contained in this trajectory sample, but a theoretical framework is needed to extract and interpret such information.

Thus transition-path theory [29] was developed to quantitatively describe the entire reaction and determine its rate constant, without assumptions of metastability for the reactant and product or any specific details of the reaction mechanism (e.g., the presence of a single transition state). Imagine a long ergodic equilibrium trajectory that travels infinitely many times between two non-overlapping subsets of microstates corresponding to the reactant and product, A and B. Next, select from this ergodic supertrajectory all the trajectory segments that leave A and next reach B without returning to A. Each trajectory segment is a *transition path* or *reactive trajectory*, and the entire ensemble of reactive trajectories is the *transition-path ensemble* (TPE). All reaction parameters as well as the reaction mechanism can be determined from this ensemble of trajectories.

The statistical description of the transition-path ensemble relies on the definition of the *committor function* $q(\phi)$ (also called the commitment or splitting probability [34], or the folding probability in the protein folding literature [35]), the probability that a trajectory initiated from a microstate ϕ next reaches B before A. The committor maps the conformation space onto the interval $q(\phi) \in [0,1]$, identifying the transition-state ensemble as conformations making up the $q(\phi) = 0.5$ isocommittor surface [36]. The committor allows calculation of the reaction rate, mean first-passage time, and mean transition-path duration from a one-dimensional description [37] providing a connection to Kramers' theory and the required one-dimensional reaction coordinate. The versatility of transition-path theory in describing a wide variety of systems makes it a leading theory today in determining reaction mechanisms in computational sciences. Additionally, transition-path theory has been used to interpret experimental data and estimate the committor function for a protein-folding reaction [3, 38]. Some quantitative and qualitative aspects of this theory are reviewed further in Section 2.2.

The committor's relevance to determining reaction parameters and reaction mechanisms has earned it the title of the "true" or "ideal" one-dimensional reaction coordinate [23, 39, 40, 41, 42]. Its importance to chemical physics has spurred work to use machine learning to estimate the committor for systems of interest [43, 44, 45] (calculating the committor is generally a computationally expensive task [23]). However, the committor does not necessarily convey mechanistically useful information, as it simply gives a number for each microstate of the system. Instead, the dominant features of the system on the isocommittor surfaces (all conformations with the same committor value) [46] are of interest. In particular, it is helpful to determine a small set of physically relevant collective variables (functions of system microstate) that provide intuitive understanding about the reaction mechanism; the committor and/or other statistics (such as the free-energy landscape) can then be determined in the given collective-variable space. To obtain accurate information about the reaction mechanism, the chosen set of collective variables must capture the relevant infor-

7

mation about the reaction. A number of approaches have been proposed to identify small sets of relevant collective variables, often relying on ideas of timescale separation to identify slow degrees of freedom by approximating eigenfunctions of the dynamical operator [47, 48] or determining a reduced representation that is predictive about future states of the system [49]. The development of methods to identify and represent relevant degrees of freedom for a reaction is currently an active area of research.

1.2 Efficient control

Considerable progress has been made in recent decades in understanding nonequilibrium systems and the statistical nature of the second law of thermodynamics, including the derivation of nonequilibrium work relations, fluctuation theorems, and thermodynamic uncertainty relations [50, 6, 51, 52, 53]. In this thesis, control protocols which drive the system between conformations corresponding to the endpoints A and B of a reaction are investigated. In contrast to transition paths, which occur due to spontaneous fluctuations from the environment occasionally giving rise to rare reactive trajectories, control protocols are enacted by an experimenter external to the system, who changes a set of *control parameters* according to a fixed schedule, driving the system between thermodynamic states which could be designed to correspond to reaction endpoints. Control protocols can be implemented in experiment [8, 9, 54, 10] or simulation [55, 56], providing molecular-level information about a wide variety of systems.

1.2.1 Driving protocols in simulation and experiment

Control parameters can be any externally tunable parameters of the system that allow the system to be driven between desired thermodynamic states. Examples from classical thermodynamics include temperature, pressure, and volume. Single-molecule experimental techniques developed over the last decades have broadened the spectrum of available control parameters. For example, optical tweezers can provide force to the ends of a protein and cause it to unfold; in this case the distance between optical tweezers holding the ends of the molecule serves as a control parameter, and it is tuned according to some protocol to drive the system between folded and unfolded conformations. Significant experimental and theoretical work has been done to learn thermodynamic properties of systems from their response to such experimental control protocols [57, 58, 22]. Computer simulations offer even greater flexibility in choosing control parameters; essentially any function of the system's atomic coordinates can be driven between relevant endpoints via some time-dependent biasing potential.

This thesis investigates protocols that minimize the average excess work done on the system during a long-but-finite duration, and probes their relationship to spontaneous transition paths. In order to explain some relevant concepts related to control protocols, imagine

an ideal gas contained in a piston with adjustable volume that is driven from a low-pressure to high-pressure state by decreasing the volume in a constant-temperature environment (volume is the control parameter in this scenario).

If the volume is changed slowly throughout the protocol (a reversible or quasistatic process), the system relaxes continuously to remain in thermal equilibrium with the environment, and the distribution of gas molecules throughout the piston remains uniform. Work is done on the system reversibly, equal to the change in free energy between the final and initial equilibrium ensembles corresponding to low- and high-pressure states of the system [12]

$$\mathcal{W}^{\text{rev}} = \Delta F^{\text{eq}} = k_{\text{B}} T \ln \frac{V_{\text{f}}}{V_{\text{i}}} \,, \tag{1.7}$$

where V_{i} and V_{f} are the initial and final volumes of the system, respectively.

However, if the volume is reduced relatively quickly, the gas particles will not instantaneously be uniformly distributed throughout the piston; they only partially relax towards the uniform distribution throughout the control protocol, and only fully equilibrate when the system is left for a long time at the final volume. The system is out of equilibrium with the environment throughout the protocol, and on average excess work is performed on the system above the equilibrium free-energy change:

$$\langle \mathcal{W}^{\text{ex}} \rangle_{\Lambda} \equiv \langle \mathcal{W} \rangle_{\Lambda} - \Delta F^{\text{eq}} \geq 0 \,, \tag{1.8}$$

where $\langle \ldots \rangle_{\Lambda}$ denotes an average over system responses to the same driving protocol Λ. The excess work is dissipated as heat to the environment when the system relaxes, quantifying the energy "wasted" during the process. This inequality for excess work derives from the second law of thermodynamics, which requires irreversible entropy production for any nonequilibrium process.

The second law of thermodynamics is a statistical law that is only true on average [51]. For systems such as biomolecules that experience large fluctuations, the excess work done on the system differs each time the protocol is performed, forming an excess-work distribution $P(\mathcal{W}^{\text{ex}}|\Lambda)$. This distribution can have nonzero probability for negative values of excess work, indicating that for some trajectories the work is less than the free-energy difference. However, the average excess work is positive (1.8).

Remarkably, the inequality Eq. (1.8) can be derived from an equality (assuming the system begins in equilibrium),

$$\langle e^{-\beta \mathcal{W}} \rangle_{\Lambda} = e^{-\beta \Delta F^{\text{eq}}} \,, \tag{1.9}$$

which was first introduced by Jarzynski in 1997 [50]. The Jarzynski equation holds for any protocol duration (even when the system is far from equilibrium during the protocol), in principle allowing calculation of the equilibrium free-energy difference between thermodynamic states from nonequilibrium system responses. Soon after, Crooks derived a work

fluctuation theorem for nonequilibrium control protocols [6]

$$\frac{P(\mathcal{W}|\Lambda)}{P(-\mathcal{W}|\tilde{\Lambda})} = e^{-\beta \mathcal{W}^{\mathrm{ex}}} \, , \tag{1.10}$$

which relates the work distributions in the forward and time-reversed protocols to the excess protocol work. The time-reversed protocol begins with the system in equilibrium at the final control parameter, then runs the protocol in reverse to reach the initial control parameter (e.g., expanding the gas in the piston rather than compressing it). When the process is performed in finite time, the system lags the equilibrium distribution during both the forward and reverse protocols, yielding different work distributions for each process. However, Eq. (1.10) places a tight constraint on the allowable work distributions for the forward and time-reversed processes in terms of the excess work. Similar fluctuations theorems have been derived for nonequilibrium processes other than control protocols [51], relating the probabilities of forward and time-reversed processes to irreversible entropy production.

These nonequilibrium work relations allow equilibrium thermodynamic information to be extracted from nonequilibrium driving protocols. For example, the free-energy difference between equilibrium ensembles can be computed using Eq. (1.9), although in practice bidirectional variations on (1.9) such as Bennett's acceptance ratio [59] are used to improve accuracy and reduce bias in the estimate [60, 61, 62]. The potential of mean force on the system along the driving coordinate can also be calculated using nonequilibrium protocols, as shown by Hummer and Szabo [63]. This makes nonequilibrium protocols a useful tool for determining quantitative information about the system's underlying free-energy landscape. The potential of mean force for DNA hairpin folding has been calculated experimentally using such methods [10].

1.2.2 Minimum-work protocols

The equilibrium free-energy difference between control-protocol endpoints is fixed because free energy is a state function, however the work done during the protocol is path dependent. For a fixed protocol duration, it is therefore of interest to find a protocol that minimizes excess work. Such minimum-work protocols (MWPs) in turn provide insight into energetically efficient operation of biological molecular machines [1], the design of methods for efficient chemical processing [64], and informs the design of novel nanotechnology [65]. Additionally, the magnitude of excess work has implications for the precision and accuracy of free-energy estimates from nonequilibrium control protocols [60, 61, 62], which are improved when less excess work is done. Exact results for minimum-work protocols for any duration have been derived in a few model systems [66]; approximations are available in the limits of long and short protocol duration [67, 7, 68, 69, 70].

In this thesis, minimum-work protocols in the limit of long-but-finite duration are considered, as derived in Ref. [7]. Here, a linear-response expansion of the system's response

to the driving protocol yields a Riemannian metric in control-parameter space measuring the excess power during the protocol. Minimum-work protocols designed using this metric are geodesics, the generalization of straight lines between protocol endpoints to curved, non-Euclidean geometry.

The metric is interpreted as generalized friction, quantifying system resistance (due to equilibrium fluctuations) to changes in the control parameter. For the example of isothermal compression of gas in a piston, this resistance comes from the particles accumulating near the piston as the volume is reduced because there has not been sufficient time for the particles to diffuse out of the way and form the equilibrium uniform distribution. This same idea is intuitive in biomolecular experiments, for example in optical-tweezer experiments where a protein is pulled apart. When the pulling is done in finite time, interactions within the protein that keep it in a folded state resist the changes in control parameter that drive these interactions to break.

The generalized friction encodes information about changes in the system's equilibrium free-energy during the protocol [67] (its thermodynamics) and equilibrium relaxation modes [68, 71] (its kinetics). Since generalized friction is related to equilibrium fluctuations, it seems natural that the geometry of control-parameter space would reflect the geometry of the underlying system energetics and dynamics.

1.3   Motivation for this thesis

To study the correspondence between driving protocols and spontaneous transitions, control protocol endpoints are chosen such that the system is restrained to reaction endpoints A or B when in equilibrium at corresponding control-parameter values. The choice of A and B and appropriate control parameters depend on the system of interest; in this thesis, calculations are performed on a particular model system but general principles will be drawn that could apply to many systems of interest.

The long-duration limit for driving protocols is considered rather than studying minimum-work protocols for all durations. The generalized friction metric that derives from a linear-response approximation of a protocol's excess work has enticing features that may enable a correspondence with spontaneous transition mechanisms. It quantifies resistance to driving the system through an ensemble of pathways in conformational space by capturing local features of the system's free energy and relaxation times in the corresponding conjugate-force space, which reflects the dynamics of collective variables the control parameters are able to control. Local features of the system's energetics and dynamics are also relevant to characterizing transient dynamics during a spontaneous transition path [72, 73, 74, 75]. If there is a significant energy barrier the system must overcome to get from A to B, it seems intuitive that a minimum-work protocol would drive the system through a saddle point on the free-energy surface (where the transition state of the spontaneous reaction is expected

to occur), requiring the minimum work possible to drive the system over the energy barrier. This suggests a correspondence between the minimum-work protocol and the spontaneous transition path of the system between A and B which will be investigated throughout this thesis.

Theoretical descriptions of minimum-work protocols and spontaneous transitions differ in several ways. Minimum-work protocols are performed in a fixed duration, while spontaneous transition paths from A to B have a range of duration. Protocols exchange work with the system throughout the protocol, while there is no work performed on the system during a spontaneous transition. Protocol endpoints are defined in control-parameter space, while endpoints of a spontaneous transition are defined in conformational (or collective-variable) space. Protocols drive the system out of equilibrium, while spontaneous transitions occur when the system is at equilibrium.

Nevertheless, there are some commonalities between the processes that are worth investigating. Both processes share the same conformational space, with internal energy coupling the system's many degrees of freedom. The energy and entropy of the conformational space, and changes in these state functions during trajectories either generated by a control protocol or selected as a transition path, can therefore be directly compared. With appropriate choice of control parameters, the endpoint distributions of the protocol can approximate the unperturbed distribution in metastable basins surrounding transition-path endpoints A and B, providing at minimum a control-parameter space capable of distinguishing metastable conformations. However, there is great interest in sets of collective variables that not only distinguish metastable basins but also provide information about the reaction mechanism. A control-parameter space is designed to bias collective variables that are presumed to be relevant to a reaction and minimum-work protocols are found and tested for their correspondence to the spontaneous reaction mechanism.

1.4   Outline of this thesis

Chapter 2 begins with some mathematical background on stochastic dynamics, transition-path theory, and minimum-work protocols. Chapter 3 provides a thermodynamic accounting of the information contained in reactive trajectories, allowing quantitative determination of the relevance of a set of collective variables to a reaction. This establishes a connection between a dynamical theory describing spontaneous transitions and thermodynamic definitions of energy and entropy along individual trajectories [76]. In Chapter 4, minimum-work protocols for driving an Ising model system between energetically stable configurations are calculated. The additional freedom provided by controlling multiple collective variables is explored and related to reduced excess work relative to a naive (constant-velocity) protocol between the same endpoints [77]. In Chapter 5, the two paradigms are drawn together, comparing the mechanism and energy flows into the system during a minimum-work protocol

and during the spontaneous transition-path ensemble, highlighting some intuitive connections between the processes and some complicating factors. Conclusions and suggestions for future work are discussed in Chapter 6.

Chapter 2

Theoretical Background

In this chapter, some necessary theoretical background is introduced which will inform the work in the remainder of this thesis. Section 2.1 begins with a review of stochastic dynamics, followed by some background into information theory in Section 2.1.2 and stochastic thermodynamics in Section 2.1.3. The fundamentals of transition-path theory are then reviewed in Section 2.2 and the utility of the committor as a reaction coordinate is discussed in more detail. In Section 2.3, the thermodynamics of control protocols are reviewed, followed by introduction of minimum-work protocols in the linear-response regime and the generalized friction.

2.1 Stochastic dynamics and thermodynamics

Equilibrium thermodynamics explains many observations in the physical sciences, including phase transitions, reaction spontaneity, and reversible processes. As interest in and ability to study nonequilibrium systems grows, integrating system dynamics into descriptions of thermodynamics has become a necessary tool to understand biological phenomena. This motivates the field of stochastic thermodynamics, which provides definitions of work, heat, and entropy for individual trajectories that are then readily applied to nonequilibrium systems. These use stochastic descriptions of system dynamics, where noise provides a random force to the system in addition to conserved and non-conserved forces, for example, that arise from internal energy and nonequilibrium environmental forces.

2.1.1 Ensemble and trajectory descriptions of stochastic dynamics

Suppose the system is in microstate ϕ at time t. In deterministic classical mechanics, $\phi(t)$ describes the position and momentum of each coordinate of the system at time t; by integrating classical equations of motion, the microstate at a later time $\phi(t + \delta t)$ can be determined [78]. Importantly, a trajectory initiated from a point ϕ follows the same path through phase space each time. This is not the case with stochastic dynamics; each step in the dynamics incorporates random forces, so a trajectory initiated from ϕ follows a

different path in each trial, leading to a distribution of possible outcomes at a later time $p(\phi, t + \delta t | \phi, t)$. However, with certain assumptions on the random noise, average properties of the ensemble of trajectories can be derived. Formal introductions to stochastic dynamics are presented elsewhere [13]. This section instead focuses on developing some intuition about stochastic dynamics, with relevant equations and notation appearing as necessary.

To begin, assume the system is initialized in state ϕ at time t_0 and the system dynamics are simulated via a Markov jump process for n discrete time steps until time t_n to produce a trajectory $\{\phi(t)\} = \{\phi_0, t_0; \phi_1, t_1; \ldots; \phi_n, t_n\}$. It is assumed that the dynamics are Markovian, that is, the next state visited by the system depends only on the current state and not past states visited by the system [13, 79]. The probability of producing a particular trajectory $\{\phi(t)\}$ can then be expressed as

$$p(\{\phi(t)\}) = p(\phi_0, t_0) \prod_{i=1}^{n} p(\phi_i, t_i | \phi_{i-1}, t_{i-1}) , \tag{2.1}$$

where $p(\phi_i, t_i | \phi_{i-1} t_{i-1})$ is the *transition probability*, the probability the system makes a transition from ϕ_{i-1} to ϕ_i in the time interval $t_{i-1} \to t_i$, and $p(\phi_0, t_0)$ is the initial probability distribution of the system.

The transition probabilities should be chosen such that at long times the system distribution reproduces the distribution of interest. To generate the dynamics of an equilibrium distribution $\pi(\phi)$, the transition probabilities should be independent of time and obey the principle of *detailed-balance* [13]

$$p(\phi | \phi')\pi(\phi') = p(\phi' | \phi)\pi(\phi) . \tag{2.2}$$

For each pair of microstates, the equilibrium probability the system is in ϕ' and makes a transition to ϕ is equal to the probability the system is in ϕ and makes a transition to ϕ'. Rearranging, the ratio of transition rates is equal to the ratio of equilibrium probabilities

$$\frac{p(\phi | \phi')}{p(\phi' | \phi)} = \frac{\pi(\phi)}{\pi(\phi')} . \tag{2.3}$$

There are many possible ways to choose transition probabilities that satisfy these conditions. Given an initial state ϕ_0 drawn from the initial probability distribution $p(\phi_0, t_0)$, a general Monte Carlo procedure generates a new trial state ϕ^* with probability $p^{\text{trial}}(\phi^* | \phi_0)$ and then draws a random number between 0 and 1 to determine whether or not the new state is accepted. The total transition probability from ϕ_0 to ϕ^*,

$$p(\phi_1 = \phi^*, t_1 | \phi_0, t_0) = p^{\text{trial}}(\phi^* | \phi_0) p^{\text{accept}}(\phi^* | \phi_0) , \tag{2.4}$$

where $p^{\text{accept}}(\phi^* | \phi_0)$ is the acceptance probability that must be chosen to ensure dynamics obey detailed balance. If the random number is less than the transition probability, the

system moves to $\phi_1 = \phi^*$ at time t_1, otherwise it stays in state ϕ_0 at time t_1. The process is repeated till a trajectory of the desired length is generated.

There are many choices for generating the trial state, depending on the system of interest. In molecular simulations, random perturbations to positions and momenta of atoms could be applied [80]. For the Ising model studied in this thesis (introduced in Section 4.2), a trial state is generated by randomly choosing one of N spins making up the system from a uniform distribution and flipping it, thus $p^{\text{trial}}(\phi^*|\phi_0) = N^{-1}$. If the trial probability is symmetric, $p^{\text{trial}}(\phi^*|\phi_0) = p^{\text{trial}}(\phi_0|\phi_*)$, the constraint on transition probabilities imposed by detailed balance falls on the acceptance probability. For physical systems with energy $V(\phi)$ that relax to a Boltzmann distribution, the acceptance probabilities must satisfy

$$\frac{p^{\text{accept}}(\phi^*|\phi_0)}{p^{\text{accept}}(\phi_0|\phi^*)} = e^{-\beta[V(\phi^*)-V(\phi_0)]} . \tag{2.5}$$

Common choices for the acceptance probability include the Metropolis criterion [81] and the Glauber criterion [82], but other choices are possible as long as they obey certain constraints to ensure the process converges to the desired distribution. The Glauber acceptance probability is used in this thesis, which has the form

$$p^{\text{accept,GL}}(\phi^*|\phi_0) = \frac{1}{1 + e^{\beta[V(\phi^*)-V(\phi_0)]}} . \tag{2.6}$$

The Glauber acceptance criterion (2.6) satisfies detailed balance,

$$\frac{p^{\text{accept,GL}}(\phi^*|\phi_0)}{p^{\text{accept,GL}}(\phi_0|\phi^*)} = \frac{1 + e^{\beta[V(\phi_0)-V(\phi^*)]}}{1 + e^{\beta[V(\phi^*)-V(\phi_0)]}} \tag{2.7a}$$

$$= \frac{e^{\beta V(\phi^*)} + e^{\beta V(\phi_0)}}{e^{\beta V(\phi_0)} + e^{\beta V(\phi^*)}} \frac{e^{\beta V(\phi_0)}}{e^{\beta V(\phi^*)}} \tag{2.7b}$$

$$= e^{-\beta[V(\phi^*)-V(\phi_0)]} . \tag{2.7c}$$

Equation (2.1) is the probability for a single trajectory generated by a Markov jump process. It is often of interest to obtain average quantities over all possible trajectories $\{\phi(t)\}$; it is therefore useful to have a description of the way the system's probability distribution evolves in time $p(\phi, t)$ from an initial distribution. This is commonly expressed as a Master equation [13]

$$\mathrm{d}_t p(\phi, t) = \sum_{\phi'} T_{\phi\phi'}(t) p(\phi', t) , \tag{2.8}$$

where $T_{\phi\phi'}(t) = \lim_{\delta t \to 0} p(\phi, t + \delta t|\phi', t)/\delta t$ is the *transition rate* of $\phi' \to \phi$ transition at time t, which is expressed as the transition probability per infinitesimal time step divided by time step δt. The $\phi \to \phi$ transition rate is defined as

$$T_{\phi\phi} = -\sum_{\phi' \neq \phi} T_{\phi'\phi} \tag{2.9}$$

16

to ensure the probability distribution remains normalized.

The Master equation describes how the probability distribution changes from the initial condition $p(\boldsymbol{\phi}_0, t_0)$ as the system evolves stochastically. When the transition rate is independent of time, the system may relax to a stationary distribution $p^{\mathrm{ss}}(\boldsymbol{\phi})$

$$0 = \mathrm{d}_t p^{\mathrm{ss}}(\boldsymbol{\phi}) = \sum_{\boldsymbol{\phi}'} T_{\boldsymbol{\phi}\boldsymbol{\phi}'} p^{\mathrm{ss}}(\boldsymbol{\phi}') \,, \tag{2.10}$$

where the probability of each state is constant. This is the global balance condition [13], ensuring that there are no changes to the probability of any state $\boldsymbol{\phi}$

$$0 = \sum_{\boldsymbol{\phi}'} \left[T_{\boldsymbol{\phi}'\boldsymbol{\phi}} p^{\mathrm{ss}}(\boldsymbol{\phi}) - T_{\boldsymbol{\phi}\boldsymbol{\phi}'} p^{\mathrm{ss}}(\boldsymbol{\phi}') \right] \,. \tag{2.11}$$

However, the stationary distribution is not necessarily an equilibrium distribution; there may be net probability flows in the system that are balanced such that there is no overall change in the probability of each state. Equilibrium requires the stronger condition of detailed balance, which can be expressed in terms of transition rates as

$$T_{\boldsymbol{\phi}\boldsymbol{\phi}'} \pi(\boldsymbol{\phi}') = T_{\boldsymbol{\phi}'\boldsymbol{\phi}} \pi(\boldsymbol{\phi}) \,, \tag{2.12}$$

for all pairs of states $\boldsymbol{\phi}$, $\boldsymbol{\phi}'$. Not only is the overall probability in each state conserved at equilibrium, but probability flux between all states is equal in both directions, resulting in no flux or probability flows throughout the state space. If global balance holds but detailed balance does not, the stationary distribution can be called a *nonequilibrium steady state* [79].

2.1.2 Entropy and information theory

Statistical mechanics treats the system state $\boldsymbol{\phi}$ as a random variable $\boldsymbol{\Phi}$. This allows an information-theoretic perspective on entropy that parallels its thermodynamic definition. The foundations of information theory were laid out by Claude Shannon at Bell Labs in 1948 [83]. Since then, it has been recognized that information is physical, as elegantly argued by Landauer [84]. Some key concepts from information theory are briefly reviewed to better elucidate this connection, which will be relevant for Chapter 3.

First, consider quantifying the uncertainty about a random variable X before taking a measurement. Throughout, the convention of capital symbols to denote a random variable and lowercase symbols to denote a particular realization of the random variable is used. The uncertainty about discrete random variable X with probability mass function $p(X = x)$ is given by its Shannon entropy, defined to be

$$H(X) \equiv -\sum_x p(x) \ln p(x) \,. \tag{2.13}$$

Notably, this is the same form for thermodynamic entropy derived by Gibbs, up to a constant k_{B} [12]. The average amount of information gained about the distribution upon measurement of the random variable is equal to its Shannon entropy, therefore information represents a reduction in uncertainty.

Information theory gains more power by considering multiple random variables X and Y. Consider X and Y jointly distributed according to the probability mass function $p(X = x, Y = y)$. For example, X and Y could be coordinates representing the state of a two-dimensional system. The joint entropy of the joint random variable (X, Y) is

$$H(X, Y) \equiv -\sum_{x,y} p(x, y) \ln p(x, y) \,. \tag{2.14}$$

Similarly, the marginal entropy of one of these random variables (without loss of generality X) is expressed in terms of its marginal distribution $p(x) = \sum_y p(x, y)$ as

$$H(X) = -\sum_x p(x) \ln p(x) \tag{2.15a}$$

$$= -\sum_{x,y} p(x, y) \ln \sum_y p(x, y) \,. \tag{2.15b}$$

This is the uncertainty about the joint state that can be reduced from observing only X. The conditional entropy of X given Y can be similarly defined as

$$H(X|Y) \equiv -\sum_{x,y} p(x, y) \ln p(x|y) \,, \tag{2.16}$$

which represents the remaining uncertainty about X when Y is known. This hints that information is shared between X and Y, which can be precisely quantified by the mutual information between X and Y,

$$I(X; Y) \equiv \sum_{x,y} p(x, y) \ln \frac{p(x, y)}{p(x)p(y)} \,. \tag{2.17}$$

Mutual information is a nonlinear statistical measure of the relationship between random variables X and Y, specifically measuring the reduction in uncertainty about one variable from measuring the other,

$$I(X; Y) = H(X) - H(X|Y) \tag{2.18a}$$

$$= H(Y) - H(Y|X) \tag{2.18b}$$

$$= I(Y; X) \,. \tag{2.18c}$$

It is a symmetric quantity that is positive semi-definite

$$I(X;Y) = -\sum_{x,y} p(x,y) \ln \frac{p(x)p(y)}{p(x,y)} \tag{2.19a}$$

$$\geq \ln \sum_{x,y} p(x,y) \frac{p(x)p(y)}{p(x,y)} \tag{2.19b}$$

$$= \ln \sum_{x,y} p(x)p(y) \tag{2.19c}$$

$$= \ln 1 \tag{2.19d}$$

$$= 0 \,, \tag{2.19e}$$

and zero if and only if X and Y are independent $[p(x,y) = p(x)p(y)]$. Eq. (2.19b) uses Jensen's inequality [83].

Information theory also has broad relevance throughout the field of stochastic thermodynamics, as an essential aspect of describing many classes of nonequilibrium systems [85, 86, 87].

2.1.3 Stochastic thermodynamics

Stochastic thermodynamics defines thermodynamic quantities for individual trajectories. Here, a trajectory generated by a Markov jump process introduced in Section 2.1.1 is considered. The transition rate for a $\phi' \to \phi$ transition is $T_{\phi\phi'}$, and the probability the system is in state ϕ at time t is $p(\phi, t)$. Let $V(\phi, t)$ denote the energy of state ϕ at time t. The first law of thermodynamics accounts for changes to the system's energy as work or heat,

$$\mathrm{d}_t V(t) = \sum_{\phi} \mathrm{d}_t \left[p(\phi, t) V(\phi, t) \right] \tag{2.20a}$$

$$= \sum_{\phi} \left[V(\phi, t) \mathrm{d}_t p(\phi, t) + p(\phi, t) \mathrm{d}_t V(\phi, t) \right] \tag{2.20b}$$

$$= \dot{Q}(t) + \dot{W}(t) \tag{2.20c}$$

where the overdot notation indicates a rate (here, of heat and work) that cannot be expressed as the time derivative of a state function. Work is performed by an external agent making changes to the energy of microstates, while heat is produced during system transitions by exchange of energy with thermodynamic reservoirs. The sign convention is used that positive heat and work correspond to energy flows into the system. Further discussion of work done on the system during a control protocol is discussed in Section 2.3.2.

Heat flow to the system for the instantaneous transition from $\phi' \to \phi$ is the change in system energy during the transition at time t, which is equal to a log ratio of transition

rates that obey detailed balance [88],

$$q_{\phi\phi'}(t) \equiv k_{\mathrm{B}}T[V(\phi,t) - V(\phi',t)] = k_{\mathrm{B}}T \ln \frac{T_{\phi'\phi}(t)}{T_{\phi\phi'}(t)} \ . \tag{2.21}$$

Assuming the reservoirs in contact with the system remain in equilibrium, the heat flow is proportional to the negative entropy change in the environment

$$q_{\phi\phi'}(t) = -k_{\mathrm{B}}T \, h_{\phi\phi'}^{\mathrm{env}}(t) \ . \tag{2.22}$$

Entropy is expressed as a dimensionless quantity throughout this thesis. The change in system entropy for a $\phi' \to \phi$ transition is the difference in specific entropy $h(\phi,t) \equiv -\ln p(\phi,t)$,

$$\Delta h_{\phi\phi'}^{\mathrm{sys}}(t) \equiv \ln \frac{p(\phi',t)}{p(\phi,t)} \ . \tag{2.23}$$

Thus the total (irreversible) entropy production for a $\phi' \to \phi$ transition is

$$h_{\phi\phi'}^{\mathrm{irr}}(t) = \Delta h_{\phi\phi'}^{\mathrm{sys}}(t) + h_{\phi\phi'}^{\mathrm{env}}(t) \tag{2.24a}$$

$$= \ln \frac{T_{\phi\phi'}(t)p(\phi',t)}{T_{\phi'\phi}(t)p(\phi,t)} \ , \tag{2.24b}$$

the log ratio of the forward and reverse flux for a transition between ϕ' and ϕ at time t, also called the thermodynamic force [88]. If the system is in equilibrium so there is equal flux of forward and reverse transitions $[T_{\phi\phi'}\pi(\phi') = T_{\phi'\phi}\pi(\phi)]$, the total entropy production is zero for each transition,

$$h_{\phi\phi'}^{\mathrm{irr}} = h_{\phi\phi'}^{\mathrm{sys}} + h_{\phi\phi'}^{\mathrm{env}} \tag{2.25a}$$

$$= -k_{\mathrm{B}}T \ln \frac{T_{\phi'\phi}}{T_{\phi\phi'}} + k_{\mathrm{B}}T \ln \frac{\pi(\phi')}{\pi(\phi)} \tag{2.25b}$$

$$= 0 \ . \tag{2.25c}$$

Assuming the system has reached steady state $p^{\mathrm{ss}}(\phi)$ with its reservoirs (which may be out of equilibrium), multiplying the thermodynamic quantities defined for each transition by the net probability flux of the transition

$$J_{\phi\phi'} \equiv T_{\phi\phi'}p^{\mathrm{ss}}(\phi') - T_{\phi'\phi}p^{\mathrm{ss}}(\phi) \ , \tag{2.26}$$

gives ensemble-averaged thermodynamic quantities. The steady-state heat flow to the system is

$$\dot{Q} = \frac{k_{\mathrm{B}}T}{2} \sum_{\phi\phi'} J_{\phi\phi'} \ln \frac{T_{\phi'\phi}}{T_{\phi\phi'}} \tag{2.27a}$$

$$= -k_{\mathrm{B}}T\dot{H}^{\mathrm{env}} \, , \tag{2.27b}$$

where the factor of $\frac{1}{2}$ accounts for the inclusion of each pair of states twice in the sum.

System entropy is unchanging at steady state (even for nonequilibrium steady states) since it is a state function of an unchanging distribution:

$$\mathrm{d}_t H^{\mathrm{sys}} = \frac{1}{2} \sum_{\phi\phi'} J_{\phi\phi'} \ln \frac{p^{\mathrm{ss}}(\phi')}{p^{\mathrm{ss}}(\phi)} \tag{2.28a}$$

$$= \frac{1}{2} \sum_{\phi\phi'} J_{\phi\phi'} \left[\ln p^{\mathrm{ss}}(\phi') - \ln p^{\mathrm{ss}}(\phi) \right] \tag{2.28b}$$

$$= \frac{1}{2} \left[\sum_{\phi'} \ln p^{\mathrm{ss}}(\phi') \sum_{\phi} J_{\phi\phi'} - \sum_{\phi} \ln p^{\mathrm{ss}}(\phi) \sum_{\phi'} J_{\phi\phi'} \right] \tag{2.28c}$$

$$= 0 \, , \tag{2.28d}$$

using $\sum_{\phi'} J_{\phi\phi'} = \sum_{\phi'} J_{\phi'\phi} = 0$ for all ϕ when the system is at steady state (Eq. (2.11)). Therefore the total entropy production rate is equal to the rate of change in environmental entropy

$$\dot{H}^{\mathrm{irr}} = \sum_{\phi>\phi'} J_{\phi\phi'} \ln \frac{T_{\phi\phi'}p^{\mathrm{ss}}(\phi')}{T_{\phi'\phi}p^{\mathrm{ss}}(\phi)} \tag{2.29a}$$

$$= \sum_{\phi\phi'} J_{\phi\phi'} \ln \frac{p^{\mathrm{ss}}(\phi')}{p^{\mathrm{ss}}(\phi)} + \sum_{\phi\phi'} J_{\phi\phi'} \ln \frac{T_{\phi\phi'}}{T_{\phi'\phi}} \tag{2.29b}$$

$$= \mathrm{d}_t H^{\mathrm{sys}} + \dot{H}^{\mathrm{env}} \tag{2.29c}$$

$$= 0 + \dot{H}^{\mathrm{env}} \geq 0 \, , \tag{2.29d}$$

and is positive semi-definite. $\dot{H}^{\mathrm{irr}} = 0$ if and only if detailed balance (2.12) holds. Irreversible entropy production is the driving force behind all spontaneous processes, and there is significant interest in understanding its role in biological processes [1, 89]. Chapter 3 shows that irreversible entropy production also plays a role in characterizing reaction mechanisms.

2.2   Theory and analysis of spontaneous transition paths

Given the stochastic dynamics described in Section 2.1.1, the theory of transition paths can now be introduced. Transition-path theory for Markov jump processes in discrete state spaces [30, 90] is presented, but the theory was originally developed for continuous sys-

tems [29]. Concepts of the transition-path ensemble, the committor, and reaction statistics computed from the committor are reviewed below.

2.2.1 Transition-path theory

Assume a system's dynamics are governed by the Master equation Eq. 2.8 with $\phi' \to \phi$ transition rate $T_{\phi\phi'}$ and that the system has reached the equilibrium distribution $\pi(\phi)$ such that detailed-balance holds (Eq. (2.12)). Select from the state space $\boldsymbol{\Phi}$ two distinct subsets of reactant microstates A and product microstates B. These subsets should be non-overlapping so the system must travel through an intermediate region I as it progresses from A to B. During a long ergodic equilibrium *supertrajectory*, the system transitions between A and B infinitely many times; the transition-path ensemble (TPE) is the set of trajectory segments (also called *reactive trajectories* or *transition paths*) that travel directly from A to B. Statistics such as reactive flux, mean first-passage time, and mean transition-path duration are derived from the frequency of $A \to B$ transition events using transition-path theory.

The joint probability that the system is in state ϕ and on a reactive trajectory segment

$$p(\phi, A \to B) = \pi(\phi)p(A \to B|\phi) \,, \tag{2.30}$$

is conveniently expressed as the product of the marginal (equilibrium) probability the system is in state ϕ and the conditional probability $p(A \to B|\phi)$ that the system is currently on an $A \to B$ reactive trajectory given that it is currently in state ϕ. It then remains to find this conditional probability.

Define the *forward committor* $q^+(\phi)$ to be the probability the system will next reach B given that it is in state ϕ, and the *backward committor* $q^-(\phi)$ to be the probability the system was more recently in A than B given that it is currently in state ϕ. These committor functions can be defined recursively for discrete-state systems:

$$\sum_{\phi'} T_{\phi'\phi} q^+(\phi') = 0 \tag{2.31a}$$

$$\sum_{\phi'} T_{\phi'\phi} \frac{\pi(\phi)}{\pi(\phi')} q^-(\phi') = 0 \tag{2.31b}$$

with boundary conditions $q^+(\phi) = 0(1)$ and $q^-(\phi) = 1(0)$ for $\phi \in A(B)$. For discrete-state systems, the committors can be calculated numerically by solving the system of linear equations 2.31 to find the committor of each state.

For equilibrium systems, the principle of microscopic reversibility [13] is used to relate the backward committor to the forward committor. The backward committor $q^-(\phi)$ is the probability the system in state ϕ was generated from a trajectory that originated in A. Each such trajectory from A to ϕ has an equally probable time-reversed trajectory from ϕ

22

to A. Therefore the ratio of $A \to \phi$ vs. $B \to \phi$ trajectories is equal to the ratio of $\phi \to A$ vs. $\phi \to B$ trajectories. Since the probability of a $\phi \to A$ trajectory is $1 - q^+(\phi)$, the forward and backward committor functions are related by the equality

$$1 - q^+(\phi) = q^-(\phi) . \tag{2.32}$$

This implies that due to time-reversal symmetry of equilibrium dynamics, either committor function (usually the forward committor is adopted) is sufficient on its own to fully describe the statistics of reactive trajectories in the intermediate state space between A and B.

With these two definitions, the joint probability the system is in state ϕ and on a reactive trajectory is expressed as

$$p(\phi, A \to B) = \pi(\phi) q^-(\phi) q^+(\phi) . \tag{2.33}$$

In words, this is the equilibrium probability the system is in state ϕ and was more recently in A and next reaches B. The independence (on the rhs) of the probability the system was more recently in A given it is currently in ϕ and the probability the system next reaches B conditioned on state ϕ can be understood from the Markov property: whether the system was more recently in A or B does not influence future dynamics since they are assumed to be memoryless and depend only on the current state ϕ. Therefore the probability the system next proceeds to B (rather than A), conditioned on current state ϕ, is independent of whether the system was more recently in A or B. Equation (2.33) readily yields the probability distribution of reactive trajectories,

$$p(\phi | A \to B) = \frac{\pi(\phi) q^-(\phi) q^+(\phi)}{p_{A \to B}} , \tag{2.34}$$

where

$$p_{A \to B} = \sum_{\phi} \pi(\phi) q^-(\phi) q^+(\phi) \tag{2.35}$$

is the probability the system is on an $A \to B$ reactive trajectory at time t during the ergodic supertrajectory.

A particularly useful output of transition-path theory is a description of the probability current of reactive trajectories. Consider the transition-path ensemble to be in a steady state where the system is initiated at the boundary of A and is constrained to reach B without returning to A. The system is instantly regenerated at the boundary of A once it reaches the boundary of B. This results in a steady-state current of trajectories flowing from A to B, with reactive probability current between states ϕ' and ϕ given by

$$J_{\phi\phi'}^{A \to B} = \pi(\phi') q^-(\phi') T_{\phi\phi'} q^+(\phi) . \tag{2.36}$$

This is the joint probability that the system is in state ϕ', was more recently in A than B, makes a transition in simulation time dt to state ϕ, and will next reach B before A. When averaged over all transitions that leave A and start a transition path, or all transitions that finish transition paths with a jump into B, the reactive probability current yields the flux of $A \to B$ transitions through the intermediate region I

$$\nu_{A \to B} = \sum_{\phi' \in A, \phi \in I} J_{\phi\phi'}^{A \to B} \tag{2.37a}$$

$$= \sum_{\phi' \in I, \phi \in B} J_{\phi\phi'}^{A \to B} . \tag{2.37b}$$

In fact, any dividing surface between A and B can be used to find the reaction rate, since the steady-state reactive flux is the same through any dividing surface. Additionally, the flux of $A \to B$ transitions is equal to the flux of $B \to A$ "reverse" transitions, since the system is at equilibrium. The reactive flux is related to the rate constants by

$$k_{A \to B} = \frac{\nu_{A \to B}}{p_A} \tag{2.38a}$$

$$k_{B \to A} = \frac{\nu_{A \to B}}{p_B} \tag{2.38b}$$

where

$$p_A = \sum_{\phi} \pi(\phi) q^-(\phi) \tag{2.39a}$$

$$p_B = \sum_{\phi} \pi(\phi)[1 - q^-(\phi)] \tag{2.39b}$$

is the probability the system was more recently in A (B) than in B (A) at some time t during the ergodic supertrajectory. The mean first-passage time to B (A) given that the system is currently in A (B) is the inverse of the rate constant

$$\tau_{A \to B}^{\mathrm{FP}} = k_{A \to B}^{-1} \tag{2.40a}$$

$$\tau_{B \to A}^{\mathrm{FP}} = k_{B \to A}^{-1} . \tag{2.40b}$$

Another quantity available from transition-path theory is the mean duration of transition paths $\tau_{A \to B}$, which is derived from the probability $p_{A \to B}$ that the system is currently on an $A \to B$ transition path and the mean round-trip time [91]

$$\tau^{\mathrm{RT}} \equiv \tau_{A \to B}^{\mathrm{FP}} + \tau_{B \to A}^{\mathrm{FP}} \tag{2.41}$$

for the system starting in A to reach B and come back to A for the first time,

$$\tau_{A \to B} = p_{A \to B}\, \tau^{\mathrm{RT}} \tag{2.42a}$$

$$= p_{A \to B} \left(\tau^{\mathrm{FP}}_{A \to B} + \tau^{\mathrm{FP}}_{B \to A} \right) \tag{2.42b}$$

$$= p_{A \to B} \left(\frac{1}{k_{A \to B}} + \frac{1}{k_{B \to A}} \right) \tag{2.42c}$$

$$= p_{A \to B} \left(\frac{p_A}{\nu_{A \to B}} + \frac{p_B}{\nu_{A \to B}} \right) \tag{2.42d}$$

$$= \frac{p_{A \to B}}{\nu_{A \to B}} \, , \tag{2.42e}$$

using

$$p_A + p_B = \sum_{\phi} \pi(\phi) q^-(\phi) + \sum_{\phi} \pi(\phi)[1 - q^-(\phi)] \tag{2.43a}$$

$$= \sum_{\phi} \pi(\phi) \tag{2.43b}$$

$$= 1 \tag{2.43c}$$

to obtain (2.42e). Analogous expressions have been derived for continuous state-space systems [29].

Transition-path theory has become the leading tool for describing reaction events, allowing calculation of average transition-path properties using the committor. Additionally, the committor is widely considered the "true" or "optimal" reaction coordinate [23, 46, 36]. Importantly, it provides language to describe reactions where an intuitive system coordinate that acts as a reaction coordinate cannot necessarily be found, and it provides quantitative information about the rate of the reaction through the reactive flux.

The committor and transition-path theory can also provide mechanistic information by highlighting regions of conformational space that are particularly relevant to the reaction [92]. The committor foliates the conformational space between A and B into a series of isocommittor surfaces (all microstates with the same value of the committor), and important features of the system on these surfaces can be determined. For example, the reactive probability distribution may lie within a "transition tube" in an appropriately chosen collective-variable space, so that the mechanism can be described by a single *most likely transition path* that in some way averages the distribution on each isocommittor surfaces. A minimum-free energy path can be found that is parallel to the free-energy gradient in collective-variable space (analogous to the minimum-energy path in the full configuration space) [93, 73, 92], or a maximum-flux transition path [74, 75] that maximizes the reactive flux on each isocommittor surface. Algorithms to compute these paths using molecular dynamics simulations in a space of many collective variables have been developed [94, 73].

Importantly, they require that the committor in chosen collective-variable space is a good approximation to the committor in the full configuration space.

2.3 Minimum-work control theory

In this section, the thermodynamic foundation necessary to understand minimum-work control protocols in the linear-response regime is discussed, beginning with a discussion of control-parameter space and its relationship to the conformational space it biases.

2.3.1 Control parameters and collective variables

As introduced in Section 1.2.1, there are many choices for control parameters to drive a system, such as temperature, pressure, and external magnetic fields. Since it is of interest to drive the system between metastable states, control parameters with clear correspondence to system collective variables are chosen. For example, a collective variable can be chosen that is presumed to have relevance to the reaction, then the system can be restrained to particular values of this collective variable using a harmonic trap. The centre of the trap (and/or its spring constant) acts as a control parameter, and the system is driven between values of the chosen collective variable corresponding to metastable states by a control protocol that moves the center of the trap [66, 95, 62].

The control parameter biases the system's internal energy $E_{\mathrm{int}}(\boldsymbol{\phi})$ with a control-parameter-dependent external energy $E_{\mathrm{ext}}(\boldsymbol{\phi}, \boldsymbol{\lambda})$, to give total energy

$$V(\boldsymbol{\phi}, \boldsymbol{\lambda}) = E_{\mathrm{int}}(\boldsymbol{\phi}) + E_{\mathrm{ext}}(\boldsymbol{\phi}, \boldsymbol{\lambda}) . \tag{2.44}$$

The total potential is related to the equilibrium probability distribution of the system at each control parameter,

$$\pi(\boldsymbol{\phi}|\boldsymbol{\lambda}) = e^{\beta[F^{\mathrm{eq}}(\boldsymbol{\lambda}) - V(\boldsymbol{\phi}, \boldsymbol{\lambda})]} , \tag{2.45}$$

with equilibrium free energy given $\boldsymbol{\lambda}$ of

$$F^{\mathrm{eq}}(\boldsymbol{\lambda}) = -k_{\mathrm{B}}T \ln \sum_{\boldsymbol{\phi}} e^{-\beta V(\boldsymbol{\phi}, \boldsymbol{\lambda})} \tag{2.46a}$$

$$= \langle V(\boldsymbol{\phi}, \boldsymbol{\lambda}) \rangle_{\boldsymbol{\lambda}}^{\mathrm{eq}} - k_{\mathrm{B}}T H^{\mathrm{eq}}(\boldsymbol{\lambda}) , \tag{2.46b}$$

where $\langle \ldots \rangle_{\boldsymbol{\lambda}}^{\mathrm{eq}}$ denotes an average over the equilibrium distribution $\pi(\boldsymbol{\phi}|\boldsymbol{\lambda})$, and

$$H^{\mathrm{eq}}(\boldsymbol{\lambda}) = -\sum_{\boldsymbol{\phi}} \pi(\boldsymbol{\phi}|\boldsymbol{\lambda}) \ln \pi(\boldsymbol{\phi}|\boldsymbol{\lambda}) \tag{2.47}$$

is the entropy of the equilibrium distribution at $\boldsymbol{\lambda}$. Appropriately chosen control parameters allow the system to preferentially sample regions of conformational space that are typically inaccessible to the system but may be visited transiently during a reaction.

Since the conformational space is typically too high-dimensional to interpret, sampling is often done in a lower-dimensional space defined by a set of collective variables $\boldsymbol{X}(\boldsymbol{\phi})$ to identify metastable conformations and free-energy barriers between them. A single point $\boldsymbol{x} = \boldsymbol{X}(\boldsymbol{\phi})$ in collective-variable space corresponds to a manifold of microstates in the higher-dimensional conformational space, so calculation of the marginal equilibrium probability of these coarse variables $\pi(\boldsymbol{X}(\boldsymbol{\phi}))$ in the absence of control ($E_{\text{ext}}(\boldsymbol{\phi}, \boldsymbol{\lambda}) = 0$) is desired. This is expressed in terms of the free energy of a particular value of collective variable $\boldsymbol{x}$,

$$F(\boldsymbol{x}) = -k_{\text{B}} T \ln \sum_{\boldsymbol{\phi}} e^{-\beta E_{\text{int}}(\boldsymbol{\phi})} \delta\left[\boldsymbol{X}(\boldsymbol{\phi}) - \boldsymbol{x}\right] \tag{2.48a}$$

$$= -k_{\text{B}} T \ln \pi(\boldsymbol{x}) - k_{\text{B}} T \ln Z , \tag{2.48b}$$

where $\delta\left[\boldsymbol{X}(\boldsymbol{\phi}) - \boldsymbol{x}\right]$ selects all microstates $\boldsymbol{\phi}$ with collective variable at point $\boldsymbol{x}$ and $Z = \sum_{\boldsymbol{\phi}} e^{-\beta E_{\text{int}}(\boldsymbol{\phi})}$ is the partition function of the (unperturbed) equilibrium distribution. If the collective variable is a one-dimensional reaction coordinate (such as the committor), Eq. (2.48) is the free energy of the reaction shown in Fig. 1.1, which incorporates both energetic and entropic aspects of the (many) system configurations consistent with the given value of the reaction coordinate.

Calculation of free energy in collective-variable space can be a computationally expensive task since sufficient sampling of the entire collective-variable space is needed to ensure converged estimates of free energy. For systems with metastable conformations, a trajectory initiated in one mesostate spends a (potentially) long time sampling microstates in that mesostate before overcoming the energy barrier to sample the other mesostate, and due to the timescale separation between the simulation timestep and reaction transitions, it becomes impractical to compute free energies via a brute-force approach.

Instead, enhanced-sampling methods use control parameters to more efficiently sample relevant regions of conformational space and compute free energies or other equilibrium properties [96]. Equilibrium samples are drawn independently from several control-parameter values that span the relevant collective-variable space. Because the control-parameter perturbations (and resulting equilibrium distributions (2.45)) that have been applied to the system are known, it is possible to reweight equilibrium samples taken from a range of control-parameter values to infer the distribution of the unperturbed system in collective-variable space [97, 96]. An analogous result was derived for nonequilibrium control protocols [63] where the system is driven along a single collective variable; the free energy along this driving protocol can be recovered by reweighting observations at each value of the collective variable using the work on the system up to that point in the protocol.

2.3.2 Thermodynamics of control protocols

Control protocols change control parameters $\boldsymbol{\lambda}$ between initial $\boldsymbol{\lambda}_{\text{initial}}$ and final $\boldsymbol{\lambda}_{\text{final}}$ values in finite time, so that entire trajectories are relevant rather than equilibrium samples from independent simulations performed at each control-parameter value.

The combination of microstate $\boldsymbol{\phi}$ and control parameter $\boldsymbol{\lambda}$ has an associated energy $V(\boldsymbol{\phi}, \boldsymbol{\lambda})$. The first law of thermodynamics requires that energy be conserved; therefore, changes in the system's energy must be accounted for as heat or work,

$$\mathrm{d}V(\boldsymbol{\phi}, \boldsymbol{\lambda}, t) = \sum_{\boldsymbol{\phi}} \left[\frac{\partial V(\boldsymbol{\phi}, \lambda, t)}{\partial \lambda} \frac{\partial \lambda}{\partial t} + \frac{\partial V(\boldsymbol{\phi}, \lambda, t)}{\partial \boldsymbol{\phi}} \frac{\partial \boldsymbol{\phi}}{\partial t} \right] \tag{2.49a}$$

$$= đ\mathcal{W}(\boldsymbol{\phi}, \boldsymbol{\lambda}, t) + đQ(\boldsymbol{\phi}, \boldsymbol{\lambda}, t) \,, \tag{2.49b}$$

where d denotes an exact differential and $đ$ an inexact differential. The work on the system is alternatively expressed as

$$đ\mathcal{W}(t) = -\sum_i đ\lambda_i(t) X_i(\boldsymbol{\phi}, \boldsymbol{\lambda}, t) \tag{2.50}$$

in terms of forces $X_i(\boldsymbol{\phi}, \boldsymbol{\lambda}, t) \equiv -\partial V(\boldsymbol{\phi}, \lambda, t)/\partial \lambda_i$ conjugate to respective control parameters λ_i.

The mean work during a protocol of duration Δt is

$$\langle \mathcal{W} \rangle_{\boldsymbol{\Lambda}} = -\int_0^{\Delta t} \mathrm{d}t \sum_i \dot{\lambda}_i(t) \langle X_i(\boldsymbol{\phi}, \boldsymbol{\lambda}, t) \rangle_{\boldsymbol{\Lambda}} \,, \tag{2.51}$$

where $\dot{\lambda}_i \equiv \partial \lambda_i/\partial t$ is the velocity of control parameter i, and $\langle \ldots \rangle_{\boldsymbol{\Lambda}}$ denotes a nonequilibrium average over the protocol, which depends on current control parameters and the entire history of control-parameter changes and system response.

When this process is done reversibly (quasistatically) so that the system is in equilibrium throughout the protocol, the mean conjugate force is related to changes in the equilibrium free energy, [67]

$$\langle X_i(\boldsymbol{\phi}, \boldsymbol{\lambda}, t) \rangle_{\boldsymbol{\Lambda}}^{\text{rev}} = \langle X_i \rangle_{\boldsymbol{\lambda}(t)}^{\text{eq}} = -\frac{\partial F^{\text{eq}}(\boldsymbol{\lambda})}{\partial \lambda_i} \,, \tag{2.52}$$

and the work on the system is equal to the free energy difference between control-parameter endpoints (2.46):

$$\mathcal{W} = \int_0^{\Delta t} \mathrm{d}t \sum_i \dot{\lambda}_i(t) \frac{\partial F^{\text{eq}}(\boldsymbol{\lambda})}{\partial \lambda_i} \tag{2.53a}$$

$$= \int_0^{\Delta t} \mathrm{d}t \frac{\mathrm{d}F^{\text{eq}}(\boldsymbol{\lambda})}{\mathrm{d}t} \tag{2.53b}$$

$$= F^{\text{eq}}(\boldsymbol{\lambda}_{\text{final}}) - F^{\text{eq}}(\boldsymbol{\lambda}_{\text{initial}}) \,. \tag{2.53c}$$

This equality between work and free energy is true for any quasistatic protocol, regardless of the exact path taken through control-parameter space.

When the control parameters are tuned in finite duration Δt, on average excess work is done on the system above the equilibrium free energy change,

$$\langle \mathcal{W}^{\mathrm{ex}} \rangle_{\mathbf{\Lambda}} = - \int_0^{\Delta t} \mathrm{d}t \sum_i \dot{\lambda}_i(t) \langle \delta X_i(\boldsymbol{\phi}, \boldsymbol{\lambda}, t) \rangle \tag{2.54}$$

where $\delta X_i(\boldsymbol{\phi}, \boldsymbol{\lambda}, t)) \equiv X_i(\boldsymbol{\phi}, \boldsymbol{\lambda}, t) - \langle X_i \rangle_{\boldsymbol{\lambda}(t)}^{\mathrm{eq}}$ is defined as the deviation of instantaneous conjugate force from its equilibrium mean at control parameter $\boldsymbol{\lambda}$.

The excess work during a protocol has special significance for nonequilibrium driving protocols. Assuming the system begins in equilibrium at $\boldsymbol{\lambda}_{\mathrm{initial}}$ and relaxes to equilibrium at $\boldsymbol{\lambda}_{\mathrm{final}}$ after the end of the protocol, excess work is equal to the irreversible entropy production of the nonequilibrium driving protocol, quantifying the energy wasted during the process.

$$\beta \langle \mathcal{W}^{\mathrm{ex}} \rangle_{\Lambda} \equiv \beta(\langle \mathcal{W} \rangle_{\Lambda} - \Delta F^{\mathrm{eq}}) \tag{2.55a}$$

$$= \beta \langle \Delta V \rangle_{\Lambda} - \beta \langle Q \rangle_{\Lambda} - \beta \langle \Delta V \rangle^{\mathrm{eq}} + \Delta H^{\mathrm{sys,eq}} \tag{2.55b}$$

$$= \beta \delta \langle V(\Delta t) \rangle_{\Lambda} + \Delta H_{\Lambda}^{\mathrm{env}} + \Delta H^{\mathrm{sys,eq}} . \tag{2.55c}$$

In (2.55b), the first law is used to equate work with change in energy and heat flow, and the free energy difference is expressed as changes in energy and entropy. Eq. (2.55c) equates heat with the environmental entropy change and defines the deviation of energy at the end of the protocol from its equilibrium value as $\delta \langle V(\Delta t) \rangle_{\Lambda} \equiv \langle V(\Delta t) \rangle_{\Lambda} - \langle V \rangle_{\boldsymbol{\lambda}_{\mathrm{final}}}^{\mathrm{eq}}$. When the system relaxes to equilibrium after the end of the protocol, $\delta \langle V(t \to \infty) \rangle_{\Lambda} = 0$. Also recognizing that $\Delta H_{\Lambda}^{\mathrm{env}} + \Delta H^{\mathrm{sys}} = \Delta H_{\Lambda}^{\mathrm{irr}}$, the excess work is equal to irreversible entropy production,

$$\beta \langle \mathcal{W}^{\mathrm{ex}} \rangle_{\Lambda} = \Delta H_{\Lambda}^{\mathrm{irr}} . \tag{2.56}$$

The connection between excess work and entropy production plays a central role in *nonequilibrium work relations*, such as the Jarzynski equality Eq. (1.9) [50] which allows estimation of the equilibrium free-energy difference ΔF^{eq} from nonequilibrium driving experiments, and Crooks' work fluctuation theorem Eq. (1.10) [6] which places a tight constraint on allowable work fluctuations.

2.3.3 Linear-response approximation to excess work

Given the role excess work plays in the nonequilibrium thermodynamics of driving protocols, it is helpful to understand how it depends on the choice of driving protocol. To this end, results have been derived to exactly or approximately calculate the excess work for different types of protocols. Exact results are analytically tractable for only a few special cases, such as for driving a system in a harmonic trap with control over the trap center and spring

constant [66]. Since exact results are difficult to derive, it is valuable to make approximations to the protocol work valid in different limits. In this thesis, the long-duration limit of protocol work is examined, using an approximation of excess work that uses linear-response theory.

The mean excess power on a system at time t is

$$\langle \mathcal{P}^{\text{ex}}(t) \rangle_{\boldsymbol{\Lambda}} = - \sum_i \dot{\lambda}_i(t) \langle \delta X_i(t) \rangle^{\text{neq}} , \tag{2.57}$$

the product of the (nonequilibrium) value of the conjugate forces and the rate of change of each control parameter at a given time. For a given control protocol, $\dot{\lambda}(t)$ is fixed and the nonequilibrium lag of conjugate forces from their equilibrium values throughout the control protocol must be determined. In Ref.[7], the lag is approximated using linear-response theory, which is outlined below.

For quasistatic protocols, the conjugate forces relax to their quasistatic values throughout the protocol. Assuming the protocols are performed in long-but-finite duration and that control parameters change continuosly with smooth first and second derivatives, the system response to external perturbations can be approximated using linear-response theory [13] as

$$\langle \delta X_i(t) \rangle^{\text{neq}} \approx \int_{-\infty}^{t} dt' \sum_j \chi_{ij}^{\boldsymbol{\lambda}(t)}(t-t') \left[\lambda_j(t') - \lambda_j(t) \right] . \tag{2.58}$$

Here, $\chi_{ij}^{\boldsymbol{\lambda}}(t-t')$ is the *linear-response function*

$$\chi_{ij}^{\boldsymbol{\lambda}}(t-t') \equiv \beta \frac{d}{dt} \langle \delta X_j(t') \delta X_i(t) \rangle_{\boldsymbol{\lambda}} \tag{2.59}$$

representing the response of conjugate force X_i at time t to a step perturbation of control parameter λ_j at time t' at control-parameter value $\boldsymbol{\lambda}$.

Integration by parts yields an approximation for the nonequilibrium conjugate force as

$$\langle \delta X_i(t) \rangle^{\text{neq}} \approx \int_{-\infty}^{t} dt' \sum_j \langle \delta X_i(0) \delta X_j(t-t') \rangle \dot{\lambda}_j(t') . \tag{2.60}$$

Assuming the control-parameter velocity changes slowly, a Taylor expansion of the control-parameter velocity $\dot{\boldsymbol{\lambda}}(t')$ around $t' = t$, truncated at the constant term yields

$$\langle \delta X_i(t) \rangle^{\text{neq}} \approx - \sum_j \zeta_{ij}(\boldsymbol{\lambda}(t)) \dot{\lambda}_j(t) , \tag{2.61}$$

where

$$\zeta_{ij}(\boldsymbol{\lambda}) \equiv \int_0^{\infty} dt \langle \delta X_i(0) \delta X_j(t) \rangle_{\boldsymbol{\lambda}} \tag{2.62}$$

is the *generalized friction matrix*, the time integral of the correlation function of conjugate forces for the system *at equilibrium* at control parameter $\boldsymbol{\lambda}$.

Combining this approximation for the mean nonequilibrium lag of conjugate forces with the control-parameter velocity yields the linear-response estimate for the mean excess power throughout the protocol, which is valid in the long-duration limit,

$$\mathcal{P}^{\mathrm{ex,LR}}(t) = \sum_{ij} \dot{\lambda}_i(t)\zeta_{ij}(\boldsymbol{\lambda}(t))\dot{\lambda}_j(t) \,. \tag{2.63}$$

This phenomenological expression for the excess power is local, depending only on instantaneous values of the control parameters and their time derivatives, rather than depending on the entire history of control-parameter perturbations. Integrating excess power over the protocol duration yields the mean excess work

$$\langle \mathcal{W}^{\mathrm{ex}} \rangle_{\boldsymbol{\Lambda}} \approx \int_0^{\Delta t} \mathrm{d}t \, \mathcal{P}^{\mathrm{ex,LR}}(t) \,. \tag{2.64}$$

2.3.4 Theoretical insight from generalized friction metric

The linear-response approximation to the excess work during a protocol has several features that provide insight into minimum-work protocols in the long-duration limit.

First, the generalized friction matrix has special mathematical properties that make it a Riemannian metric, providing a thermodynamic measure of the distance between two equilibrium distributions in control-parameter space which accounts for fluctuations required for the system to relax between the two distributions [67]. Minimum-work protocols are *geodesics* of the generalized friction in control-parameter space, which are shortest paths when distance is measured in terms of the friction metric rather than Euclidean projections. The friction defines a thermodynamic length for a given protocol,

$$\mathcal{L}_{\boldsymbol{\Lambda}} = \int_0^{\Delta t} \mathrm{d}t \sqrt{\mathcal{P}^{\mathrm{ex,LR}}(t)} \,, \tag{2.65}$$

which is independent of duration Δt. Using the Cauchy-Schwarz inequality, thermodynamic length provides a lower bound on protocol excess work

$$\mathcal{W}_{\boldsymbol{\Lambda}}^{\mathrm{ex,LR}} = \int_0^{\Delta t} \mathrm{d}t \, \mathcal{P}^{\mathrm{ex,LR}}(t) \tag{2.66a}$$

$$= \frac{1}{\Delta t} \int_0^{\Delta t} \mathrm{d}t \, 1 \cdot \int_0^{\Delta t} \mathrm{d}t \left[\sqrt{\mathcal{P}^{\mathrm{ex,LR}}(t)} \right]^2 \tag{2.66b}$$

$$\geq \frac{1}{\Delta t} \left[\int_0^{\Delta t} \mathrm{d}t \sqrt{\mathcal{P}^{\mathrm{ex,LR}}(t)} \right]^2 \tag{2.66c}$$

$$= \frac{\mathcal{L}^2}{\Delta t} \,. \tag{2.66d}$$

This lower bound is realized for protocols with constant excess power [7]. Therefore, minimum-work protocols in the linear-response regime are independent of protocol duration (up to a constant scaling of control-parameter velocity) and have constant excess power.

The generalized friction can be decomposed into components measuring the size of force fluctuations and the timescale of relaxation in the equilibrium distribution at control parameter $\boldsymbol{\lambda}$:

$$\zeta_{ij}(\boldsymbol{\lambda}) = \langle \delta X_i \delta X_j \rangle_{\boldsymbol{\lambda}} \tau_{ij}(\boldsymbol{\lambda}) \,, \tag{2.67}$$

where $\langle \delta X_i \delta X_j \rangle_{\boldsymbol{\lambda}}$ is the equilibrium covariance between conjugate force X_i and X_j at control parameter $\boldsymbol{\lambda}$, and

$$\tau_{ij}(\boldsymbol{\lambda}) = \int_0^\infty \mathrm{d}t \frac{\langle \delta X_i(0) \delta X_j(t) \rangle_{\boldsymbol{\lambda}}}{\langle \delta X_i \delta X_j \rangle_{\boldsymbol{\lambda}}} \tag{2.68}$$

is the integral relaxation time between conjugate forces X_i and X_j. The relevance of the force covariance matrix to control protocols is discussed in Ref. [67].

A major advantage of the linear-response approximation to excess work is the local description of the conjugate force lag from local perturbations. This greatly simplifies the original situation where the nonequilibrium conjugate force depends on the entire protocol history. The protocol that minimizes the linear-response approximation of excess work can be found using calculus of variations [78]. This amounts to solving the Euler-Lagrange equation [78] for $\boldsymbol{\lambda}(t)$

$$\zeta_{ki} \ddot{\lambda}_i + \frac{\partial \zeta_{ki}}{\partial \lambda_j} \dot{\lambda}_j \dot{\lambda}_i - \frac{1}{2} \frac{\partial \zeta_{ij}}{\partial \lambda_k} \dot{\lambda}_j \dot{\lambda}_i = 0 \,. \tag{2.69}$$

with endpoints $\boldsymbol{\lambda}(0) = \boldsymbol{\lambda}_{\text{initial}}$ and $\boldsymbol{\lambda}(\Delta t) = \boldsymbol{\lambda}_{\text{final}}$. For most systems, this Euler-Lagrange equation is not exactly solvable. Section 2.3.5 provides a method for generating numerical solutions to this equation. However, for a one-dimensional control parameter λ, the Euler-Lagrange equation can be solved for arbitrary generalized friction [7], yielding

$$\dot{\lambda}^{\text{MWP}}(t) \propto \zeta(\lambda(t))^{-1/2} \tag{2.70}$$

where the proportionality constant is chosen such that the protocol is completed in duration Δt. The resulting design principle of one-dimensional minimum-work control says that minimizing work requires slower changes to the control parameter wherever the resistance to driving (i.e., generalized friction) is large. The performance of one-dimensional driving using this framework have been studied numerically in several model systems [98, 99, 100].

2.3.5 Numerical calculations of minimum-work protocols

Solving the Euler-Lagrange equation for minimum-work protocols (Eq. (2.69)) is a non-trivial task. Analytical forms for the generalized friction are only available for a handful of systems [7, 101, 98], necessitating numerical estimation of the generalized friction at a range of control-parameter values. Optimization of a one-dimensional protocol is fairly

straightforward since the path is fixed and only the control-parameter velocity must be optimized. This can be done by numerically estimating the friction coefficient for a discrete set of M control parameter values $\lambda(m)$ between protocol endpoints, then solving Eq. (2.70) to find the time $t^{\mathrm{MWP}}(\lambda(m))$ for each control-parameter value $\lambda(m)$ during the protocol. Beginning with $t^{\mathrm{MWP}}(\lambda_0) = 0$, subsequent times can be found by expressing Eq. (2.70) as an iterative relation of discrete time intervals

$$t^{\mathrm{MWP}}(\lambda(m+1)) = t^{\mathrm{MWP}}(\lambda(m)) + \mathrm{d}t^{\mathrm{MWP}}_{m \to m+1} \tag{2.71a}$$

$$= t^{\mathrm{MWP}}(\lambda(m)) + \left[\frac{\zeta(\lambda(m)) + \zeta(\lambda(m+1))}{2} \right]^{1/2} [\lambda(m+1) - \lambda(m)] . \tag{2.71b}$$

All times $t^{\mathrm{MWP}}(\lambda(m))$ can be divided by the final time $t^{\mathrm{MWP}}(\lambda(M-1))$ to obtain the optimal scaled protocol duration for each control-parameter point. This algorithm gives the protocol time for control-parameter points with equal spatial separation; in Appendix B.2.1, a method to calculate a one-dimensional minimum-work protocol with discrete points equally spaced in time is discussed.

Designing multidimensional control protocols is more involved, since the spatial path through multidimensional control-parameter space must also be optimized. Additionally, it is typically impractical to sample the friction matrix over control-parameter space for more than a few control parameters. A few numerical methods have been developed to calculate minimum-work protocols, some relying on the generalized friction framework [102, 103] and others on importance sampling [104] or machine-learning approaches [105]. An approach based on the string method found in Ref. [103] is outlined below. The string method was originally developed to find minimum-energy paths between metastable states in configuration space [93].

The string method divides a control protocol into M discrete segments, with each node at a point in control-parameter space. The string is initialized as a straight line between control-parameter endpoints, then iteratively updated to solve a discrete version of the Euler-Lagrange equation (2.69). The string is updated from the nth to $(n+1)$th iteration by solving the set of linear equations for each component λ_α at each discrete point m along the string:

$$\lambda_\alpha^{n+1}(m) - \lambda_\alpha^n(m) = \Delta r \Bigg\{ \mathrm{D}^2 \lambda_\alpha^{n+1}(m) \tag{2.72}$$

$$+ \sum_{ijk} [\zeta^{-1}]_{\alpha k}^n(\boldsymbol{\lambda}(m)) \, \mathrm{D}\lambda_i^n(m) \, \mathrm{D}\lambda_j^n(m) \left[\partial_i \zeta_{kj}(\boldsymbol{\lambda}(m)) - \frac{1}{2} \partial_k \zeta_{ij}(\boldsymbol{\lambda}(m)) \right] \Bigg\}$$

where $\partial_i \zeta_{kj}(\boldsymbol{\lambda}) = \partial \zeta_{kj}(\boldsymbol{\lambda})/\partial \lambda_i$ is the partial derivative of the kjth component of the friction matrix with respect to λ_i (assuming all other control parameters are fixed) at control-

parameter value $\boldsymbol{\lambda}$, Δr is a parameter controlling the size of the string update, and

$$\mathrm{D}\lambda_\alpha(m) \equiv \frac{\lambda_\alpha(m+1) - \lambda_\alpha(m-1)}{2\delta t} \tag{2.73a}$$

$$\mathrm{D}^2\lambda_\alpha(m) \equiv \frac{\lambda_\alpha(m+1) + \lambda_\alpha(m-1) - 2\lambda_\alpha(m)}{\delta t^2} \tag{2.73b}$$

are finite-difference estimators of time derivatives for the string, with δt denoting the difference in scaled protocol time between adjacent string points. The string is repeatedly updated using Eq. (2.72) until it converges to user-specified precision.

The string update step requires estimation of the friction matrix and its derivative in each control-parameter direction for each point in the current iteration of the string. If these quantities can be estimated from short equilibrium simulations at each control-parameter value, the estimates can be obtained in parallel to reduce computational cost. The estimates from all string points are inserted into Eq. (2.72) to update the string. However, accurate calculation of the friction matrix and its derivative for a single control-parameter value can take significant sampling time. If the friction matrix has constant integral relaxation time for all components across control-parameter space, the force variance matrix $\zeta_{ij}(\boldsymbol{\lambda}) \approx \langle \delta X_j \delta X_i \rangle_{\boldsymbol{\lambda}}$ takes the place in Eq. (2.64) of the friction matrix [103]. In this case, the derivative of the force variance matrix in control-parameter space is $\partial_i \langle \delta X_j \delta X_k \rangle_{\boldsymbol{\lambda}} = \langle \delta X_i \delta X_j \delta X_k \rangle_{\boldsymbol{\lambda}}$. Estimating second and third cumulants from an equilibrium distribution take significantly less sampling time than estimating the integrated correlation function of the conjugate forces; if the constant-relaxation-time approximation holds, there is a significant reduction in computational cost for computing minimum-work protocols using this method. This approximation was used in [103] to design minimum-work protocols in a 100-dimensional control-parameter space.

Minimum-work protocols are calculated using the string method in Chapter 4, with some modifications to account for the varying integral relaxation time throughout control-parameter space for the chosen model system, that precludes using the force variance matrix alone to design minimum-work protocols.

Chapter 3

Information Thermodynamics of Transition Paths

The reaction coordinate describing a transition between reactant and product is a fundamental concept in the theory of chemical reactions. As discussed in Section 1.1.2, the committor is widely considered to be the "true" reaction coordinate, describing every aspect of the transition from $A \rightarrow B$. In this chapter, a novel information-theoretic justification of the committor as the reaction coordinate is derived. Additionally, it is shown that selecting the transition-path ensemble (the set of trajectories from reactant to product) from a long ergodic equilibrium trajectory results in entropy production that precisely equals the information generated by system dynamics about the reactivity of trajectories. The components of entropy production and information generation due to an arbitrary system coordinate are also equal; this reveals equivalent thermodynamic and information-theoretic measures of the suitability of low-dimensional collective variables that encode information relevant for describing reaction mechanisms, the need for which was discussed in Section 2.2. The committor is a single coordinate that preserves all system entropy production and distills all system information about reactivity, giving further support for its role as the reaction coordinate. The equality between entropy production and dynamical information generation also holds at the level of arbitrary individual coordinates, providing parallel measures of the coordinate's relevance to the reaction. The results derived in this chapter use ideas from the fields of information theory (Section 2.1.2), stochastic thermodynamics (Section 2.1.3), and transition-path theory (Section 2.2).

3.1 Information-theoretic perspective on the transition path ensemble

Consider a multidimensional system $\boldsymbol{\Phi}$ evolving according to Markovian dynamics governed by the Master equation [13], $\mathrm{d}_t p(\boldsymbol{\phi}) = \sum_{\boldsymbol{\phi}'} T_{\boldsymbol{\phi}\boldsymbol{\phi}'} p(\boldsymbol{\phi}')$, where $T_{\boldsymbol{\phi}\boldsymbol{\phi}'}$ is the transition rate from state $\boldsymbol{\phi}' \rightarrow \boldsymbol{\phi}$ and $p(\boldsymbol{\phi})$ is the probability of state $\boldsymbol{\phi}$. It is assumed that the transition

rates obey detailed balance [13] and the system is in equilibrium with its environment so that $p(\phi) = \pi(\phi)$, the equilibrium probability of ϕ. The transition-path ensemble (TPE) is studied, the set of trajectories that leave one subset of states $A \in \mathbf{\Phi}$ and next visit a distinct subset $B \in \mathbf{\Phi} \backslash A$ before returning to A. In most applications, A and B are metastable states separated by a dynamical barrier; following Refs.[30, 90] it is only assumed that A and B do not overlap and lack direct transitions; i.e., $T_{\phi\phi'} = 0$ for $\phi' \in A$ and $\phi \in B$.

The TPE is formed by selecting from a long ergodic equilibrium *supertrajectory* the trajectory segments that leave A and reach B before A. Transition paths are therefore selected based on the trajectory *outcome* S_+ [the next mesostate (A or B) visited by the system] and *origin* S_- (the mesostate most recently visited by the system). This partitions the supertrajectory into four trajectory subensembles, each with particular $\boldsymbol{s} \equiv (s_-, s_+)$: The forward (reverse) transition-path ensemble is the set of trajectory segments with $\boldsymbol{s} = (A, B)$ [$\boldsymbol{s} = (B, A)$], and the stationary subensemble from $A \to A$ ($B \to B$) has $\boldsymbol{s} = (A, A)$ [$\boldsymbol{s} = (B, B)$], as depicted in Fig. 3.1. Every trajectory segment in the forward TPE has a corresponding equally probable time-reversed trajectory segment in the reverse TPE.

At any time during the equilibrium supertrajectory, random variables $\mathbf{\Phi}$ and $\boldsymbol{S}$ are defined that respectively denote the current system state and trajectory subensemble, with $p(\phi, \boldsymbol{s})$ the joint distribution that the system is currently in state ϕ and is currently on a trajectory segment with respective origin and outcome $\boldsymbol{s} = \{s_-, s_+\}$. Since the system dynamics are Markovian, the trajectory outcome and origin are conditionally independent given current state ϕ, so the joint distribution can be factored as [30]

$$p(\phi, \boldsymbol{s}) = \pi(\phi)p(s_+|\phi)p(s_-|\phi) \,. \tag{3.1}$$

The conditional probabilities of trajectory outcome and origin given current state ϕ are

$$p(S_+ = B \,|\, \phi) = q^+(\phi) \tag{3.2a}$$

$$p(S_- = A \,|\, \phi) = q^-(\phi) \,. \tag{3.2b}$$

Here $q^+(\phi)$ is the forward *committor*, the probability that the system currently in state ϕ will next reach B before A, and $q^-(\phi)$ is the backward committor, the probability that the system (currently in ϕ) was more recently in mesostate A than in B. The committors obey boundary conditions $q^+(\phi) = 0$ and $q^-(\phi) = 1$ for $\phi \in A$, and $q^+(\phi) = 1$ and $q^-(\phi) = 0$ for $\phi \in B$. Since the system is in equilibrium and the transition rates obey detailed balance, $q^-(\phi) = 1 - q^+(\phi)$ [30], a single committor (without loss of generality, the forward committor $q^+(\phi)$) provides information about both the outcome and origin of the trajectory segment, so $q^+(\phi)$ is referred to as the reaction coordinate.

During the equilibrium supertrajectory, the system continually evolves from A to B and B to A, completing a unidirectional cycle through each trajectory subensemble with
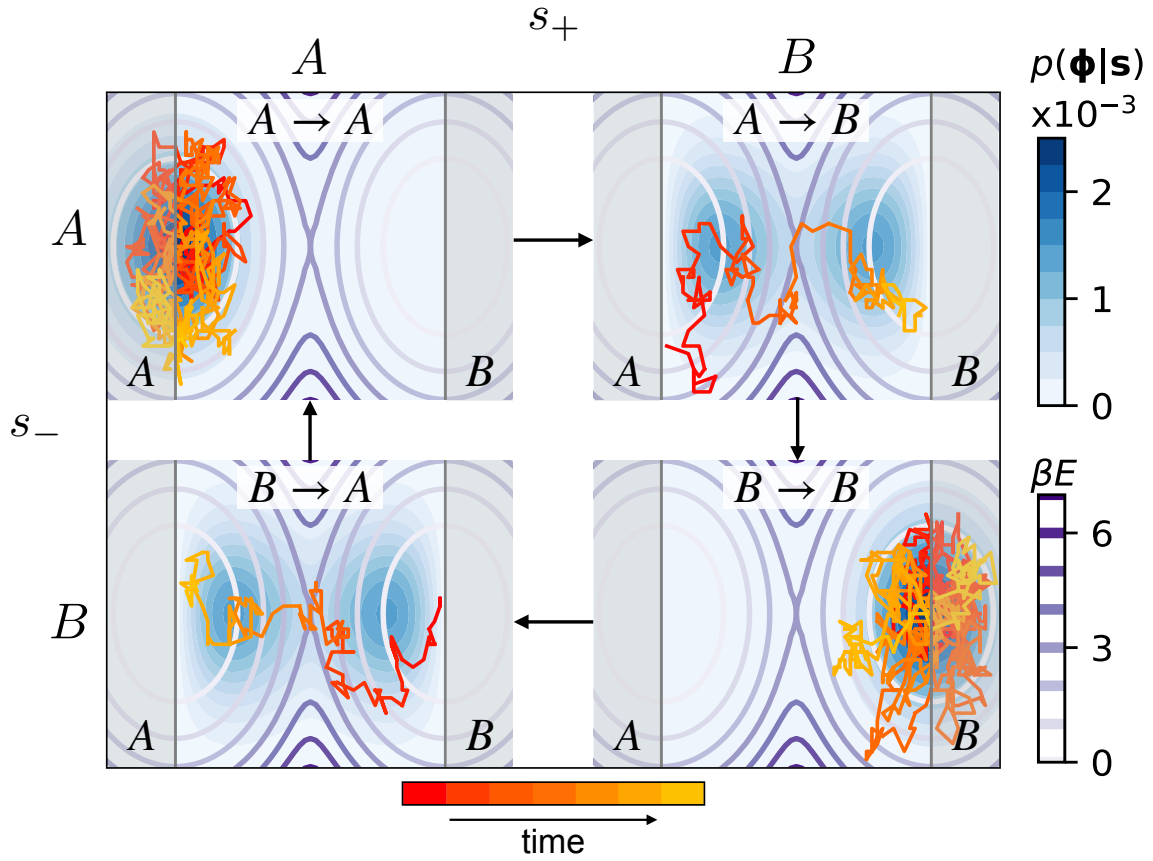
Figure 3.1: Partitioning a long ergodic equilibrium supertrajectory into subensembles based on trajectory outcome S_+ and origin S_-. Contours: example double-well potential energy βE_{int}. Shaded regions: A is the region to the left of the left energy minimum and B is the region to the right of the right minimum. Heat map: probability distribution $p(\boldsymbol{\phi}|\boldsymbol{s})$ of system state conditioned on trajectory subensemble $\boldsymbol{s}$. Solid curves: representative trajectory from each subensemble. The forward (reverse) TPE in the top-right (bottom-left) panel has net flux of trajectories from $A \to B$ ($B \to A$). The top-left (bottom-right) panel shows the stationary subensemble from $A \to A$ ($B \to B$).

stochastic transition times depending on underlying microscopic dynamics. Transition-path theory [90, 91, 29] derives quantitative properties (reaction rate and free-energy difference) of the $A \rightarrow B$ reaction from the equilibrium probability flux of subensemble transitions

$$\nu_{\boldsymbol{S}} = \sum_{\phi \notin A, \phi' \in A} T_{\phi\phi'} \pi(\phi') q^+(\phi) \tag{3.3}$$

and the respective marginal probabilities $p(s_+)$ and $p(s_-)$:

$$k_{AB} = \frac{\nu_{\boldsymbol{S}}}{p(S_- = A)} = \frac{\nu_{\boldsymbol{S}}}{p(S_+ = A)} \tag{3.4a}$$

$$k_{BA} = \frac{\nu_{\boldsymbol{S}}}{p(S_- = B)} = \frac{\nu_{\boldsymbol{S}}}{p(S_+ = B)} \tag{3.4b}$$

$$\beta \Delta F_{AB} = \ln \frac{p(S_- = A)}{p(S_- = B)} = \ln \frac{p(S_+ = A)}{p(S_+ = B)} , \tag{3.4c}$$

where k_{AB} (k_{BA}) is the rate constant for the $A \rightarrow B$ ($B \rightarrow A$) transition and $\Delta F_{AB} \equiv F_B - F_A$ is the free-energy difference. The second equality in each line uses $p(S_- = A) = p(S_+ = A)$ which follows from detailed balance. Mesoscopic reaction properties are thus derived from information about the subensembles, specifically the proportion of time spent in each subensemble and how frequently the subensemble switches.

Therefore, the reaction coordinate should be maximally informative about the current subensemble. This is precisely quantified by mutual information, a non-linear statistical measure of the relationship between two random variables, specifically quantifying the reduction of uncertainty (given by Shannon entropy $H(X) \equiv -\sum_x p(x) \ln p(x)$) about one random variable from measuring another [83]:

$$I(\boldsymbol{S}; \boldsymbol{\Phi}) \equiv \sum_{\phi, s} p(\phi, s) \ln \frac{p(\phi, s)}{\pi(\phi) p(s)} , \tag{3.5}$$

where $p(\boldsymbol{s}) = \sum_\phi p(\phi, s)$ is the marginal probability that the system is currently on a trajectory segment with outcome and origin $\boldsymbol{s} = (s_-, s_+)$. Operationally, $p(\boldsymbol{s})$ can be estimated from the proportion of time $\tau_{\boldsymbol{s}}$ spent in subensemble $\boldsymbol{s}$ during an equilibrium supertrajectory of length τ, $p(\boldsymbol{s}) = \lim_{\tau \to \infty} \tau_{\boldsymbol{s}}/\tau$. The mutual information between trajectory outcome and origin and a general state function $f(\phi)$ is constrained by the data processing inequality [83]

such that $I(\boldsymbol{S}; \boldsymbol{\Phi}) \geq I(\boldsymbol{S}; f(\boldsymbol{\Phi}))$. The committor satisfies this equality:

$$I(\boldsymbol{S}; \boldsymbol{\Phi}) = \sum_{\phi, s} \pi(\phi) p(s|\phi) \ln \frac{p(s|\phi)}{p(s)} \tag{3.6a}$$

$$= \sum_{\phi, s} \sum_{q_o} \pi(\phi) p(s|\phi) \ln \frac{p(s|\phi)}{p(s)} \delta[q^+(\phi) - q_o] \tag{3.6b}$$

$$= \sum_{q_o, s} p(s|q_o) \ln \frac{p(s|q_o)}{p(s)} \left[\sum_{\phi} \pi(\phi) \delta[q^+(\phi) - q_o] \right] \tag{3.6c}$$

$$= \sum_{q_o, s} \pi(q_o) p(s|q_o) \ln \frac{p(s|q_o)}{p(s)} \tag{3.6d}$$

$$= \sum_{q_o, s} p(q_o, s) \ln \frac{p(q_o, s)}{\pi(q_o) p(s)} \tag{3.6e}$$

$$= I(\boldsymbol{S}; q^+(\boldsymbol{\Phi})) . \tag{3.6f}$$

In (3.6a), the mutual information is expressed in terms of conditional probability $p(\boldsymbol{s}|\phi)$. In (3.6b), a sum over all discrete committor values appearing in the system is included, allowing separation of all states with the same committor value in Eq. (3.6c), where it is also recognized that the probability the system is in a particular subensemble given state ϕ depends only on its committor value, $p(\boldsymbol{s}|\phi) = p(\boldsymbol{s}|q^+(\phi))$. This allows the mutual information between trajectory subensemble and system state to be expressed as the mutual information between trajectory subensemble and the committor of system state in Eq. (3.6f). Since $I(\boldsymbol{S}; \boldsymbol{\Phi}) = I(\boldsymbol{S}; q^+(\boldsymbol{\Phi}))$, the committor is a sufficient statistic for the mutual information between trajectory subensemble and full system state [83]. In this sense, the committor is the "optimal" reaction coordinate, since it is maximally informative about the trajectory subensemble given a measurement of system state. This is the first major result.

3.1.1 Physical origin of subensemble random variables

Physically, the trajectory outcome and origin (and hence the committors) represent uncertainty in the state of the environment. Classical mechanics assumes a constant-energy universe (system $\boldsymbol{\Phi}$ plus environment $\boldsymbol{\Psi}$) governed by deterministic dynamics so that the outcome and origin of the trajectory initiated from a given state of system *and* environment are deterministic (and can be determined by integrating the state of the universe forward and backward in time until the system reaches A or B), i.e., $p(\boldsymbol{s}|\phi, \psi)$ is either 0 or 1. This partitions the state space of the universe into four quadrants corresponding to each trajectory subensemble, with each state (ϕ, ψ) belonging to only one subensemble; thus the uncertainty about the trajectory subensemble given a state of the universe is zero,

$$H(\boldsymbol{S}|\boldsymbol{\Phi}, \boldsymbol{\Psi}) \equiv - \sum_{\phi, \psi, s} p(\phi, \psi, s) \ln p(\boldsymbol{s}|\phi, \psi) = 0 . \tag{3.7}$$

In this case, the mutual information between the universe and trajectory subensemble is the uncertainty about the trajectory subensemble,

$$I(\boldsymbol{S}; \boldsymbol{\Phi}, \boldsymbol{\Psi}) = H(\boldsymbol{S}) - H(\boldsymbol{S}|\boldsymbol{\Phi}, \boldsymbol{\Psi}) = H(\boldsymbol{S}) \,, \tag{3.8}$$

thus measurement of the state of the universe fully determines the trajectory outcome and origin.

However, the microstate of the environment is not typically resolved, instead its interaction with the system is coarse-grained into friction and fluctuations [13]. Measurement of the system state alone does not fully determine the trajectory outcome and origin, which become random variables with positive conditional Shannon entropy

$$H(\boldsymbol{S}|\boldsymbol{\Phi}) \equiv -\sum_{s,\phi} p(\phi, s) \ln p(s|\phi) > 0 \tag{3.9}$$

reflecting uncertainty in the state of the environment that is relevant to classification of the current subensemble.

3.2 Joint system and subensemble dynamics and irreversible entropy production

The main utility of introducing random variables for the trajectory outcome and origin is that reactive events are now part of an ergodic trajectory that unidirectionally cycles through subensembles and forms a nonequilibrium steady state, allowing the use of tools and concepts from stochastic thermodynamics outlined in Section 2.1.3 to discuss the statistics of reactive trajectories. In particular, the entropy production of the transition-path ensemble is now derived and its equivalence with information generated about the trajectory outcome S_+ is proved.

The joint dynamics of $(\boldsymbol{\Phi}, \boldsymbol{S})$ are described using the Master equation

$$\mathrm{d}_t p(\phi, s) = \sum_{\phi', s'} T_{\phi\phi'}^{ss'} \, p(\phi', s') \tag{3.10}$$

where $T_{\phi\phi'}^{ss'}$ is the $(\phi', s') \to (\phi, s)$ transition rate. Conditional transition rates for the system in the $A \to B$ subensemble were derived in Ref. [90], and they are adapted for all subensembles below.

Following [90], the transition probability over time dt for a $\phi' \to \phi$ transition given the trajectory remains in subensemble s is

$$T^{s=s'}_{\phi\phi'} dt = p(\phi, s_+, s_- | \phi', s'_+, s'_-) \tag{3.11a}$$

$$= p(s_+, s_- | \phi, \phi', s'_+, s'_-) \, p(\phi | \phi', s'_+, s'_-) \tag{3.11b}$$

$$= p(\phi | \phi', s'_+, s'_-) \tag{3.11c}$$

$$= p(\phi | \phi', s'_+) \tag{3.11d}$$

$$= \frac{p(\phi, s'_+ | \phi')}{p(s'_+ | \phi')} \tag{3.11e}$$

$$= \frac{p(s'_+ | \phi) \, p(\phi | \phi')}{p(s'_+ | \phi')} \, . \tag{3.11f}$$

Equation (3.11b) splits the joint probability into conditional and marginal probabilities. In (3.11c), $p(s_+, s_- | \phi, \phi', s'_+, s'_-) = 1$ when the subensemble does not change. In (3.11d), the Markov property is used to eliminate the dependence of the next state on trajectory origin. In (3.11e), the conditional probability is expressed as the ratio of joint and marginal probabilities. In (3.11f), it is recognized that the trajectory outcome depends only on the most recent state ϕ and drop the dependency on ϕ'. Finally, recalling that $s_+ = s'_+$ for these transitions and re-expressing the transition probability as a transition rate, the rate of a $\phi' \to \phi$ transition is subensemble s is

$$T^{s=s'}_{\phi\phi'} = \frac{p(s_+ | \phi)}{p(s_+ | \phi')} T_{\phi\phi'} \, . \tag{3.12}$$

Transition rates for the four sets of transitions that change subensemble are similarly derived, where either the trajectory origin or outcome changes while the other is constant. The transition probability in time dt for a $\phi' \to \phi$ transition out of A where the subensemble changes from $s' = (A, A) \to s = (A, B)$ is:

$$T^{s=(A,B),s'=(A,A)}_{\phi\phi'} dt = p(\phi, S_+ = B, S_- = A | \phi', S'_+ = A, S'_- = A) \tag{3.13a}$$

$$= p(S_+ = B, S_- = A | \phi, \phi', S'_+ = A, S'_- = A) \, p(\phi | \phi', S'_+ = A, S'_- = A) \tag{3.13b}$$

$$= p(S_+ = B | \phi, \phi', S'_+ = A, S'_- = A) \, p(\phi | \phi', S'_+ = A, S'_- = A) \tag{3.13c}$$

$$= p(S_+ = B | \phi) \, p(\phi | \phi') \, . \tag{3.13d}$$

(3.13c) uses the fact that $S_- = A$ when $\phi' \in A$. (3.13d) uses the Markov property to simplify the conditional distributions. The transition probability is then re-expressed as a transition rate

$$T^{s=(A,B),s'=(A,A)}_{\phi\phi'} = p(S_+ = B | \phi) \, T_{\phi\phi'} \, . \tag{3.14}$$

41

Similarly, the transition rate for a $\phi' \to \phi$ transition where the subensemble outcome changes from $\boldsymbol{s}' = (B, B) \to \boldsymbol{s}' = (B, A)$ is

$$T_{\phi\phi'}^{\boldsymbol{s}=(B,A),\boldsymbol{s}'=(B,B)} = p(S_+ = A|\phi)\, T_{\phi\phi'} \,. \tag{3.15}$$

The trajectory origin changes when the system finishes a transition path at the boundary of A or B. The probability for a $\phi' \to \phi$ transition into A where the subensemble changes from $\boldsymbol{s}' = (B, A) \to \boldsymbol{s} = (A, A)$ is:

$$T_{\phi\phi'}^{\boldsymbol{s}=(A,A),\boldsymbol{s}'=(B,A)}\mathrm{d}t = p(\phi, S_+ = A, S_- = A|\phi', S_+' = A, S_-' = B) \tag{3.16a}$$

$$= p(S_+ = A, S_- = A|\phi, \phi', S_+' = A, S_-' = B)\, p(\phi|\phi', S_+' = A, S_-' = B) \tag{3.16b}$$

$$= p(\phi|\phi', S_+' = A) \tag{3.16c}$$

$$= \frac{p(\phi, S_+' = A|\phi')}{p(S_+' = A|\phi')} \tag{3.16d}$$

$$= \frac{p(S_+' = A|\phi, \phi')\, p(\phi|\phi')}{p(S_+' = A|\phi')} \tag{3.16e}$$

$$= \frac{p(\phi|\phi')}{p(S_+' = A|\phi')} \,. \tag{3.16f}$$

(3.16c) recognizes that $S_+ = A$ and $S_- = A$ for $\phi \in A$ and uses the Markov property to eliminate dependence on S_-' in $p(\phi|\phi', S_+' = A, S_-' = A)$. (3.16f) uses $S_+' = A$ for $\phi \in A$. Finally, the transition probability is re-expressed as the transition rate

$$T_{\phi\phi'}^{\boldsymbol{s}=(A,A),\boldsymbol{s}'=(B,A)} = \frac{1}{p(S_+ = A|\phi')}T_{\phi\phi'} \,, \tag{3.17}$$

and similarly derive the transition rate for a $\phi' \to \phi$ transition where the system enters B and finishes a forward TPE ($\boldsymbol{s}' = (A, B) \to \boldsymbol{s} = (B, B)$) as

$$T_{\phi\phi'}^{\boldsymbol{s}=(B,B),\boldsymbol{s}'=(A,B)} = \frac{1}{p(S_+ = B|\phi')}T_{\phi\phi'} \,. \tag{3.18}$$

When averaged over all such transitions, these transition rates yield the unidirectional probability flux $\nu_{\boldsymbol{S}}$ between trajectory subensembles (3.4),

$$\nu_{\boldsymbol{S}} = \sum_{\boldsymbol{\phi} \notin A, \boldsymbol{\phi}' \in A} p(S_+ = B|\boldsymbol{\phi})\, T_{\boldsymbol{\phi}\boldsymbol{\phi}'} \pi(\boldsymbol{\phi}') \tag{3.19a}$$

$$= \sum_{\boldsymbol{\phi} \notin B, \boldsymbol{\phi}' \in B} p(S_+ = A|\boldsymbol{\phi})\, T_{\boldsymbol{\phi}\boldsymbol{\phi}'} \pi(\boldsymbol{\phi}') \tag{3.19b}$$

$$= \sum_{\boldsymbol{\phi} \in A, \boldsymbol{\phi}' \notin A} T_{\boldsymbol{\phi}\boldsymbol{\phi}'} \pi(\boldsymbol{\phi}')\, p(S_- = B|\boldsymbol{\phi}') \tag{3.19c}$$

$$= \sum_{\boldsymbol{\phi} \in B, \boldsymbol{\phi}' \notin B} T_{\boldsymbol{\phi}\boldsymbol{\phi}'} \pi(\boldsymbol{\phi}')\, p(S_- = A|\boldsymbol{\phi}')\,. \tag{3.19d}$$

Note that each RHS of (3.19) has an implicit conditional probability for the other element of the trajectory subensemble, each of which equals unity on the relevant system subspace, e.g., $p(S_- = A|\boldsymbol{\phi}') = 1$ for $\boldsymbol{\phi}' \in A$ in (3.19a).

To summarize, the transition rates for joint system-subensemble dynamics are:

$$T_{\boldsymbol{\phi}\boldsymbol{\phi}'}^{\boldsymbol{s}\boldsymbol{s}'} = \begin{cases} T_{\boldsymbol{\phi}\boldsymbol{\phi}'}^{\boldsymbol{s}} \equiv T_{\boldsymbol{\phi}\boldsymbol{\phi}'} \dfrac{p(s_+|\boldsymbol{\phi})}{p(s_+|\boldsymbol{\phi}')}, & \boldsymbol{s}' = \boldsymbol{s} \\[2ex] T_{\boldsymbol{\phi}\boldsymbol{\phi}'} p(S_+ = B|\boldsymbol{\phi}), & \begin{cases} \boldsymbol{\phi}' \in A\,, \boldsymbol{\phi} \notin A\,, \\ \boldsymbol{s}' = (A,A)\,, \boldsymbol{s} = (A,B) \end{cases} \\[3ex] T_{\boldsymbol{\phi}\boldsymbol{\phi}'} p(S_+ = A|\boldsymbol{\phi}), & \begin{cases} \boldsymbol{\phi}' \in B\,, \boldsymbol{\phi} \notin B\,, \\ \boldsymbol{s}' = (B,B)\,, \boldsymbol{s} = (B,A) \end{cases} \\[3ex] T_{\boldsymbol{\phi}\boldsymbol{\phi}'} / p(S_+ = A|\boldsymbol{\phi}'), & \begin{cases} \boldsymbol{\phi}' \notin A\,, \boldsymbol{\phi} \in A\,, \\ \boldsymbol{s}' = (B,A)\,, \boldsymbol{s} = (A,A) \end{cases} \\[3ex] T_{\boldsymbol{\phi}\boldsymbol{\phi}'} / p(S_+ = B|\boldsymbol{\phi}'), & \begin{cases} \boldsymbol{\phi}' \notin B\,, \boldsymbol{\phi} \in B\,, \\ \boldsymbol{s}' = (A,B)\,, \boldsymbol{s} = (B,B) \end{cases} \\[3ex] - \sum_{\substack{\boldsymbol{\phi}'' \neq \boldsymbol{\phi}' \\ \boldsymbol{s}'' \neq \boldsymbol{s}'}} T_{\boldsymbol{\phi}''\boldsymbol{\phi}'}^{\boldsymbol{s}''\boldsymbol{s}'}, & \boldsymbol{\phi} = \boldsymbol{\phi}'\,, \boldsymbol{s} = \boldsymbol{s}' \\[3ex] 0 & \text{otherwise} \end{cases} \tag{3.20}$$

The top transition does not change the subensemble, and biases transitions within subensemble $\boldsymbol{s}$ toward states with higher probability of trajectory outcome s_+. The middle four transitions switch subensembles and are unidirectional, contributing to the probability flux $\nu_{\boldsymbol{S}}$ (Eqs. (3.3), (3.19)). The $(\boldsymbol{\phi}', \boldsymbol{s}') \to (\boldsymbol{\phi}, \boldsymbol{s})$ transition rate is defined such that the probability distribution remains normalized under time evolution (Eq. 2.9). These joint dynamics are Markovian: since the underlying system dynamics are Markovian, the transition rates (3.20) do not depend on the trajectory origin s_-, and the outcome s_+ does not induce dependence

on earlier system states. Considered alone, system dynamics are at equilibrium and microscopically reversible; adding the trajectory outcome and origin variables (that are not functions of system state and explicitly depend on the past and future) breaks time-reversal symmetry, producing absolutely irreversible trajectory-subensemble transitions and time-asymmetric system transitions within a given subensemble.

3.2.1 Time-asymmetry of transition-path dynamics

To quantify the time-asymmetry for a particular $\phi' \to \phi$ transition in subensemble s, (3.20), Bayes' rule, and the equilibrium detailed-balance relation $T_{\phi\phi'}\pi(\phi') = T_{\phi'\phi}\pi(\phi)$ are combined to derive a local detailed-balance relation,

$$\frac{T_{\phi\phi'}^s p(\phi'|s)}{T_{\phi'\phi}^s p(\phi|s)} = \frac{p(s_-|\phi')p(s_+|\phi)}{p(s_-|\phi)p(s_+|\phi')} \, . \tag{3.21}$$

The $A \to A$ (and analogously $B \to B$) stationary subensemble has $s_+ = s_- = A$, and due to system detailed balance $p(S_- = A|\phi) = p(S_+ = A|\phi)$, so the RHS is unity and detailed balance holds for transitions within stationary subensembles. The reactive subensembles (forward or reverse TPE) have different trajectory outcome and origin so the RHS side differs from unity, leading to a detailed-balance-breaking flux (and hence entropy production) along particular transitions within these subensembles.

Within a fixed subensemble s, the net trajectory flux is

$$J_{\phi\phi'}^s = T_{\phi\phi'}^s p(\phi', s) - T_{\phi'\phi}^s p(\phi, s) \tag{3.22a}$$

$$= \left[p(s_-|\phi')p(s_+|\phi) - p(s_-|\phi)p(s_+|\phi') \right] T_{\phi\phi'}\pi(\phi') \, . \tag{3.22b}$$

The second equality follows from $p(\phi, s) = p(s|\phi)\pi(\phi)$; the conditional independence of s_+ and s_- given state ϕ, i.e., $p(s_+, s_-|\phi) = p(s_+|\phi)p(s_-|\phi)$; and substitution for $T_{\phi'\phi}^s$ using (3.20). The stationary subensembles ($A \to A$ and $B \to B$) have no net flux because each trajectory segment and its time-reversed counterpart occur at equal rates within the same subensemble. In contrast, the forward and reverse TPEs have net trajectory flux since each transition path and its time-reversed counterpart occur in different subensembles.

3.2.2 Entropy production for joint dynamics

Time-asymmetry within the transition-path ensembles implies that there is irreversible entropy production for dynamics in these subensembles. The change in joint entropy $H(\mathbf{\Phi}, \boldsymbol{S}) \equiv$

$-\sum_{\phi,s} p(\phi,s)\ln p(\phi,s)$ is decomposed into three terms [106]

$$0 = \mathrm{d}_t H(\mathbf{\Phi},\mathbf{S}) = \underbrace{\sum_{\phi,\phi',s} T^s_{\phi\phi'} p(\phi',s)\ln\frac{T^s_{\phi\phi'} p(\phi',s)}{T^s_{\phi'\phi} p(\phi,s)}}_{\dot{H}^{\mathrm{irr}}(\mathbf{\Phi},\mathbf{S})} - \underbrace{\sum_{\phi,\phi',s} T^s_{\phi\phi'} p(\phi',s)\ln\frac{T^s_{\phi\phi'}}{T^s_{\phi'\phi}}}_{\dot{H}^{\mathrm{env}}(\mathbf{\Phi},\mathbf{S})}$$
$$+ \underbrace{\sum_{\phi,\phi',s\neq s'} T^{ss'}_{\phi\phi'} p(\phi',s')\ln\frac{p(\phi',s')}{p(\phi,s)}}_{\dot{H}^{\mathrm{sub}}(\mathbf{\Phi},\mathbf{S})}\,, \tag{3.23}$$

where $\dot{H}^{\mathrm{irr}}(\mathbf{\Phi},\mathbf{S})$ is the irreversible entropy production and $\dot{H}^{\mathrm{env}}(\mathbf{\Phi},\mathbf{S})$ the environmental entropy change for transitions that do not change the subensemble, and $\dot{H}^{\mathrm{sub}}(\mathbf{\Phi},\mathbf{S})$ is the change in joint entropy due to transitions that change the subensemble.

To simplify the entropy change for transitions that change the subensemble, the joint transition rates (Eq. 3.20) for these four sets of transitions are inserted into the third term of Eq. (3.23) to obtain:

$$\dot{H}^{\mathrm{sub}}(\mathbf{\Phi},\mathbf{S}) = \sum_{\phi\in A,\phi'\notin A} T_{\phi\phi'}\pi(\phi')p(S_-=B|\phi')\ln\frac{p(\phi',\mathbf{S}=(B,A))}{p(\phi,\mathbf{S}=(A,A))} \tag{3.24a}$$
$$+ \sum_{\phi\notin A,\phi'\in A} T_{\phi\phi'}\pi(\phi')p(S_+=B|\phi)\ln\frac{p(\phi',\mathbf{S}=(A,A))}{p(\phi,\mathbf{S}=(A,B))}$$
$$+ \sum_{\phi\in B,\phi'\notin B} T_{\phi\phi'}\pi(\phi')p(S_-=A|\phi')\ln\frac{p(\phi',\mathbf{S}=(A,B))}{p(\phi,\mathbf{S}=(B,B))}$$
$$+ \sum_{\phi\notin B,\phi'\in B} T_{\phi\phi'}\pi(\phi')p(S_+=A|\phi)\ln\frac{p(\phi',\mathbf{S}=(B,B))}{p(\phi,\mathbf{S}=(B,A))}$$
$$= \sum_{\phi\notin A,\phi'\in A} T_{\phi'\phi}\pi(\phi)p(S_-=B|\phi)\ln\frac{p(\phi,\mathbf{S}=(B,A))}{\pi(\phi')} \tag{3.24b}$$
$$- \sum_{\phi\notin A,\phi'\in A} T_{\phi\phi'}\pi(\phi')p(S_+=B|\phi)\ln\frac{p(\phi,\mathbf{S}=(A,B))}{\pi(\phi')}$$
$$+ \sum_{\phi\notin B,\phi'\in B} T_{\phi'\phi}\pi(\phi)p(S_-=A|\phi)\ln\frac{p(\phi,\mathbf{S}=(A,B))}{\pi(\phi')}$$
$$- \sum_{\phi\notin B,\phi'\in B} T_{\phi\phi'}\pi(\phi')p(S_+=A|\phi)\ln\frac{p(\phi,\mathbf{S}=(B,A))}{\pi(\phi')}$$
$$= 0\,. \tag{3.24c}$$

In (3.24a), the indices in the first and third terms are rewritten, and the simplifications $p(\phi,\mathbf{S}=(A,A))=\pi(\phi)$ for $\phi\in A$ and $p(\phi,\mathbf{S}=(B,B))=\pi(\phi)$ for $\phi\in B$ are used. The first and second terms and the third and fourth terms cancel because of the equilibrium relationships $p(S_+=A|\phi)=p(S_-=A|\phi)$, $p(S_+=B|\phi)=p(S_-=B|\phi)$, $p(\phi,\mathbf{S}=$

$(B, A)) = p(\phi, \mathbf{S} = (A, B))$, and $T_{\phi\phi'}\pi(\phi') = T_{\phi'\phi}\pi(\phi)$. Therefore subensemble transitions do not change the joint entropy.

Substituting in the joint transition rates (3.20) for dynamics within a subensemble $T_{\phi\phi'}^{s} = T_{\phi\phi'}p(s_+|\phi)/p(s_+|\phi')$, the environmental entropy change is

$$\dot{H}^{\text{env}}(\mathbf{\Phi}, \mathbf{S}) = \sum_{\phi,\phi',s} T_{\phi\phi'}^{s} p(\phi', s) \left[\ln \frac{T_{\phi\phi'}}{T_{\phi'\phi}} + 2 \ln \frac{p(s_+|\phi)}{p(s_+|\phi')} \right] \tag{3.25a}$$

$$= 2 \sum_{\phi,\phi',s} T_{\phi\phi'}^{s} p(\phi', s) \ln \frac{p(s_+|\phi)}{p(s_+|\phi')} \tag{3.25b}$$

$$= 2 \sum_{\phi,\phi',s_+} T_{\phi\phi'}^{s_+} p(\phi', s_+) \ln \frac{p(s_+|\phi)}{p(s_+|\phi')} \tag{3.25c}$$

$$= 2 \dot{I}^{\mathbf{\Phi}}(S_+; \mathbf{\Phi}) , \tag{3.25d}$$

where $\dot{I}^{\mathbf{\Phi}}(S_+; \mathbf{\Phi}) \geq 0$ is the rate of change in mutual information between the trajectory outcome and system state due to system dynamics in a fixed subensemble [107]. Eq. (3.25b) is obtained by summing over $\mathbf{s}$ to cancel contributions from the equilibrium transition rates, and (3.25c) is obtained by summing over the trajectory origin S_-.

Finally, the irreversible entropy production $\dot{H}^{\text{irr}}(\mathbf{\Phi}, \mathbf{S})$ of the cycle is expressed in terms of the irreversible entropy production of dynamics given fixed subensemble $\mathbf{s}$, $\dot{\Sigma}_{\mathbf{s}}$, which quantifies the time irreversibility of system dynamics *within* that subensemble [108]:

$$\dot{H}^{\text{irr}}(\mathbf{\Phi}, \mathbf{S}) = \sum_{\mathbf{s}} p(\mathbf{s}) \sum_{\phi,\phi'} T_{\phi\phi'}^{s} p(\phi'|\mathbf{s}) \ln \frac{T_{\phi\phi'}^{s} p(\phi'|\mathbf{s})}{T_{\phi'\phi}^{s} p(\phi|\mathbf{s})} \tag{3.26a}$$

$$= \sum_{\mathbf{s}} p(\mathbf{s}) \dot{\Sigma}_{\mathbf{s}} \tag{3.26b}$$

$$\equiv \langle \dot{\Sigma} \rangle , \tag{3.26c}$$

where $\langle \dot{\Sigma} \rangle \equiv \sum_{\mathbf{s}} p(\mathbf{s}) \dot{\Sigma}_{\mathbf{s}} \geq 0$ is the subensemble-weighted average of the irreversible entropy production rate within each subensemble.

Rearranging (3.23) gives:

$$0 \leq \langle \dot{\Sigma} \rangle = 2 \dot{I}^{\mathbf{\Phi}}(S_+; \mathbf{\Phi}) \tag{3.27a}$$

$$= \dot{I}^{\mathbf{\Phi}}(S_+; \mathbf{\Phi}) - \dot{I}^{\mathbf{\Phi}}(S_-; \mathbf{\Phi}) , \tag{3.27b}$$

where $\dot{I}^{\mathbf{\Phi}}(S_-; \mathbf{\Phi}) \leq 0$ is the rate of change in mutual information between trajectory origin and system state due to $\mathbf{\Phi}$ dynamics in a fixed subensemble. The rate of change in mutual information between system state and trajectory *outcome* due to $\mathbf{\Phi}$ dynamics is of equal magnitude and opposite sign from the rate of change in mutual information between system

state and trajectory *origin* due to $\boldsymbol{\Phi}$ dynamics:

$$\dot{I}^{\boldsymbol{\Phi}}(S_+; \boldsymbol{\Phi}) = \sum_{\phi,\phi',s_+} T^{s_+}_{\phi\phi'} p(\phi', s_+) \ln \frac{p(s_+|\phi)}{p(s_+|\phi')} \tag{3.28a}$$

$$= \sum_{\phi,\phi',s_+} T_{\phi\phi'} \pi(\phi') p(s_+|\phi) \ln \frac{p(s_+|\phi)}{p(s_+|\phi')} \tag{3.28b}$$

$$= \sum_{\phi,\phi',s_-} T_{\phi\phi'} \pi(\phi') p(s_-|\phi) \ln \frac{p(s_-|\phi)}{p(s_-|\phi')} \tag{3.28c}$$

$$= \sum_{\phi,\phi',s_-} T_{\phi'\phi} \pi(\phi) p(s_-|\phi) \ln \frac{p(s_-|\phi)}{p(s_-|\phi')} \tag{3.28d}$$

$$= - \sum_{\phi,\phi',s_-} T_{\phi'\phi} p(\phi, s_-) \ln \frac{p(s_-|\phi')}{p(s_-|\phi)} \tag{3.28e}$$

$$= -\dot{I}^{\boldsymbol{\Phi}}(S_-; \boldsymbol{\Phi}) . \tag{3.28f}$$

In (3.28b) the joint transition rate is expanded using the subensemble-specific transition rates (3.20); (3.28c) relates outcome and origin probabilities using equilibrium relations $p(S_+ = A|\phi) = p(S_- = A|\phi)$, $p(S_+ = B|\phi) = p(S_- = B|\phi)$; and (3.28d) uses detailed balance, $T_{\phi\phi'}\pi(\phi') = T_{\phi'\phi}\pi(\phi)$.

These information rates reflect the dependence of the trajectory outcome and origin variables on the past and future states of the system: as the system evolves, uncertainty about the outcome S_+ diminishes, increasing the information the current system state carries about S_+, while uncertainty (given current system state $\boldsymbol{\Phi}$) about the origin S_- increases, decreasing information $\boldsymbol{\Phi}$ carries about S_-.

Since the stationary subensembles have no net flux, they also have no entropy production ($\dot{\Sigma}_{\boldsymbol{s}=(A,A)} = \dot{\Sigma}_{\boldsymbol{s}=(B,B)} = 0$), hence (3.27a) reduces to an equation for a single subensemble,

$$0 \leq p_{\mathrm{R}} \dot{\Sigma}_{\mathrm{R}} = \dot{I}^{\boldsymbol{\Phi}}(S_+; \boldsymbol{\Phi}) . \tag{3.29}$$

This equates the rate $\dot{I}^{\boldsymbol{\Phi}}(S_+; \boldsymbol{\Phi})$ of generating information about the outcome with the product of the entropy production rate of a reactive subensemble $\dot{\Sigma}_{\mathrm{R}} = \dot{\Sigma}_{\boldsymbol{s}=(A,B)} = \dot{\Sigma}_{\boldsymbol{s}=(B,A)}$ and that subensemble's marginal probability $p_{\mathrm{R}} = p(\boldsymbol{S} = (A,B)) = p(\boldsymbol{S} = (B,A))$. Although the supertrajectory is at equilibrium with no entropy production, $\dot{\Sigma}_{\mathrm{R}}$ physically represents the dissipation that would be necessary in a system evolving according to the TPE's detailed-balance-breaking transition rates (top line of (3.20) for $s_- \neq s_+$). Equation (3.29) is the second major result: the entropy production in a reactive subensemble equals the information generated about the reactivity of trajectories.

3.2.3 Thermodynamic metric

The TPE entropy production (3.29) can also be cast geometrically, by deriving a Fisher-information metric $\mathcal{I}(\phi)$ that imposes an information geometry on the state space [109, 110]. Suppose the state space $\mathbf{\Phi}$ is continuous and represented by a set of coordinates $\{\phi_i\}$. The Master equation $d_t p(\phi) = \sum_{\phi'} T_{\phi\phi'} p(\phi')$ represents a discrete approximation of the continuous dynamics [13]. The difference in mutual information rates (3.27b) is rearranged to obtain the transition-weighted relative entropy [83]

$$D[p(s|\phi')||p(s|\phi)] \equiv \sum_s p(s|\phi') \ln \frac{p(s|\phi')}{p(s|\phi)} \tag{3.30}$$

between the conditional subensemble distributions $p(s|\phi')$ and $p(s|\phi)$ before and after the transition, respectively, then expand in small state changes $\phi - \phi'$:

$$p_R \dot{\Sigma}_R = \dot{I}^{\mathbf{\Phi}}(S_+; \mathbf{\Phi}) - \dot{I}^{\mathbf{\Phi}}(S_-; \mathbf{\Phi}) \tag{3.31a}$$

$$= \sum_{\phi,\phi',s_+} T_{\phi\phi'}\pi(\phi')p(s_+|\phi) \ln \frac{p(s_+|\phi)}{p(s_+|\phi')} - \sum_{\phi,\phi',s_-} T_{\phi'\phi}\pi(\phi)p(s_-|\phi) \ln \frac{p(s_-|\phi')}{p(s_-|\phi)} \tag{3.31b}$$

$$= \sum_{\phi,\phi',s_+,s_-} T_{\phi\phi'}\pi(\phi')p(s_+|\phi)p(s_-|\phi) \ln \frac{p(s_+|\phi)}{p(s_+|\phi')}$$

$$- \sum_{\phi,\phi',s_+,s_-} T_{\phi'\phi}\pi(\phi)p(s_-|\phi)p(s_+|\phi) \ln \frac{p(s_-|\phi')}{p(s_-|\phi)} \tag{3.31c}$$

$$= \sum_{\phi,\phi',s_+,s_-} T_{\phi\phi'}\pi(\phi')p(s_+|\phi)p(s_-|\phi) \ln \frac{p(s_+|\phi)p(s_-|\phi)}{p(s_+|\phi')p(s_-|\phi')} \tag{3.31d}$$

$$= \sum_{\phi,\phi'} T_{\phi\phi'}\pi(\phi')D[p(s|\phi')||p(s|\phi)] \tag{3.31e}$$

$$\approx \sum_{\phi,\phi'} T_{\phi\phi'}\pi(\phi')\tfrac{1}{2}\sum_{i,j}(\phi_i - \phi'_i)\mathcal{I}_{ij}(\phi')(\phi_j - \phi'_j) \,. \tag{3.31f}$$

In (3.31c), each term is multiplied by unity ($\sum_{s_-} p(s_-|\phi)$ and $\sum_{s_+} p(s_+|\phi)$ respectively), then detailed balance ($T_{\phi\phi'}\pi(\phi') = T_{\phi'\phi}\pi(\phi)$) is used to combine terms in (3.31d). The ith component of the state-space vector is denoted by ϕ_i, and $\mathcal{I}_{ij}(\phi)$ is the Fisher information

of the trajectory outcome/origin distribution at state $\boldsymbol{\phi}$,

$$\mathcal{I}_{ij}(\boldsymbol{\phi}) \equiv \sum_{\boldsymbol{s}} p\left(\boldsymbol{s}|\boldsymbol{\phi}\right) \frac{\partial \ln p(\boldsymbol{s}|\boldsymbol{\phi})}{\partial \phi_i} \frac{\partial \ln p(\boldsymbol{s}|\boldsymbol{\phi})}{\partial \phi_j} \tag{3.32a}$$

$$= \sum_{\boldsymbol{s}} \frac{1}{p\left(\boldsymbol{s}|\boldsymbol{\phi}\right)} \frac{\partial p(\boldsymbol{s}|\boldsymbol{\phi})}{\partial \phi_i} \frac{\partial p(\boldsymbol{s}|\boldsymbol{\phi})}{\partial \phi_j} \tag{3.32b}$$

$$= \frac{2}{\left(1 - q^+(\boldsymbol{\phi})\right)q^+(\boldsymbol{\phi})} \frac{\partial\left(1 - q^+(\boldsymbol{\phi})\right)q^+(\boldsymbol{\phi})}{\partial \phi_i} \frac{\partial\left(1 - q^+(\boldsymbol{\phi})\right)q^+(\boldsymbol{\phi})}{\partial \phi_j} \tag{3.32c}$$

$$+ \frac{1}{\left(1 - q^+(\boldsymbol{\phi})\right)^2} \frac{\partial\left(1 - q^+(\boldsymbol{\phi})\right)^2}{\partial \phi_i} \frac{\partial\left(1 - q^+(\boldsymbol{\phi})\right)^2}{\partial \phi_j} + \frac{1}{\left(q^+(\boldsymbol{\phi})\right)^2} \frac{\partial\left(q^+(\boldsymbol{\phi})\right)^2}{\partial \phi_i} \frac{\partial\left(q^+(\boldsymbol{\phi})\right)^2}{\partial \phi_j}$$

$$= \left[\frac{2\left(1 - 2q^+(\boldsymbol{\phi})\right)^2}{\left(1 - q^+(\boldsymbol{\phi})\right)q^+(\boldsymbol{\phi})} + \frac{4\left(1 - q^+(\boldsymbol{\phi})\right)^2}{\left(1 - q^+(\boldsymbol{\phi})\right)^2} + \frac{4\left(q^+(\boldsymbol{\phi})\right)^2}{\left(q^+(\boldsymbol{\phi})\right)^2}\right] \frac{\partial q^+(\boldsymbol{\phi})}{\partial \phi_i} \frac{\partial q^+(\boldsymbol{\phi})}{\partial \phi_j} \tag{3.32d}$$

$$= \frac{2}{q^+(\boldsymbol{\phi})\left(1 - q^+(\boldsymbol{\phi})\right)} \frac{\partial q^+(\boldsymbol{\phi})}{\partial \phi_i} \frac{\partial q^+(\boldsymbol{\phi})}{\partial \phi_j} . \tag{3.32e}$$

In (3.32c), each term from (3.32b) is expressed in terms of the forward committor $q^+(\boldsymbol{\phi})$. The chain rule is used to pull out a common factor in (3.32d), and the expression is simplified in (3.32e).

In information geometry [109, 110], Fisher information arises as a distance metric in the parameter space of a probability distribution. Here, the conditional probability $p(\boldsymbol{s}|\boldsymbol{\phi})$ changes as the system evolves, where the system state $\boldsymbol{\phi}$ parameterizes the conditional probability distribution through the committor $q^+(\boldsymbol{\phi})$. When the system is in A (B), there is no uncertainty about trajectory outcome S_+ and origin S_-, so the conditional probability $p(\boldsymbol{s}|\boldsymbol{\phi} \in A)$ ($p(\boldsymbol{s}|\boldsymbol{\phi} \in B)$) is unity for $\boldsymbol{s} = (A, A)$ ($\boldsymbol{s} = (B, B)$) and zero for all other subensembles. As the system evolves, the conditional distribution over subensembles $p(\boldsymbol{s}|\boldsymbol{\phi})$ corresponding to the current system state changes. The information-geometric distance between successive points in state space where there are changes to coordinates ϕ_i and ϕ_j is quantified by the Fisher information metric, with squared line element [111]

$$\mathrm{d}\ell^2_{\boldsymbol{\phi}\boldsymbol{\phi}'} \equiv \tfrac{1}{2} \sum_{i,j} (\phi_i - \phi'_i) \mathcal{I}_{ij}(\boldsymbol{\phi}')(\phi_j - \phi'_j) . \tag{3.33}$$

Equation (3.31f) is then the rate of mean square distance accumulated by the system evolving at equilibrium,

$$\frac{\langle \mathrm{d}\ell^2 \rangle}{\mathrm{d}t} = \sum_{\boldsymbol{\phi},\boldsymbol{\phi}'} T_{\boldsymbol{\phi}\boldsymbol{\phi}'} \, \pi(\boldsymbol{\phi}') \, \mathrm{d}\ell^2_{\boldsymbol{\phi}\boldsymbol{\phi}'} . \tag{3.34}$$

Multiplying by the mean round-trip time $\tau^{\mathrm{RT}} = \tau^{\mathrm{FP}}_{A \to B} + \tau^{\mathrm{FP}}_{B \to A} = (\nu_s)^{-1}$ to complete a cycle through the subensembles (the sum of mean first-passage times from A to B and B to A [91]), the squared reaction-coordinate length $\mathcal{L}^2_{AB}$ is obtained as the mean-squared metric distance for a round trip $A \to B \to A$:

$$\mathcal{L}^2_{AB} \equiv (\tau_A + \tau_B)\frac{\langle \mathrm{d}\ell^2 \rangle}{\mathrm{d}t} \tag{3.35a}$$

$$\approx 2\tau^{\mathrm{RT}} p_{\mathrm{R}} \dot{\Sigma}_{\mathrm{R}} \tag{3.35b}$$

$$= 2\tau_{\mathrm{R}} \dot{\Sigma}_{\mathrm{R}} \tag{3.35c}$$

$$\dot{\Sigma}_{\mathrm{R}} \approx \frac{\mathcal{L}^2_{AB}}{2\tau_{\mathrm{R}}} \,, \tag{3.35d}$$

for mean transition-path duration $\tau_{\mathrm{R}} = \tau^{\mathrm{RT}} p_{\mathrm{R}}$, given by the mean time to complete a cycle through all subensembles multiplied by the marginal probability of a reactive subensemble [90, 91]. The reaction-coordinate length $\mathcal{L}_{AB}$ roughly quantifies the mean number of fluctuations [112, 67] required for the system to complete a round trip. Therefore, TPE entropy production provides a meaningful measure of the distance between A and B [113], quantifying the time-irreversibility in the transition.

3.3 Quantitative measure of coordinate's relevance to reaction

The TPE entropy production quantitatively measures the relevance of an arbitrary coordinate to the reaction, which is demonstrated below for a simple model system. The system is assumed to have bipartite dynamics [86, 87], essentially that instantaneous transitions only happen in either a one-dimensional coordinate X or in all other degrees of freedom $\boldsymbol{Y}$ making up the system state $\boldsymbol{\Phi} = (X, \boldsymbol{Y})$:

$$T_{xx',\boldsymbol{yy'}} = \begin{cases} T_{xx',\boldsymbol{y}} & x \neq x' \,, \boldsymbol{y} = \boldsymbol{y'} \\ T_{x,\boldsymbol{yy'}} & x = x' \,, \boldsymbol{y} \neq \boldsymbol{y'} \\ -\sum\limits_{\substack{x'' \neq x' \\ \boldsymbol{y''} \neq \boldsymbol{y'}}} T_{x''x',\boldsymbol{y''y'}} & x = x' \,, \boldsymbol{y} = \boldsymbol{y'} \\ 0 & \text{otherwise} \end{cases} \tag{3.36}$$

Dynamics that do not obey the bipartite assumption introduce further complications in unambiguously partitioning the entropy production between coordinates [114].

Combining (3.2), (3.22a), and (3.23) gives the full TPE entropy production as a function of the forward committor,

$$p_{\mathrm{R}} \dot{\Sigma}_{\mathrm{R}} = \tfrac{1}{2} \sum_{\phi, \phi'} T_{\phi \phi'} \pi(\phi')(q^+(\phi) - q^+(\phi')) \ln \frac{q^+(\phi)(1 - q^+(\phi'))}{q^+(\phi')(1 - q^+(\phi))} \,, \qquad (3.37)$$

which splits into contributions from the two transition types:

$$p_{\mathrm{R}} \dot{\Sigma}_{\mathrm{R}} = p_{\mathrm{R}} \dot{\Sigma}_{\mathrm{R}}^X + p_{\mathrm{R}} \dot{\Sigma}_{\mathrm{R}}^{\boldsymbol{Y}} \qquad (3.38a)$$

$$= \tfrac{1}{2} \sum_{x, x', \boldsymbol{y}} T_{xx', \boldsymbol{y}} \pi(x', \boldsymbol{y})[q^+(x, \boldsymbol{y}) - q^+(x', \boldsymbol{y})] \ln \frac{q^+(x', \boldsymbol{y})[1 - q^+(x, \boldsymbol{y})]}{q^+(x, \boldsymbol{y})[1 - q^+(x', \boldsymbol{y})]} \qquad (3.38b)$$

$$+ \tfrac{1}{2} \sum_{x, \boldsymbol{y}, \boldsymbol{y}'} T_{x, \boldsymbol{y} \boldsymbol{y}'} \pi(x, \boldsymbol{y}')[q^+(x, \boldsymbol{y}) - q^+(x, \boldsymbol{y}')] \ln \frac{q^+(x, \boldsymbol{y})[1 - q^+(x, \boldsymbol{y}')]}{q^+(x, \boldsymbol{y}')[1 - q^+(x, \boldsymbol{y})]} \,.$$

The same decomposition holds for the information rate in Eq. (3.27a) [107], so that TPE entropy production due to X dynamics is equal to the information rate (due to X dynamics) between $\boldsymbol{\Phi}$ and S_+:

$$p_{\mathrm{R}} \dot{\Sigma}_{\mathrm{R}}^X = \dot{I}^X(S_+; \boldsymbol{\Phi}) \,, \qquad (3.39)$$

where

$$\dot{I}^X(S_+; \boldsymbol{\Phi}) \equiv \sum_{x, x', \boldsymbol{y}, s_+} T_{xx', \boldsymbol{y}}^{s_+} p(x', \boldsymbol{y}, s_+) \ln \frac{p(s_+ | x, \boldsymbol{y})}{p(s_+ | x', \boldsymbol{y})} \,. \qquad (3.40)$$

This is the third major result: the entropy production due to dynamics of coordinate X equals the mutual information generated by X dynamics, thereby quantifying the relevance of X transitions to identifying the current subensemble and highlighting those transitions that are 'correlated' with reactive trajectories and therefore important to the reaction mechanism.

In particular, for X^* determining the committor and $\boldsymbol{Y}^*$ orthogonal degrees of freedom that are therefore not relevant to the reaction $[q(x, \boldsymbol{y}) = q(x)]$, the entropy production rate due to $\boldsymbol{Y}^*$ dynamics is (simplifying (3.38b)):

$$\dot{\Sigma}_{\mathrm{R}}^{\boldsymbol{Y}^*} = \sum_{x, \boldsymbol{y}, \boldsymbol{y}'} T_{x, \boldsymbol{y} \boldsymbol{y}'} \pi(x, \boldsymbol{y}')[q^+(x) - q^+(x)] \ln \frac{q^+(x)[1 - q^+(x)]}{q^+(x)[1 - q^+(x)]} = 0 \,. \qquad (3.41a)$$

Therefore $\dot{\Sigma}_{\mathrm{R}}^{X^*} = \dot{\Sigma}_{\mathrm{R}}$. This is additional confirmation that the committor is the reaction coordinate, in that it provides a thermodynamically complete coarse-grained representation of the transition-path ensemble, fully accounting for its entropy production [108].

3.3.1 Illustrative example: 2D bistable potential

The ability of TPE entropy production to identify relevant reaction coordinates is illustrated with overdamped dynamics in a double-well energy landscape (Fig. 3.2a). The bistable energy potential is separable into terms only depending on the reaction coordinate r and on the bath mode b:

$$E_{\text{int}}(r, b) = -k_{\text{B}}T \ln \left[e^{-\frac{1}{2}\beta k_{\text{m}}(r+r_{\text{m}})^2} + e^{-\frac{1}{2}\beta k_{\text{m}}(r-r_{\text{m}})^2} \right] + \tfrac{1}{2}kb^2 \,, \tag{3.42}$$

where $r_{\text{m}} = 1$ defines the locations of the energy minima, and k_{m} is the landscape curvature near those energy minima, chosen such that the energy barrier $E_{\text{int}}(0, 0) - E_{\text{int}}(r_{\text{m}}, 0)$ is $4k_{\text{B}}T$.

To exemplify the typical situation where the reaction coordinate is not known *a priori* and coordinates are thus chosen based on convenience or intuition, fixed system coordinates (x, y) lie at an angle θ to the correct reaction coordinate, the linear coordinate passing through both energy minima:

$$r = x \cos \theta + y \sin \theta \tag{3.43a}$$

$$b = -x \cos \theta + y \sin \theta \,. \tag{3.43b}$$

The state space (x, y) is divided into discrete segments $\text{d}x = \text{d}y = 0.04r_{\text{m}}$, and the Master equation with transition rates $T_{xx',yy'} = \Gamma * \min \left[1, e^{-\beta \Delta E_{\text{int}}}\right]$ with diffusion prefactor $\Gamma = 0.1 \, \text{d}t^{-1}$ and energy change $\Delta E_{\text{int}} = E_{\text{int}}(x, y) - E_{\text{int}}(x', y')$ [115] is used to evolve the system's dynamics. The committor is solved on the discrete state space using the recursion relations [30]

$$\sum_{\phi'} T_{\phi'\phi} q^{+}(\phi') = 0 \tag{3.44a}$$

$$\sum_{\phi'} T_{\phi\phi'} \frac{\pi(\phi')}{\pi(\phi)} q^{-}(\phi') = 0 \,. \tag{3.44b}$$

Reaction endpoints are defined by $A = \{x, y \mid r(x, y) \leq -r_m\}$ and $B = \{x, y \mid r(x, y) \geq r_m\}$, so that the committor is independent of b. The TPE entropy production is calculated from (3.38).

For $\theta = 0°$, X is the reaction coordinate, Y is an orthogonal *bath mode* [116], and X dynamics fully capture the TPE entropy production without Y contribution. Figure 3.2b shows that as the underlying energy landscape is rotated relative to system coordinates, the X-coordinate entropy production decreases and Y-coordinate entropy production increases, with equal contribution at $\theta = 45°$. The entropy production for each coordinate is proportional to the squared Euclidean distance between A and B projected onto each coordinate, $\dot{\Sigma}_{\text{R}}^{X}(\theta) \propto \cos^2 \theta$ and $\dot{\Sigma}_{\text{R}}^{Y}(\theta) \propto \sin^2 \theta$. This proportionality is an interesting result in light
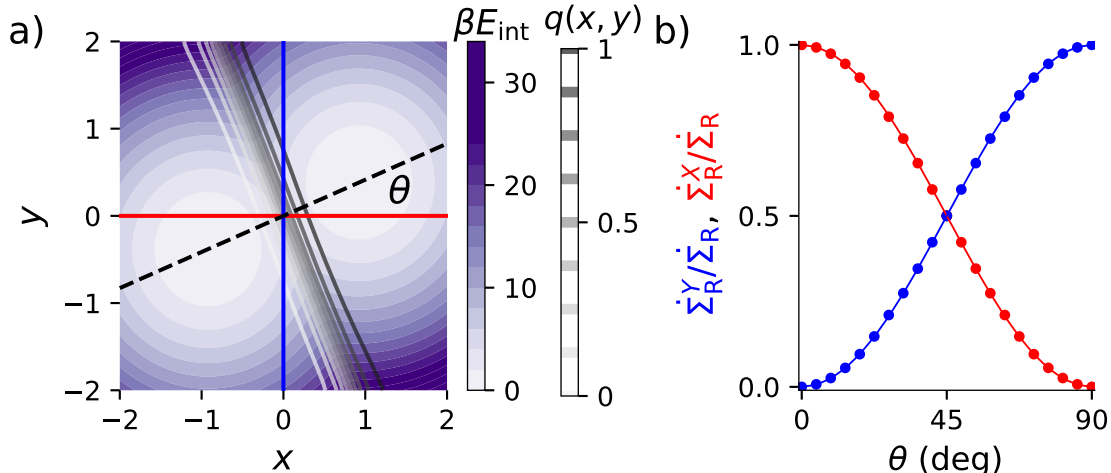
Figure 3.2: TPE entropy production rate for 2D double-well potential with rotating coordinate system. a) Double-well energy landscape (purple) with fixed x-axis (red solid) and y-axis (blue solid), and rotated reaction coordinate (black dashed). Gray contours: forward committor. b) Share of TPE entropy production due to coordinate X (red dots) and Y (blue dots), as a function of reaction-coordinate angle θ. Red solid: $\cos^2\theta$. Blue solid: $\sin^2\theta$.

of the information metric derived in Section 3.2.3, and could warrant further investigation into geometric aspects of reaction dynamics and information.

3.4 Discussion

The information thermodynamics of a system undergoing reactions between distinct state-space subsets A and B has been derived, making a fundamental connection between transition-path theory, information theory, and stochastic thermodynamics. Partitioning a long ergodic equilibrium trajectory into reactive and non-reactive subensembles results in entropy production for system dynamics in the reactive subensembles (physically representing the dissipation needed to implement the detailed-balance-breaking transition rates of the TPE), which in turn identifies transitions that are relevant to the overall reaction mechanism. This rigorous equality between TPE entropy production and informativeness of dynamics also holds for an arbitrary coordinate, revealing parallel stochastic-thermodynamic and information-theoretic measures of the relevance of collective variables to the system reaction, that are each maximized by the committor.

This work has implications for the identification of important collective variables and analysis of reaction mechanisms. While the committor provides a microscopically detailed reaction coordinate that maps each system microstate to a scalar value, it does not immediately identify physically meaningful collective variables (e.g., internal molecular coordinates) that are relevant to the reaction [46, 23, 75]. These results indicate that relevant coordinates

are identified as having nonzero entropy production in the transition-path ensemble; thus partitioning the entropy production between multiple relevant collective variables for that provide physical intuition can yield a low-dimensional model that allows increased insight into the reaction mechanism.

More concretely, this connection that has been established between transition-path theory and stochastic thermodynamics suggests a novel method for rigorously grounded inference of reaction coordinates: generate an ensemble of transition paths using transition-path sampling [26, 27] or related algorithms [31, 32, 28]; estimate entropy production along chosen coordinates [89, 117, 118] or identify linear combinations of coordinates producing the most entropy using dissipative-components analysis [119]; use these most dissipative coordinates to enhance sampling of transition paths; and through further iteration identify system coordinates producing the most entropy in the transition-path ensemble and hence of most relevance to the reaction.

Machine-learning approaches to solve for high-dimensional committor coordinates [43, 44, 45] or find low-dimensional reaction models that retain predictive power [120, 49] are active areas of research [121]. The information-theoretic and thermodynamic perspectives on reactive trajectories described in this chapter provide guidance to the development of data-intensive automated methods to infer these models and their corresponding reaction mechanisms.

Chapter 4

Multidimensional Minimum-Work Protocols

4.1 Introduction

In the previous chapter, a thermodynamic perspective on reaction dynamics that allows quantification of reaction information was discussed. This provides a concrete theoretical framework for determining small sets of collective variables that are relevant to a reaction. As discussed in Section 2.3.1, there is interest in understanding the correspondence between relevant collective variables and efficient driving protocols. However, there has been little work in the literature to date discussing minimum-work protocols that drive multiple spatial coordinates of a system between metastable states. In this chapter, protocols in a multidimensional control-parameter space are examined to establish some intuition about principles of minimum-work driving.

Many systems of interest have slow relaxations between metastable mesostates, and finding good control parameters to drive the system can be non-trivial [11]. For finite-duration protocols where the system has insufficient time to fully relax, the system can get stuck in metastable mesostates, causing the nonequilibrium system distribution to lag the equilibrium distribution, resulting in excess work. A longer-duration protocol would allow further time for the system to relax from these metastable mesostates towards the equilibrium distribution and reduce excess work, but it may not be feasible or desirable to increase duration.

A similar problem arises in enhanced-sampling contexts, where external control parameters are used to bias system degrees of freedom that are relevant to characterizing metastable mesostates, reducing the time required to sample the system space and compute free energies [96, 122]. A collective variable is sought that captures the slowest relaxation mode, otherwise the system can get stuck in a metastable region, resulting in slow convergence of relevant statistics (e.g., free-energy profiles) [11, 56]. Using multiple biasing coordinates that

couple to relevant relaxation modes can significantly improve sampling speed by allowing the system to more easily circumvent dynamical barriers [123, 124, 125, 126, 127].

Inspired by these results, multidimensional minimum-work protocols are investigated in this chapter in more detail, using external magnetic fields to drive a 3×3 Ising system between energetically stable configurations. Reducing the size of the system and control-parameter space greatly simplifies the calculations, since analytical calculations of the Boltzmann distribution and static equilibrium properties throughout control-parameter space are possible. Of particular interest is the development of a physically intuitive picture of the geometric aspects of control protocols and the generalized friction matrix. The Ising model is introduced in Section 4.2, then the calculations of minimum-work protocols are outlined in Section 4.3. Section 4.4 presents the results of the calculations and further analysis on the differences between naive and designed protocols. Section 4.5 concludes with a broader discussion of the results and their overall relevance to this thesis.

4.2 Ising system

The Ising model is a workhorse model used throughout statistical mechanics to study a variety of systems and phenomena, such as phase transitions, nucleation, and gene-regulatory networks [128]. Its basic components are a set of up or down spins (or more generally, a Bernoulli random variable) with pairwise coupling energy J. If $J > 0$, spin pairs have lower energy when they are aligned, and the Ising model is called ferromagnetic. Additionally, a magnetic field $\boldsymbol{\lambda}$ can be applied to spins to affect the relative energies of the up and down orientation of each spin.

Calculations of minimum-work protocols to drive a large 2D Ising model between the all-down and all-up configurations using multiple control parameters have been done in Refs. [102, 103, 104]. In particular, Ref. [103] drove the Ising model using a set of magnetic fields applied to blocks of spins in the 2D array, which pushes on collective variables given by the magnetization of spins under the control of a field. The minimum-work protocols designed for this control-parameter space resemble the minimum-free energy path (Sec. 2.2) calculated for this system [94].

Our system is a 2-dimensional 3×3 Ising model with fixed anti-symmetric boundary conditions [103], illustrated in Fig. 4.1. The spins are ferromagnetically coupled, with spin configuration $\{\boldsymbol{\sigma}\}$ having internal energy

$$E_{\text{int}}(\{\boldsymbol{\sigma}\}) \equiv -J \sum_{\{i,j\}} \sigma_i \sigma_j \,, \tag{4.1}$$

where $J = 1\,k_{\text{B}}T$ is the coupling coefficient for Boltzmann constant k_{B} and temperature T, $\sigma_i \in \{-1, 1\}$ is the orientation of spin i, and $\sum_{\{i,j\}}$ denotes a sum over nearest-neighbor spin pairs. The system evolves under single-spin-flip Glauber dynamics [82], where at the end of
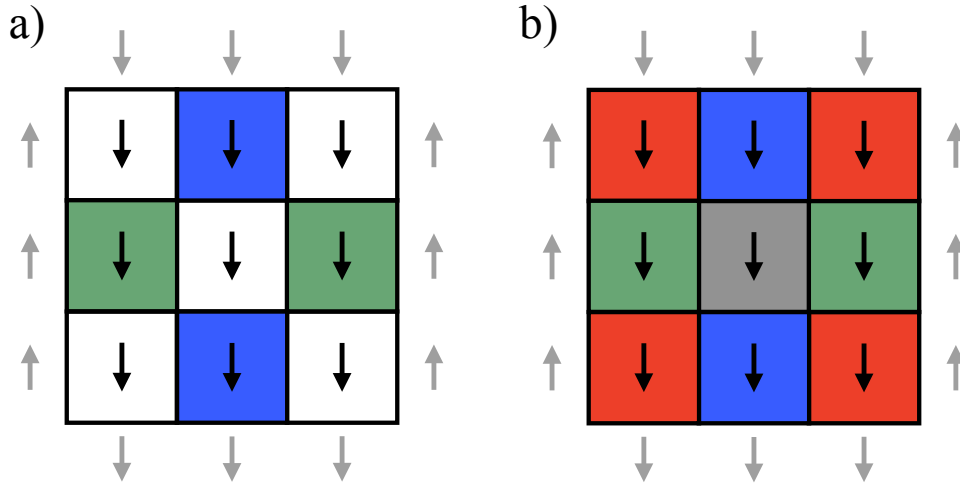
Figure 4.1: Schematic of 3×3 Ising model with 9 fluctuating spins (black, shown in initial all-down configuration) and 12 fixed boundary spins (gray). a) 2D control parameter given by two external fields on spins colored blue and green, respectively. b) 4D control parameter given by four external fields on spins colored according to their symmetry type. Spins with the same color are influenced by a common magnetic field.

a fixed time interval a spin is chosen at random and a spin-flip is attempted, with acceptance probability given by the Glauber criterion (Eq. 2.6). The system has two energetically metastable configurations, with spins all down or all up, and the dynamical barrier separating the metastable configurations has activation energy $\Delta E^\ddagger = 8k_\text{B}T$; the spin-inversion transition from all down to all up is a relatively rare event, with mean first-passage time of $\sim 2 \times 10^3$ attempted spin flips [90].

The spin-inversion reaction to flip the system configuration from all-down to all-up is driven by a set of external magnetic fields $\boldsymbol{h}$ applied to sets of spins (background colors in Fig. 4.1), chosen to reflect the symmetry of the boundary conditions. The applied magnetic fields $\boldsymbol{\lambda}$ bias the total energy as

$$V(\{\boldsymbol{\sigma}\}, \boldsymbol{\lambda}) \equiv E_\text{int}(\{\boldsymbol{\sigma}\}) - \boldsymbol{\lambda}^\text{T} \cdot \boldsymbol{X}(\{\boldsymbol{\sigma}\}) \,, \tag{4.2}$$

where $\boldsymbol{X}(\{\boldsymbol{\sigma}\})$ is a vector of conjugate forces for configuration $\{\boldsymbol{\sigma}\}$, with force $X_i \equiv -\partial V/\partial \lambda_i$ conjugate to control parameter (external field) λ_i, equaling the total magnetization of the spin set controlled by λ_i. The system magnetization is inverted by changing the magnetic field from $\lambda_{i,\text{initial}} = -0.5\,k_\text{B}T$ for all i (favoring the all-down configuration) to $\lambda_{i,\text{final}} = 0.5\,k_\text{B}T$ (favoring the all-up configuration).

Two multidimensional control-parameter spaces are explored, associated with a 2D magnetic field applied to the blue and green spins (Fig. 4.1a, the spins that are most biased by

the boundary conditions) or a 4D magnetic field applied to all spins (Fig. 4.1b, the highest dimensionality that does not break the symmetry imposed by the boundary conditions). The 2D control-parameter space is a two-dimensional manifold of the 4D control-parameter space with $\lambda_{\text{red}} = \lambda_{\text{black}} = 0\,k_{\text{B}}T$.

4.3 Methods

In both 2D and 4D control-parameter spaces three types of protocols between the same endpoints are considered. A naive protocol changes all magnetic fields with constant velocity. The time-optimized protocol changes all magnetic fields together (i.e., $\lambda_i(t) = \lambda_j(t)$ for all i, j, t), with velocity optimized to minimize (within the linear-response approximation) the work (see Section 2.3.5). Both these protocol types are equivalent to one-dimensional control using a single magnetic field to drive the system, with scalar conjugate force equal to the total magnetization of controlled spins (only blue and green spins' magnetization for 2D, and total magnetization for 4D). This yields a one-dimensional friction coefficient, the sum of all elements of the 2D or 4D friction matrices $\zeta^{\text{1D}}(\boldsymbol{\lambda}) = \sum_{ij} \zeta_{ij}(\boldsymbol{\lambda})$. During the time-optimized protocols, all fields are changed with velocity proportional to the inverse square-root of the one-dimensional friction coefficient [7], $\dot{\lambda}^{\text{1D}} \propto \zeta^{\text{1D}}(\boldsymbol{\lambda})^{-1/2}$. The fully optimized protocols solve the Euler-Lagrange equation [78] for the excess work (2.64) in all control-parameter dimensions, which is done numerically using the string method [103] (see Section 2.3.5).

The 4×4 friction matrix (2.62) is estimated on a discrete grid in 4D control-parameter space, then used to design fully optimized 2D and 4D protocols. Sampling the friction in a multidimensional space has a significant computational cost since the number of grid points scales exponentially with the number of control parameters; in contrast, the time-optimized protocol can be calculated by estimating the friction matrix along the one-dimensional protocol path, which has a significantly smaller computational cost. (The naive protocol requires no prior sampling.) Each protocol type is simulated for a range of protocol durations to collect an ensemble of system responses. Appendix B provides details on numerical methods.

4.3.1 Reference relaxation time

Throughout, work and power are reported divided by the number N_i of spins controlled by field i ($N_{\text{red}} = 4$, $N_{\text{blue}} = N_{\text{green}} = 2$, and $N_{\text{black}} = 1$). The protocol duration is scaled by a reference relaxation time τ_{rel} discussed below.

Since the system is driven from all-down to all-up, a reference relaxation time is chosen to reflect the timescale of this transition. The rate constant $k(\boldsymbol{\lambda})$ for the reaction from all-down to all-up at equilibrium for control-parameter $\boldsymbol{\lambda}$ is calculated from transition-path theory [90] (Section 2.2). This gives the mean number of all-down to all-up transitions observed per unit time, given that the system is initially in the all-down configuration.

This rate constant is averaged over control-parameter values along the 4D naive protocol to obtain an average rate constant $\bar{k}$ for the spin inversion during this protocol:

$$\bar{k} = \frac{1}{\Delta t} \int_0^{\Delta t} \mathrm{d}t \; k(\boldsymbol{\lambda}(t)) \, . \tag{4.3}$$

$\bar{k}$ represents the mean number of transitions from all-down to all-up during a naive 4D protocol of duration Δt, assuming the system is in local equilibrium in the initial metastable basin throughout the protocol. $\tau_{\mathrm{rel}} = \bar{k}^{-1}$ then is the protocol time required for the system to make on average one spin-inversion transition. For this system, $\tau_{\mathrm{rel}} = 1102 \, \mathrm{d}t$, where during each $\mathrm{d}t$ one spin flip is attempted. The moderate-duration protocols shown in Figs. 4.5, 4.8, 4.9, and 4.11 have duration $\Delta t = 1000 \, \mathrm{d}t \approx 0.91 \, \tau_{\mathrm{rel}}$.

4.4 Results

4.4.1 Avoiding high friction

To illustrate general features of the multidimensional friction matrix for this system, Fig. 4.2a shows the trace (a scalar capturing the essential features) of the 2×2 friction matrix in the control-parameter space defined by axes $(\lambda_{\mathrm{blue}}, \lambda_{\mathrm{green}})$. The friction is maximized at $(1, -1)$, where the fields cancel the fixed boundary conditions (Fig. 4.1), stabilizing the all-down and all-up configurations relative to the unperturbed system ($\lambda_i = 0$ for all i). Both components of the friction (2.67), the force covariance (Fig. 4.2b) and the integral relaxation time (Fig. 4.2c), are similarly peaked.

Figure 4.2 also shows the 2D time-optimized and fully optimized protocols. (The naive protocol follows the same path as the time-optimized protocol but with constant protocol velocity.) The naive and time-optimized protocols follow a straight line between the fixed endpoints and drive the system through a high-friction region, with the time-optimized protocol slowing down where the friction along the path is large, near control parameters corresponding to the unperturbed system. In contrast, the fully optimized protocol avoids this by making an excursion through low-friction regions of control-parameter space.

Figure 4.3 shows the three protocol types in 2D and 4D control-parameter spaces. The time-optimized protocols have high initial and final velocity, and low velocity in the middle of the protocol, where fields are close to zero and friction is large. This is consistent with the minimum-work protocol for one-dimensional barrier crossing [95], where the protocol slows to allow time for stochastic fluctuations to kick the system over an energy barrier, thereby remaining closer to equilibrium relative to the naive protocol.

The fully (both temporally and spatially) optimized protocols depart dramatically from the naive and time-optimized protocols. Consistent with results in a larger Ising model [103], the green field (whose corresponding spins are initially anti-aligned with their boundary conditions and therefore energetically frustrated) is increased rapidly in the early stages of
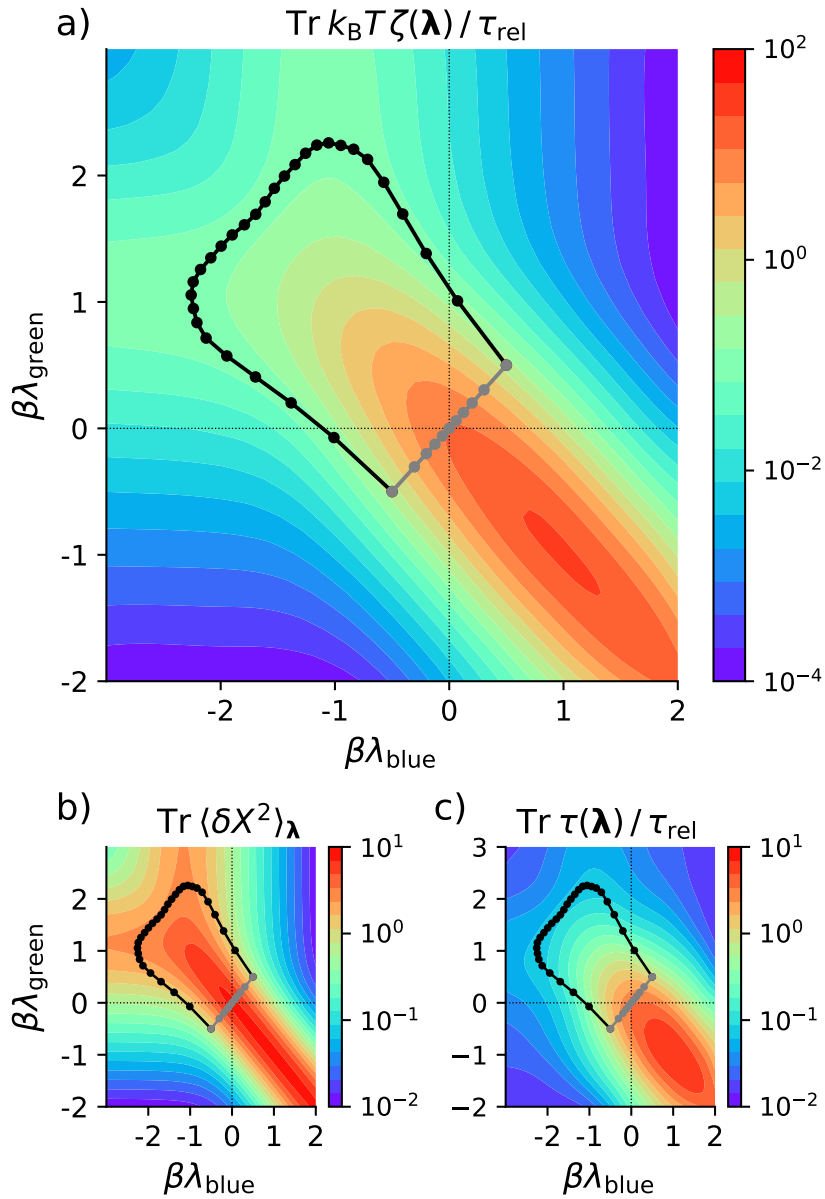
Figure 4.2: 2D fully optimized (black) and time-optimized (gray) protocols overlaid on a heatmap showing the trace of the a) friction (2.62), b) force covariance, or c) integral relaxation time (2.68) matrix. Points denote control-parameter values evenly spaced in time on the given protocol.

the protocol, while the blue field (whose corresponding spins are initially aligned with their boundary conditions) is decreased initially, then increased later. 2D and 4D protocols show
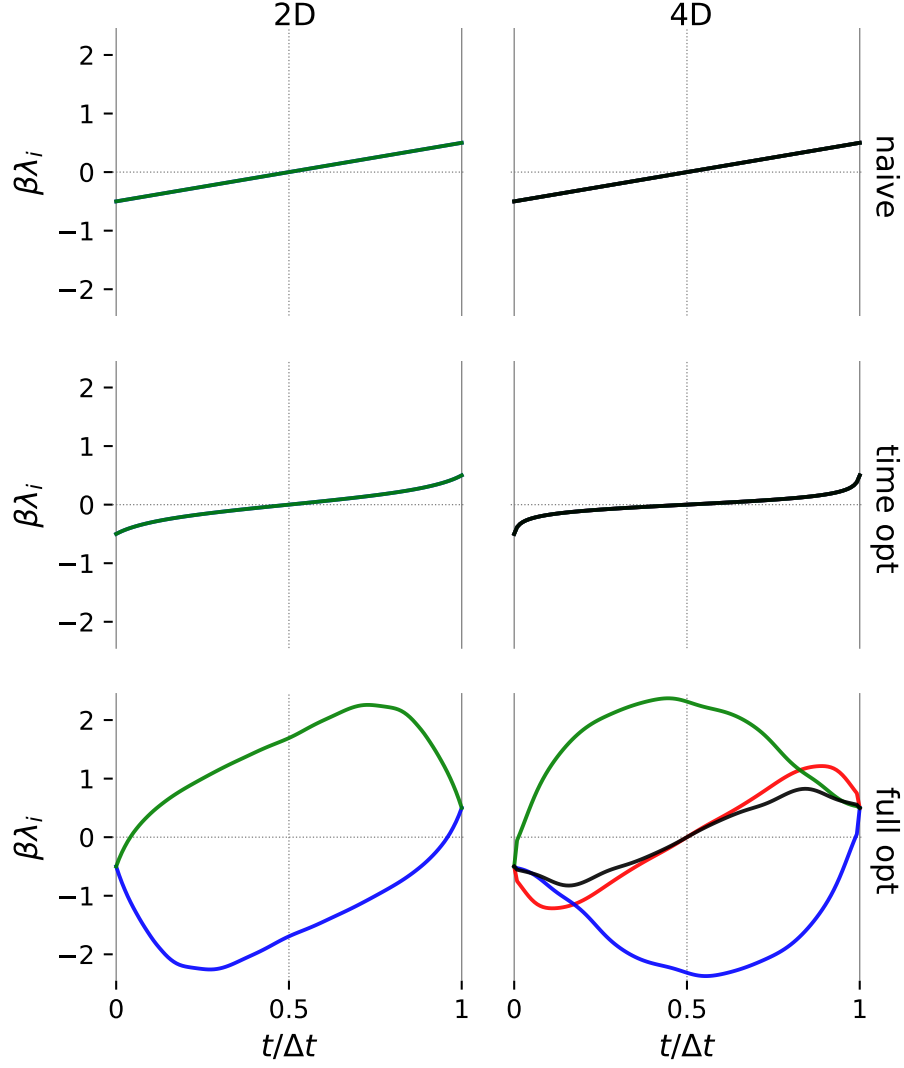
Figure 4.3: Designed protocols for driving spin-inversion in the 3×3 Ising model. Magnetic fields λ_i as a function of scaled protocol time $t/\Delta t$ during naive (top row), time-optimized (middle), and fully optimized (bottom) protocols in 2D (left column) and 4D (right) control-parameter spaces. Colors correspond to spin sets in Fig. 4.1.

the same trend, but the addition of time-varying red and black fields in 4D protocols shifts the relative timing of changes in green and blue fields.

A striking feature of both the 2D and 4D fully optimized protocols is the non-monotonic magnetic fields at early and late stages. In the linear-response regime, a 1D minimum-work protocol must be monotonic since a non-monotonic protocol drives through the same control-parameter value more than once, producing unnecessary dissipation; in the metric language of the linear-response approximation, the shortest curve between two points cannot
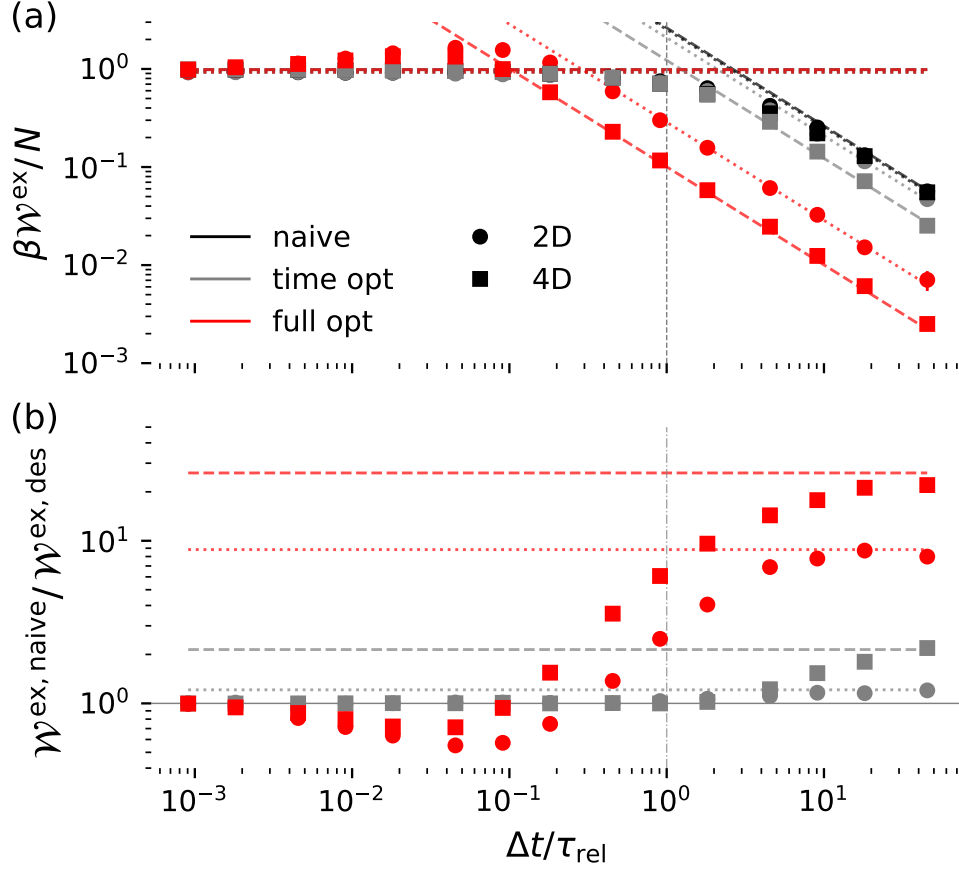
Figure 4.4: Comparison of excess work in naive and designed protocols. (a) Excess work (2.64) scaled by number N of controlled spins (i.e., $N = 4$ controlled spins for 2D protocols and $N = 9$ for 4D) as a function of protocol duration Δt, for each of the six protocol types (Fig. 4.3): 2D (circles) and 4D (squares) control-parameter spaces; and naive (black), time-optimized (gray), and fully optimized (red). (b) Ratio of excess work during naive protocols to excess work during time-optimized (gray) and fully optimized (red) protocols for 2D (circles) and 4D (squares) control. Horizontal dotted lines: excess work for an instantaneous protocol. Dashed lines: linear-response approximations, accurate at long duration. Protocol durations are scaled by reference relaxation time τ_{rel}.

cross itself. But individual fields for minimum-work protocols can be non-monotonic as long as the protocol path in multidimensional space does not loop back on itself.

4.4.2 Reducing excess work and keeping close to equilibrium

Next, the excess work for all protocols is compared to assess their relative performance. Figure 4.4a shows the excess work as a function of protocol duration. In the limit of long duration, the excess works approach their respective linear-response approximations and scale as Δt^{-1}. For short protocols, the excess works approach that of an instantaneous pro-
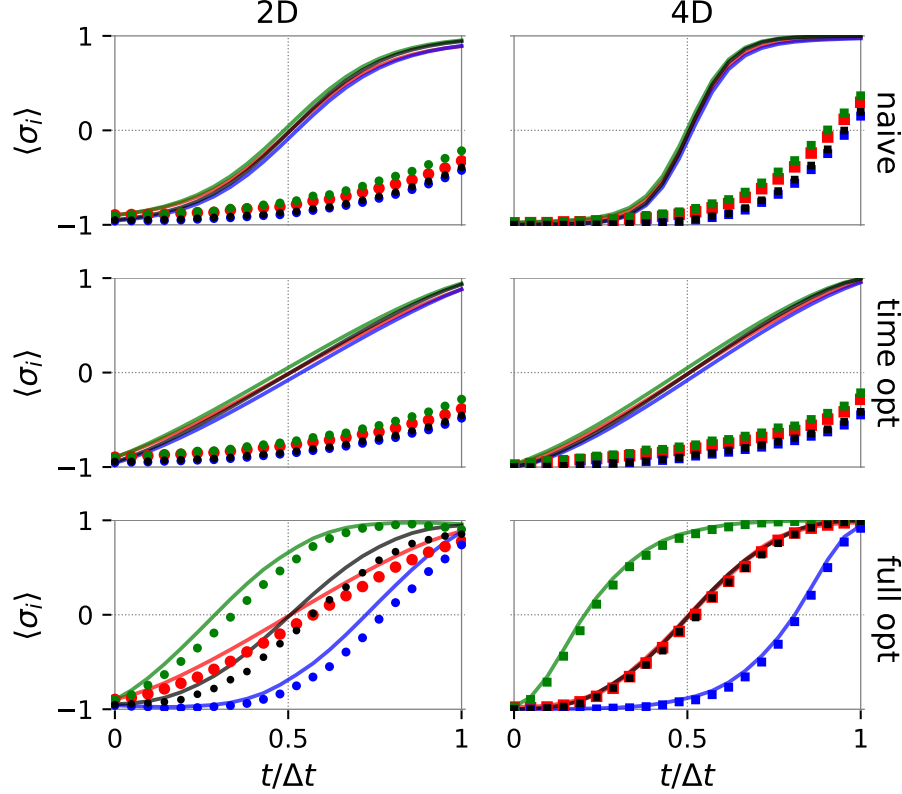
Figure 4.5: Quasistatic (curves) and nonequilibrium (points) mean magnetization of each spin type (different colors, see Fig. 4.1) as a function of elapsed time t along a moderate-duration ($\Delta t \approx \tau_{\text{ref}}$) protocol of each type (see Fig. 4.3).

tocol, given by the relative entropy of the initial and final equilibrium distributions [69]. Figure 4.4b shows the ratio of excess work for naive relative to time-optimized and fully optimized protocols. While the 2D and 4D time-optimized protocols respectively do $\sim 1.2\times$ and $\sim 2\times$ less excess work than the corresponding naive protocols, the fully optimized protocols respectively do $\sim 9\times$ and $\sim 26\times$ less excess work than the naive protocols. (Fully optimized control protocols do more work than all other protocols (including an instantaneous protocol) for short durations ($\Delta t / \tau_{\text{rel}} \lesssim 0.1$).)

Fully optimized protocols keep the system much closer to equilibrium than naive and time-optimized protocols for a fixed duration, resulting in significant work reduction. Figure 4.5 shows for moderate protocol duration ($\Delta t \approx \tau_{\text{rel}}$) the quasistatic and nonequilibrium mean magnetizations $\langle \sigma_i \rangle$ of each spin type. In fully optimized protocols, there is a rough ordering of spin flips matching the corresponding field increases (Fig. 4.3): green spins flip first, red and black spins flip next, and blue spins flip last (also depicted schematically in Fig. 4.6). This temporal separation of spin flips contrasts with naive and time-optimized protocols, where the different spin types have nearly equal mean magnetization throughout the protocol.
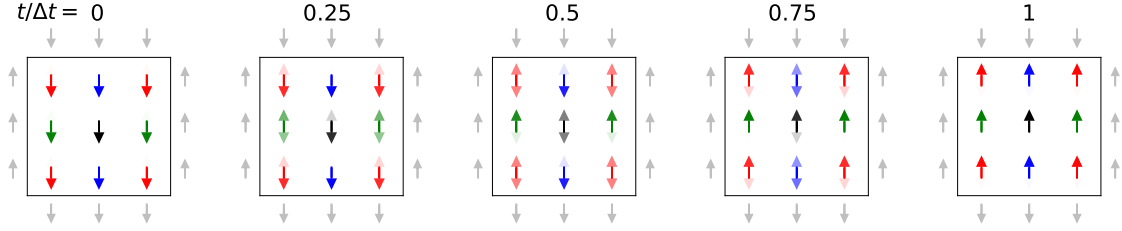
Figure 4.6: Spin-inversion mechanism at five evenly spaced times during the 4D fully optimized protocol. The relative equilibrium population of up and down orientations are indicated by the transparency of the arrowheads.

Throughout the naive and time-optimized protocols, the system remains bistable, with the external fields primarily biasing the relative energies of the all-down and all-up configurations. For the moderate duration shown, the system gets stuck in the initial metastable basin so nonequilibrium mean magnetizations significantly lag the quasistatic limit. In contrast, the fully optimized protocols keep the system much closer to equilibrium, with modest lag during the 2D protocol and nearly no lag during the 4D protocol.

4.4.3   Reducing spin covariance and flattening the energy landscape

Since the fully optimized protocols drive the system through configuration space differently than the naive and time-optimized protocols, it is helpful to understand general features that are associated with reduced protocol work in the fully optimized protocols.

Figure 4.7 shows the equilibrium variance of each spin and covariance between neighboring and non-neighboring spins. The 9×9 spin-covariance matrix is "coarse-grained" (by summing spin-covariance elements for the same spin types into a single entry) to form the 4×4 force-covariance matrix, a factor in the friction (2.67) that contributes to excess power for 4D protocols. Since each spin magnetization is a Bernoulli random variable, it is a quadratic function of the mean, $\langle \delta\sigma_i^2 \rangle = 1 - \langle\sigma_i\rangle^2$, where each variance peaks when its mean changes sign (Fig. 4.5). At some point during a long-duration protocol, the mean magnetization of each spin must change sign and hence the corresponding diagonal element of the force-covariance matrix must reach a fixed maximal value; however, the contribution from off-diagonal elements (covariance between distinct spins) can be reduced. In the naive and time-optimized protocols, where the system primarily occupies the all-up and all-down configurations, all spin covariances (even for non-neighboring spin pairs) are large where those configurations are equally populated. The fully optimized protocols disrupt this bistability and lower the total energy of many other configurations; this reduces spin covariance, with low covariance between neighboring spins and near-zero covariance between non-neighboring spins. The fully optimized 4D protocol, with additional controls, reduces covariance more than the 2D protocol.
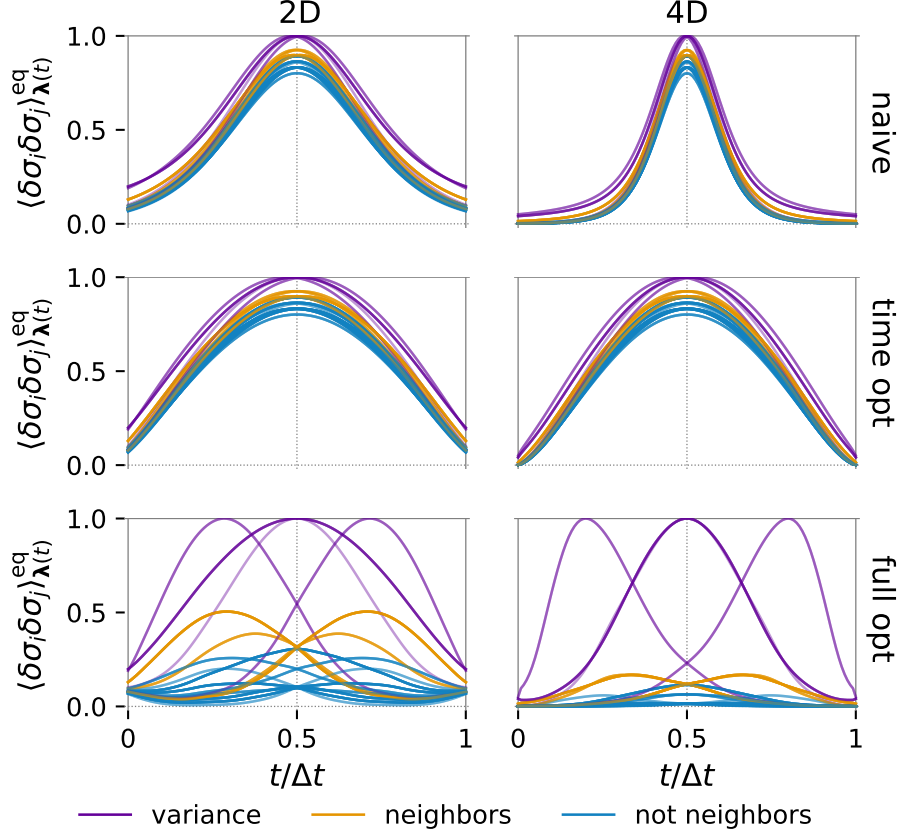
64

Figure 4.7: Equilibrium spin covariance as a function of elapsed time t along a protocol of each type (Fig. 4.3). Purple: spin variance. Gold: covariance between neighboring spins. Blue: covariance between non-neighboring spins.

The covariance between spins reflects the internal energy that couples their orientations; high covariance stems from ensembles dominated by low-internal-energy configurations with aligned spins, while low covariance results from the system accessing configurations with anti-aligned spins and hence higher internal energy. Next, the energetic and entropic properties of instantaneous system distributions along the control protocol are calculated.

Figure 4.8 shows the free energy along each protocol, as well as its component energies (4.2) and entropy $H \equiv -\sum_{\{\sigma\}} p(\{\sigma\}) \ln p(\{\sigma\})$ for (equilibrium or nonequilibrium) probability distribution $p(\{\sigma\})$. The fully optimized protocols drive the system through distributions with high mean internal energy (4.1). This indicates significant population of high-internal-energy configurations (i.e., with anti-aligned spins), relative to naive and time-optimized protocols where the internal energy is relatively constant (consistent with the system primarily occupying the all-down or all-up configuration). These configurations with high internal energy are stabilized by stronger external fields, flattening the total-energy landscape relative to naive and time-optimized protocols. The fully optimized protocols also significantly increase the entropy, indicating many accessible system configurations, in
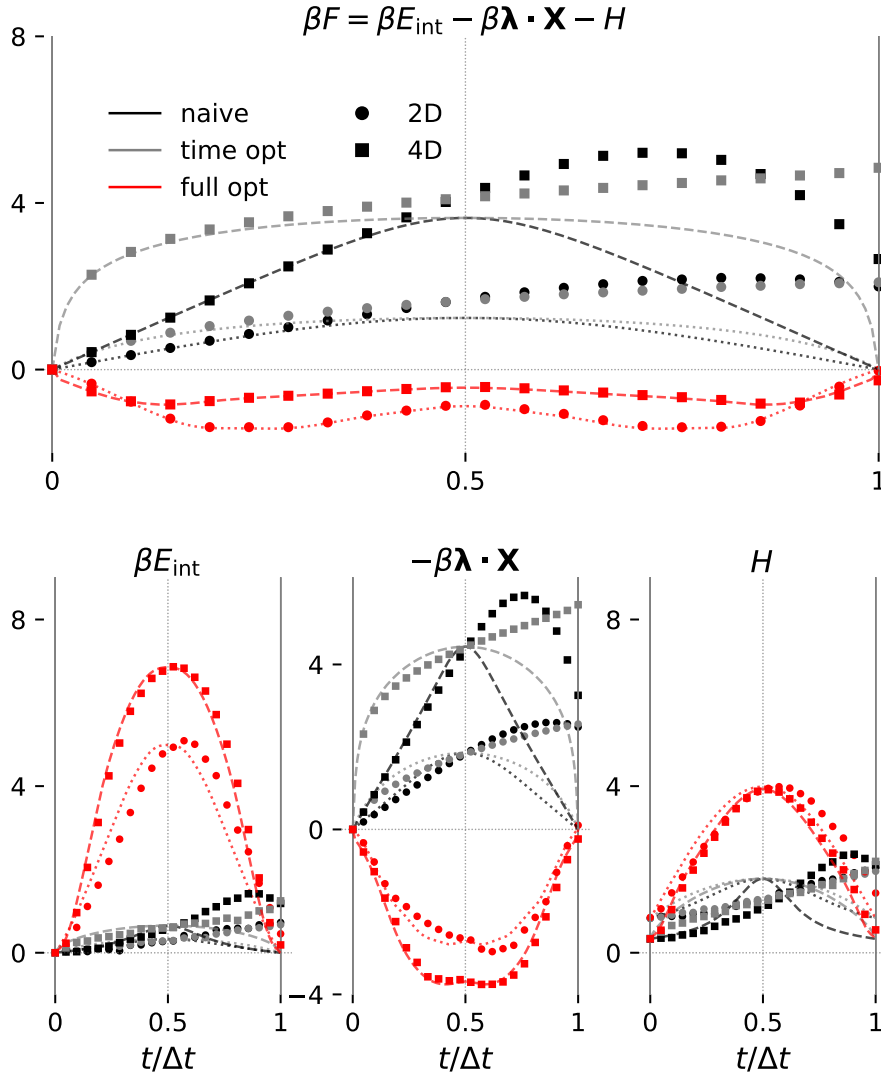
Figure 4.8: Free energy (top), and its components the mean internal energy (4.1) (bottom left), mean external energy (bottom middle), and entropy (bottom right), of equilibrium (curves) and nonequilibrium (points) distributions, as a function of elapsed time t along a moderate-duration ($\Delta t \approx \tau_{\mathrm{rel}}$) protocol of each type.

contrast to naive and time-optimized protocols which primarily fluctuate near the all-up and all-down configurations.

The flattened total-energy landscape and increased entropy during the fully optimized protocol results in relatively constant free energy, requiring comparatively low-magnitude quasistatic work throughout the protocol. In contrast, the naive and time-optimized protocols drive the system through a substantial free-energy barrier, requiring quasistatic work input to the system during the first half of the protocol and quasistatic work extraction during the second half.

4.4.4 Heterogeneously inputting and extracting work, coinciding with system relaxation

I now investigate how fully optimized protocols drive each spin set differently than naive and time-optimized protocols, by splitting total work into quasistatic and excess works by each control-parameter component:

$$\mathcal{W} = \sum_i [\mathcal{W}_i^{\text{qs}} + \mathcal{W}_i^{\text{ex}}] \tag{4.4a}$$

$$= \sum_i \left[\int_0^{\Delta t} \mathrm{d}t\, \mathcal{P}_i^{\text{qs}}(t) + \int_0^{\Delta t} \mathrm{d}t\, \mathcal{P}_i^{\text{ex}}(t) \right] \tag{4.4b}$$

$$\approx \sum_i \left[-\int_0^{\Delta t} \mathrm{d}t\, \dot{\lambda}_i(t) \langle X_i \rangle_{\boldsymbol{\lambda}(t)}^{\text{eq}} + \int_0^{\Delta t} \mathrm{d}t \sum_j \dot{\lambda}_i(t) \zeta_{ij}(\boldsymbol{\lambda}(t)) \dot{\lambda}_j(t) \right]. \tag{4.4c}$$

The final line uses the linear-response approximation (2.64). The quasistatic works of all components sum to the path-independent equilibrium free-energy difference between control-protocol endpoints, $\sum_i \mathcal{W}_i^{\text{qs}} = \Delta F$; however, the quasistatic work of each component is path-dependent.

Figure 4.9 shows components of the quasistatic power $\mathcal{P}_i^{\text{qs}}(t)$ and finite-duration power $\mathcal{P}_i(t)$, for moderate protocol duration ($\Delta t \approx \tau_{\text{rel}}$). During naive and time-optimized protocols, each field inputs and extracts work equally, inputting work in roughly the first half of the protocol to increase the system's free energy and extracting it in the second half to reduce the free energy (Fig. 4.8). For finite duration, the mean conjugate force lags its equilibrium mean (Fig. 4.5), and the quasistatic work is not fully extracted in the second half of the protocol, resulting in excess work. The fully optimized protocols input and extract work differently through the different fields, with zero power for each component where either the field velocity is zero or the spin magnetization for that field is zero (i.e., controlled spins on average flip from down to up).

The 2D fully optimized protocol is roughly divided into three stages according to sign changes of the power. Until $t/\Delta t \approx 0.3$, the green field is increased while the blue field is decreased (Fig. 4.3), inputting power to the green spins to increase their mean magnetization (Fig. 4.5) while power is extracted from the blue spins (which do not significantly change mean magnetization), so that overall nearly no power is input to the system and the free energy is relatively constant (Fig. 4.8). Second (until $t/\Delta t \approx 0.7$), both fields are increased and work is input to the blue spins to increase their magnetization, while work is extracted from the green spins as they are stabilized in spin-up. Finally, the green field is reduced and work is input to the green spins, while increasing the blue field stabilizes the spin-up configuration of the blue spins and extracts work: the two fields act in concert to keep the free energy more constant. This elaborate and heterogeneous schedule of work input and extraction could reflect the system's energetic requirements to fluctuate into configurations
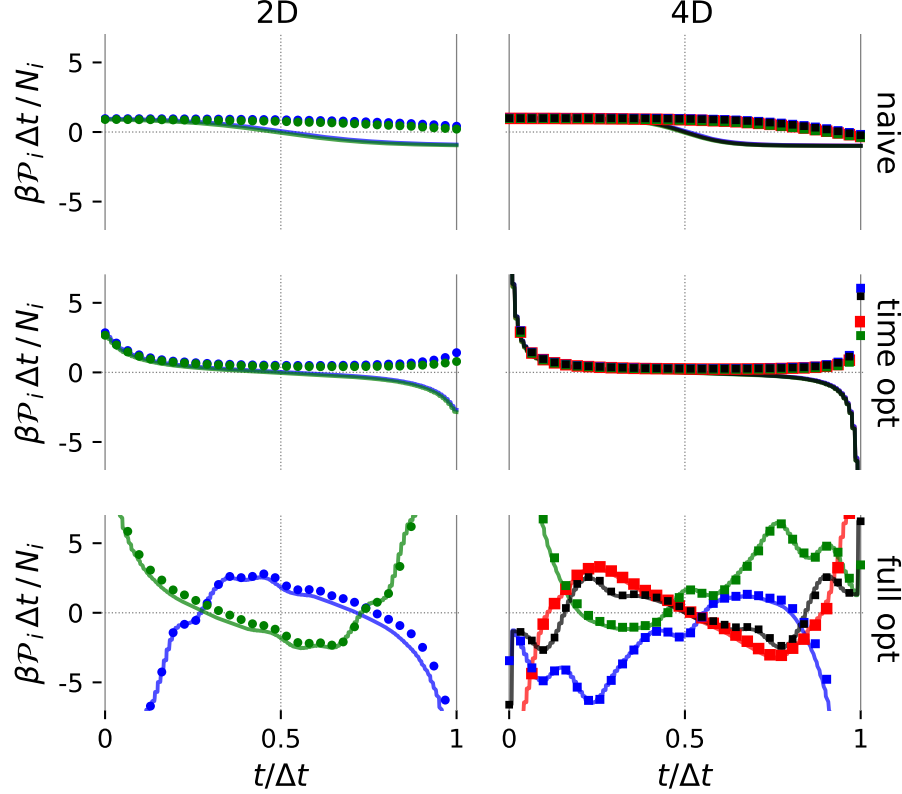
Figure 4.9: Quasistatic power (curves) and finite-duration power (Eq. (4.4b), points) of each field (colors, see Fig. 4.1) as a function of elapsed time t along a moderate-duration ($\Delta t \approx \tau_{\text{rel}}$) protocol of each type (see Fig. 4.3). Power $\mathcal{P}_i$ is scaled by the number N_i of spins controlled by field i.

with high internal energy: during naive and time-optimized protocols the system must wait for appropriate fluctuations from the environment to overcome the internal energy barrier separating the all-down and all-up configurations in order to relax to equilibrium; in contrast, the fully optimized protocols externally provide the appropriate energy to each spin as work, allowing the system to access high-internal-energy configurations and significantly decreasing the time required for system relaxation. The 4D fully optimized protocol shows similar features to 2D in the green and blue power, but events happen relatively earlier in green and later in blue.

Fig. 4.10 shows the work $\mathcal{W}_i$ by each field i as a function of protocol duration, asymptoting to the instantaneous work ($-[\lambda_{i,\text{final}} - \lambda_{i,\text{initial}}]\langle X_i \rangle^{\text{eq}}_{\boldsymbol{\lambda}(t=0)} \approx 1\,k_{\text{B}}T$ per spin) for short durations and reaching the (path-dependent) quasistatic work in the long-duration limit. For the naive and time-optimized protocols, each field does the instantaneous work until $\Delta t \approx \tau_{\text{rel}}$ before crossing over to near-zero quasistatic values for longer durations.

The fully optimized protocols input work via the green field and extract work via the blue field in the quasistatic limit. Work input from the green field increases from instanta-
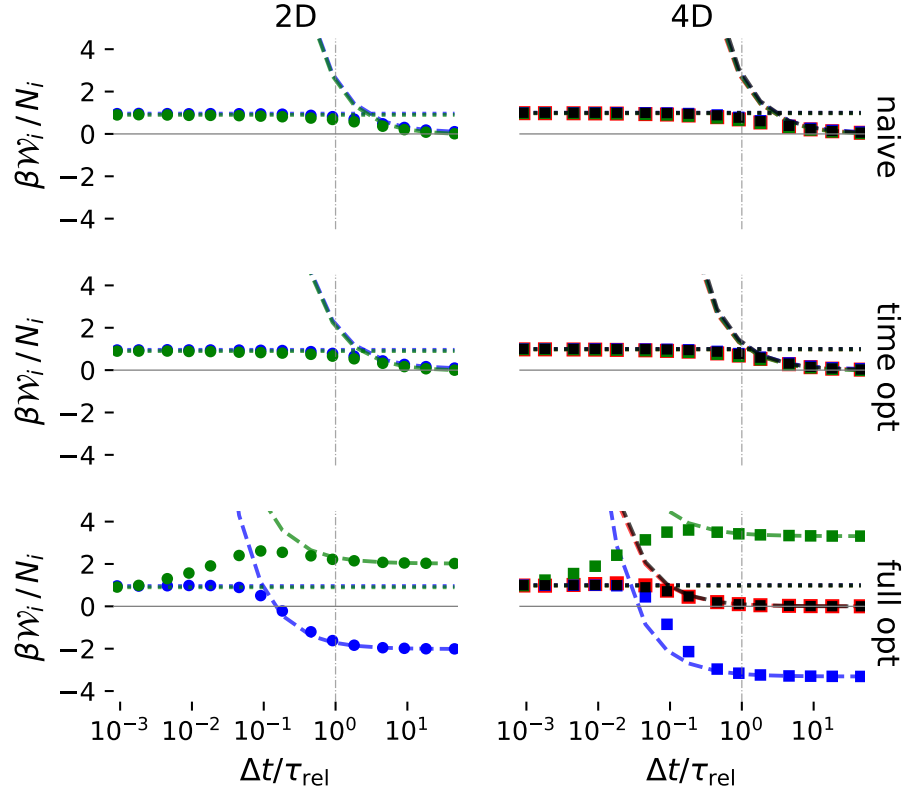
Figure 4.10: Work (4.4a) as a function of protocol duration Δt for each field (colors, see Fig. 4.1), for each protocol type (see Fig. 4.3). Dotted lines: instantaneous work. Dashed curves: linear-response approximation. Protocol durations are scaled by reference spin-inversion relaxation time τ_{rel}.

neous to quasistatic values at short durations while work inputs from other fields remain near instantaneous values. When the duration is sufficiently long for the system to relax ($\Delta t \approx 0.1\tau_{\mathrm{rel}}$), the work from the remaining fields crosses over from instantaneous to quasistatic values, with the blue field extracting work from the system for moderate-to-long duration. This suggests that work input by the green field early in the protocol is transduced through the system to be extracted by the blue field later in the protocol. Driving the green spins to flip early in the protocol initiates the transition mechanism, with further work input to red and black spins continuing the complete spin inversion by allowing the system to access high-internal-energy configurations, and final work extraction from blue spins stabilizing the system in the low-internal-energy all-up configuration. Providing work to specifically support this transition mechanism—rather than simply homogeneously inputting and extracting work to each spin type—keeps the system close to equilibrium for durations ~10× shorter than naive and time-optimized protocols.

Figure 4.11 shows the excess power by each field for moderate duration ($\Delta t \approx \tau_{\mathrm{rel}}$), as well as linear-response approximations (4.4b). The excess power during naive and time-
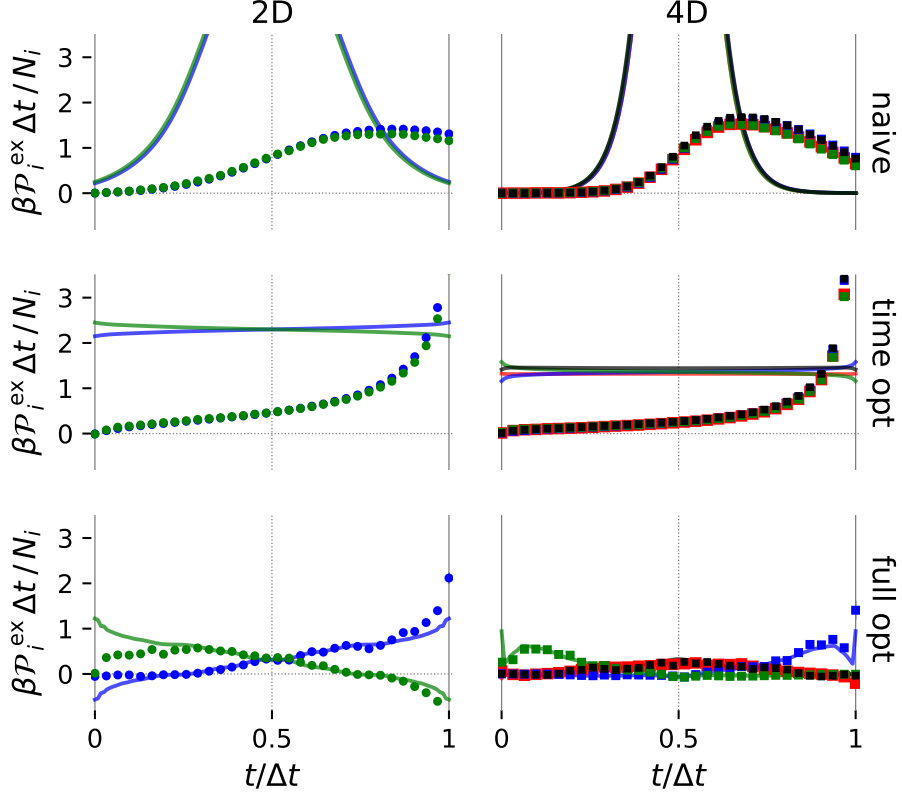
Figure 4.11: Exact excess power (points, Eq. (4.4b)) and linear-response approximation (curves, Eq. (4.4c)) for each field (colors, see Fig. 4.1) as a function of elapsed time t along a moderate-duration ($\Delta t \approx \tau_{\mathrm{rel}}$) protocol of each type (Fig. 4.3).

optimized protocols is essentially equal for each field, consistent with the field velocities (Fig. 4.3) and lag of conjugate forces (Fig. 4.5) being equal for each component in these protocols. The linear-response approximation for the excess power during naive protocols peaks in the middle where friction is largest (Fig. 4.2). In time-optimized protocols, the linear-response approximation to the excess power is essentially the same for each field and constant, flattening the excess-power peak in naive protocols by reducing protocol velocity in this region. The excess power at this moderate duration (significantly distant from equilibrium) unsurprisingly differs from the linear-response approximation for both naive and time-optimized protocols. Nevertheless, the excess power during the fully optimized protocol is significantly reduced relative to both naive and time-optimized protocols and agrees well with the linear-response approximation, indicating that the system remains much closer to equilibrium.

4.5 Discussion

The naive and time-optimized protocols (which follow the same path through multidimensional control-parameter space) change all fields simultaneously, biasing the total magnetization to drive the system from all-down to all-up. The barrier separating the energetically metastable configurations remains large throughout the protocol, and for moderate-duration protocols, the slow transitions between these endpoints prevent the system from relaxing to equilibrium, resulting in significant lag and large excess work. The time-optimized protocol reduces velocity where the friction is large to allow more time for system relaxation, which for moderate-to-long durations reduces the excess work relative to the naive protocol.

In contrast, fully optimized protocols in multidimensional control-parameter spaces have more flexibility to manipulate the system's total energy (4.2); they use this flexibility to flatten the total-energy landscape through a heterogeneous schedule of work input and extraction through each spin (Figs. 4.8, 4.9), also speeding up system transitions from all-down to all-up. Both these equilibrium thermodynamic and kinetic properties are captured by the generalized friction matrix through the respective matrices of force covariance and integral relaxation time (2.67). Both components of the friction play a role in improved understanding of minimum-work control in multidimensional control-parameter spaces.

The force covariance matrix is a property of the equilibrium distribution without any reference to the system's dynamics. It quantifies the curvature of equilibrium free energy in control-parameter space [67]. For systems where the integral relaxation time is constant, the force covariance matrix replaces the generalized friction as the metric on control-parameter space that quantifies excess work, indicating that constant free-energy paths minimize work. Although the integral relaxation time is not constant throughout control-parameter space for this system (Fig. 4.2c), fully optimized protocols still have relatively constant free energy relative to naive and time-optimized protocols (Fig. 4.8), a factor that helps to reduce work.

Off-diagonal elements of the force covariance matrix in fully optimized protocols are significantly reduced relative to naive and time-optimized protocols (Fig. 4.7), reflecting that the system accesses configurations which are otherwise inaccessible due to high internal energy. This has interesting connections to strategies for optimal inference of system parameters [129]. There, a spin system is perturbed by external fields to new equilibrium ensembles that allow more efficient inference of system coupling constants (i.e., fewer samples are required to estimate parameters to a given precision), and the perturbation is updated using the Fisher information matrix for desired system parameters. Inference is optimal at field strengths that decorrelate spins, allowing the system to occupy otherwise-inaccessible configurations. Since the excess work during the protocol is known to affect the efficiency of free-energy estimation [60, 61, 62] and the force covariance matrix is proportional to the Fisher information matrix [67], it is not surprising that parallels in design strategy exist. However, a key distinction is the dynamical nature of control protocols, where the relax-

ation kinetics between equilibrium distributions of successive control parameters is relevant; thus the importance of the generalized friction, which captures relaxation timescales via the integral relaxation time matrix.

The fully optimized protocols reveal a clear mechanism for spin inversion. First, work is input to green spins (initially energetically frustrated) causing them to flip, which together with direct work input by red and black fields drives neighboring red and black spins to flip. At the midpoint of the 4D fully optimized protocol, both green spins are up, red and black spins are evenly split between up and down, and blue spins are both down. Past the midpoint, work is extracted from red and black spins by increasing the corresponding fields to stabilize spins in the up orientation, which together with work input to neighboring blue spins causes blue spins to flip. Finally, work is extracted from blue spins to stabilize the system in the all-up configuration.

Using this strategy, the system reaches the linear-response regime for durations $\sim 10\times$ shorter than in the naive and time-optimized protocols (Fig. 4.10). Excess work scales as Δt^{-1} for long-duration protocols where the system is in the linear-response regime. A protocol that enables fast relaxation so the system reaches this regime and excess-work scaling at relatively short duration results in work reduction relative to a naive protocol that reaches the linear-response regime at longer duration. The fully optimized protocols appear to facilitate such fast-relaxing paths in configuration space, reducing the time spent in the initial metastable basin before making a transition over the energy barrier. Earlier studies showed an empirical similarity between the minimum-work protocol and the minimum free-energy path [103, 94], which represents the spontaneous (in the absence of driving) transition path between metastable mesostates. The minimum-work protocols here also drive the system along a plausible spontaneous transition path: green spins flip first, followed by red and black spins, and finally blue spins. There may be a deeper connection between minimum-work protocols and spontaneous transition paths of the unperturbed system, that would allow extraction of kinetic information about transition paths from minimum-work protocols or use kinetic information to design efficient protocols. A more detailed quantitative comparison is necessary to elucidate this connection.

Chapter 5

Connections between Minimum-Work Protocols and Transition Paths

5.1 Introduction

Here the correspondence between the system's response to a minimum-work protocol and its dynamics during a spontaneous transition is investigated. The 3×3 ferromagnetic Ising system (Fig. 5.1) is examined as it spontaneously transitions from the all-down to all-up configuration with no external magnetic fields ($\boldsymbol{\lambda} = 0$), and as it is driven between all-down and all-up configurations by a set of time-dependent magnetic fields (Chapter 4). This chapter focuses on the 4D fully optimized protocol since it drives all spins in the system and is the highest-dimension protocol calculated for this system. A sufficiently long protocol duration, $\Delta t = \tau_{\mathrm{ref}}$, such that the system is in the linear-response regime is considered so that the system is near equilibrium (although excess work is still done on the system).

Two aspects of the reaction are of interest to this study: the reaction mechanism and its transient energetic costs. In this context, the reaction mechanism is a sequence of spin flips that brings the system from all down to all up; since there are 9 spins, there are 9 steps to the reaction mechanism. However, the system response to control protocols and its spontaneous transition paths are both ensembles of trajectories where any spin can (and frequently does) flip orientations multiple times throughout a single trajectory. Section 5.2 first considers how information about the reaction mechanism can be extracted from the trajectory ensembles.

The energetic cost of the mechanism is also characterized. Control protocols do work on the system, providing an energetic bias that produces a system response. Chapter 4 shows that during the fully optimized 4D protocol, work is done by green fields early in the protocol, which flips green spins and increases the system's internal energy, and then later blue fields extract work from blue spins as they flip and decrease the system's internal energy (Fig. 4.10). Spontaneous transition paths similarly take in heat to overcome the energy

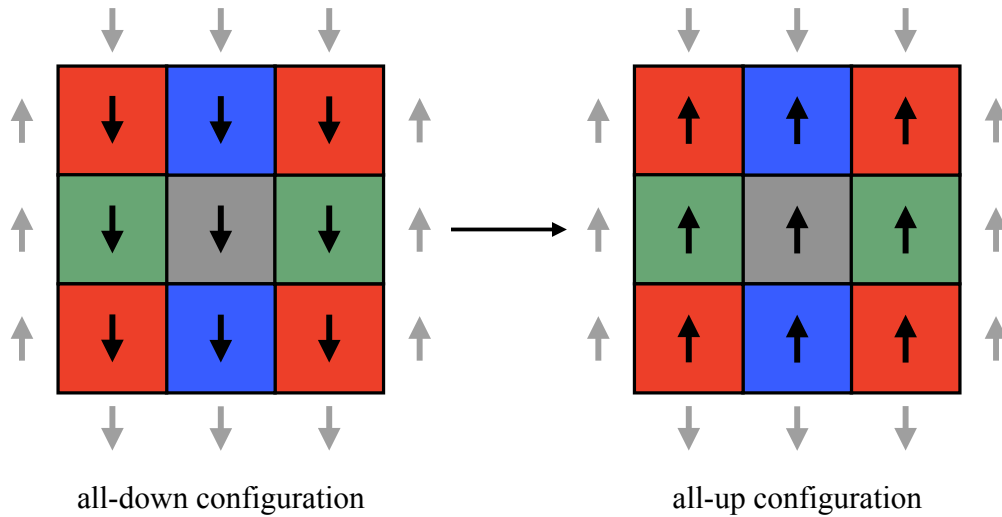all-down configuration   all-up configuration

Figure 5.1: Schematic of 3×3 Ising model with 9 fluctuating spins (black, shown in initial all-down configuration) and 12 fixed boundary spins (gray). Spins are colored according to their symmetry type.

barrier, but must wait for the appropriate fluctuations from the environment to complete a successful transition path. The random fluctuations that do give rise to transition paths are characterized by transition rates which are biased relative to the equilibrium detailed-balance transition rates, as described in Section 3.2, and therefore the heat flows associated with these conditional transition rates can be determined. Energy flows into the system can also be assigned to different degrees of freedom, which provides some insight into the reaction mechanism [116, 130]. These energetic changes in the Ising system will be examined during both the driven and spontaneous processes.

5.2 Trajectory ensembles from minimum-work protocols and the transition-path ensemble

By repeating a control protocol many times, an ensemble of system responses can be generated that (roughly) transit from the all-down to all-up configurations. A sample of 5000 trajectories is generated in response to a control protocol by simulating the 4D fully optimized protocol for $\Delta t = \tau_{\mathrm{rel}}$ as described in Appendix B.

Similarly, the transition-path ensemble (TPE) is an ensemble of reactive trajectories from all down to all up. To generate an ensemble of transition paths, the forward committor is calculated by solving the linear equation (Eq. (2.31))

$$\sum_{\phi'} T_{\phi'\phi} q(\phi') = 0 \tag{5.1}$$

with boundary conditions $q(\phi) = 0$ for ϕ the all-down configuration and $q(\phi) = 1$ for ϕ the all-up configuration. Because the forward and backward committors for an equilibrium system are related by detailed balance, the superscript in $q^+(\phi)$ is eliminated and the forward committor is referred to as the committor throughout this chapter.

After calculating the committor for the Ising system, trajectory segments belonging to the transition-path ensemble are generated using modified transition rates (Eq. 3.20, [90]). A trajectory is initialized in the all-down configuration $\phi' = A$, then the next state ϕ outside of A is chosen with probability

$$p^{\text{initial}}(\phi) = \frac{T_{\phi\phi'} q(\phi)}{\sum_{\phi''} T_{\phi''\phi'} q(\phi'')} \,, \tag{5.2}$$

where the denominator normalizes this probability over all possible transitions out of A. Subsequent states are generated using modified transition rates, which for the forward TPE are

$$T^{\text{R}}_{\phi\phi'} = T_{\phi\phi'} \frac{q(\phi)}{q(\phi')} \,. \tag{5.3}$$

Trajectories are terminated once they reach B for the first time. A sample of $40,000$ transition paths are generated using this method. Figure 5.2a shows a sample of these transition paths and trajectories generated from protocol simulations by projecting the state space onto the committor coordinate.

The trajectory ensemble generated by protocol simulations and the transition-path ensemble differ in the duration of trajectories and the configuration at the beginning and end. In the TPE, the endpoint configurations of each trajectory are the same by definition (the all-down and all-up configurations, respectively), but the duration of each trajectory differs. In contrast, each trajectory generated by the protocol is of equal duration, but the start and end configurations are not constrained, as discussed below.

The difference in endpoints between the processes can be approximately resolved by choosing control parameters and protocol endpoints that restrain the system to configurations corresponding to the reaction endpoints. For the Ising model, the reactant A is defined to be the all-down configuration and the product B to be the all-up configuration; these configurations have committor values $q(\phi \in A) = 0$ and $q(\phi \in B) = 1$, respectively. For the Ising model, the 4D fully optimized protocol starts with all fields at $\lambda_{i,\text{initial}} = -0.5k_{\text{B}}T$ and ends with $\lambda_{i,\text{final}} = 0.5k_{\text{B}}T$, whose equilibrium distributions are dominated by the all-down or all-up configuration, respectively (which leads to mean system magnetization close to -1 and 1). The system does not reach the equilibrium distribution by the end of the protocol, but for sufficiently long protocols will be "close" to this endpoint. The system's distance from the equilibrium distribution depends on the protocol duration, but since minimum-work protocols in the linear-response regime are independent of protocol duration, a single protocol duration ($\Delta t = \tau_{\text{rel}}$) in this regime is used as a representative of the process for
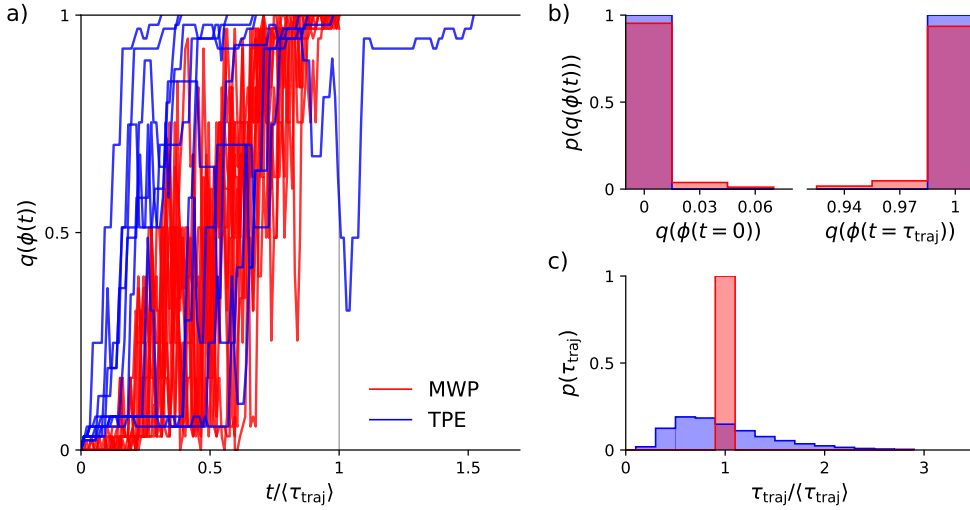
Figure 5.2: Trajectory ensembles from the minimum-work protocol and transition-path ensemble. a) 10 sample trajectories generated by an MWP (red) and from the transition-path ensemble (blue), projected onto the committor $q(\phi)$. The trajectory time is scaled by the mean duration τ_{traj} of trajectories in the respective ensemble. b) Distribution of start and end configurations for trajectories in each ensemble. c) Distribution of trajectory durations for trajectories in each ensemble.

all protocols in the linear-response regime. Figure 5.2b shows that the distribution of committor values for configurations at the start and end of the protocol performed in duration $\Delta t = \tau_{\text{ref}}$ are highly peaked in the $q = 0$ and $q = 1$ states respectively, indicating that on average the trajectories generated by a protocol of this duration transit between the reaction endpoints.

Resolving the range of transition-path durations requires a different approach. Because the system must wait for appropriate fluctuations from the environment to make progress during a transition from all down to all up, each step requires a stochastic amount of time and the overall transition paths have a range of durations. Additionally, the system can move "backward" during the trajectory, leading to several recrossings of a dividing surface between A and B. (Note that these are the same recrossing events that are ignored in transition-state theory and minimized in variational transition-state theory through optimization of the dividing surface). Combining information from multiple trajectories is therefore challenging, as at a given time t since the start of the transition path, one trajectory could be significantly closer to the product than another. This contrasts with control protocols where the external fields are the same at time t for all trajectories, and for long-duration protocols, the system state will be close to equilibrium and therefore relatively consistent across multiple trajectory samples.

76

If the reactive trajectories are described by a one-dimensional coordinate $x(t)$ (for example, from single-molecule force spectroscopy [131]), there are methods to combine temporal information from trajectories of multiple durations to obtain a mean transition path time for each value of the one-dimensional coordinate, $\langle t(x) \rangle$, using various ideas based on mean first-passage times [132, 133, 134, 135]. However, an average one-dimensional trajectory is not of interest to the present analysis since general mechanistic information is desired in terms of physically intuitive collective variables (here given by each spin's magnetization).

Two approaches are therefore used for averaging information from the transition-path ensemble, one that aligns the time axis of each trajectory and one that aligns committor values. The time in each trajectory can be linearly rescaled by its total duration τ_{traj} to get a reaction coordinate $t/\tau_{\mathrm{traj}} \in [0, 1]$ that represents the proportion of time elapsed during the trajectory, which is termed the *scaled transition-path time*. In practice, the scaled transition-path time t/τ_{traj} is divided into 20 discrete bins and the mean properties in each bin for each trajectory is determined, then averaged over all trajectories, ensuring that long trajectories are not disproportionately weighted in the ensemble average [134]. This is a relatively *ad hoc* approach to rescaling transition paths of multiple durations that would benefit from further improvement, but it nevertheless captures mechanistic information.

Another approach uses the committor, which in principle contains the information about the reaction mechanism in the transition-path ensemble since it is the "ideal" reaction coordinate. For a continuous system, visualizing the isocommittor surfaces throughout a collective-variable space would provide information about the order of events and the bottlenecks of the reaction. However, the Ising model has a discrete state space and there is no guarantee that a particular isocommittor surface is populated by any system configurations (for example, there is no configuration with $q(\boldsymbol{\phi}) = 0.5$). Instead, the range of committor values is divided into discrete bins and all configurations from each trajectory that fall in the same bin are grouped together. Such an approach allows calculation of the mean properties of each spin during the transition path. Averaging over the committor coordinate directly tends to under-emphasizes configurations at the start and end of the transition mechanism, and therefore a nonlinear transformation of the committor is used to make these configurations more visible. Inspired by the mathematical form of TPE entropy production in Chapter 3, the committor is transformed by the mapping

$$f(\boldsymbol{\phi}) = \ln \frac{q(\boldsymbol{\phi})}{1 - q(\boldsymbol{\phi})} \, , \tag{5.4}$$

and configurations that fall into 9 evenly spaced bins for this coordinate are grouped together. Note that an invertible transformation of the committor does not affect the information contained in the one-dimensional coordinate, and therefore $f(\boldsymbol{\phi})$ is still a "perfect" reaction coordinate.
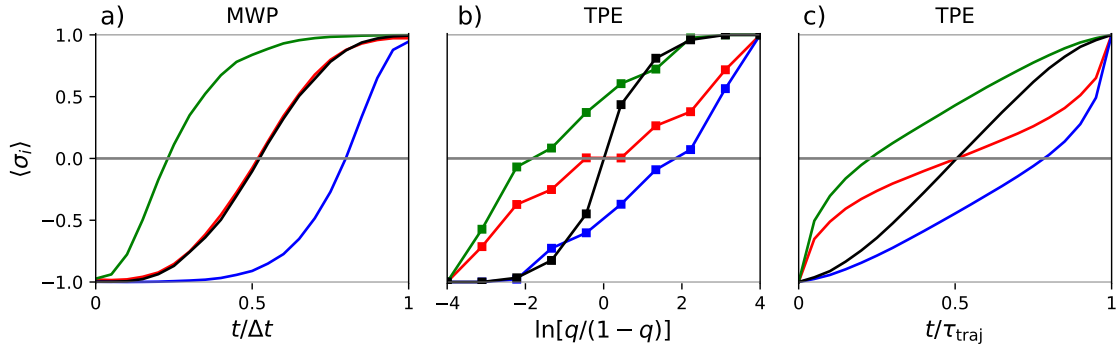
Figure 5.3: Mean magnetization of each spin set (colors in Fig. 5.1) a) at scaled time $t/\Delta t$ during the minimum-work protocol, b) binned to the rescaled committor $\ln[q/(1-q)]$ during the transition-path ensemble, and c) at scaled transition-path time t/τ_{traj} during the transition-path ensemble.

5.3 Reaction mechanism for spin inversion

Figure 5.3 shows the mean state of each spin set during the 4D fully optimized protocol and during the transition path as parameterized by the scaled committor or scaled transition-path time. The two methods of averaging the mean state in the transition-path ensemble look very similar, but using scaled transition-path time allows one to use more discrete bins and obtain smoother curves. The mean state of the system during the 4D fully optimized protocol and during the transition-path ensemble show similar characteristics, but differ in their finer details. In general, the green spins (which are initially energetically frustrated due to the adjacent boundary spins of opposite sign) flip relatively early in both processes, and the blue spins (which end in an energetically frustrated orientation) flip relatively late. The red and black spins flip throughout the middle of the protocol. However, the finer details of the processes differ; the MWP seems to flip both green spins around the same time (early in the protocol), while in the TPE at least one red spin flips relatively early, on average before the second green spin.

The differences in the detailed mechanisms may be due to the constraint on driving in the 4D control parameter space where, e.g., all four red fields are changed in the same way throughout the protocol. This constraint means it is not possible for the protocol to drive one spin to the up orientation independently of other spins in the same spin set. The resulting protocol in the 4D space is therefore symmetric in the identity of each of the spins. In contrast, trajectories in the transition-path ensemble seem to break this symmetry, but the average state of each spin as a function of the rescaled committor or scaled transition-path time does not show this symmetry-breaking.

Therefore, a method is developed to determine mechanistic information that distinguishes between spins in the same spin set from the trajectories generated by MWPs or
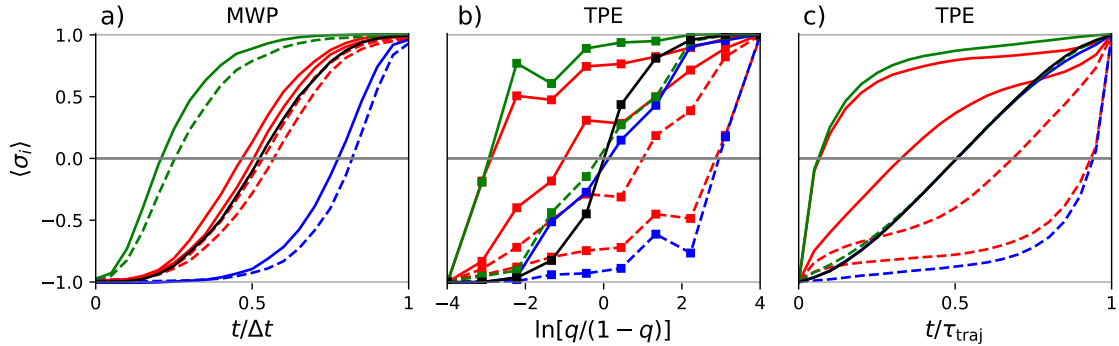
Figure 5.4: Symmetry-broken mechanism for the spin-inversion transition. The mean state of each spin during the a) MWP, b) TPE as a function of the scaled committor, and c) TPE as a function of scaled trajectory time t/τ_{traj}. The first spin(s) to flip in each spin set are indicated by solid curves and the last spin(s) are indicated by dashed curves.

that form the TPE. Since the order of spin flips during the mechanism is of interest, the order l_α of spin flips (with $\alpha = 1 \dots 9$) is determined for each trajectory that gives the identity of the α-th spin to flip during a trajectory. Each spin may flip multiple times in a single trajectory, so to get an average order for the trajectory, the proportion of time spent in the up orientation for each spin is calculated, then all spins are ordered in descending proportion: a spin that spends more time during the trajectory in the up orientation than another is identified as flipping earlier in the reaction (has a lower value of α). There are four spin types identified by colors in Fig. 5.1. To break this symmetry, nine spin types are identified: the first, second, third, and fourth red spin; the first and second green spin; the first and second blue spin; and the black spin. These coarse variables still do not contain exact geometric information (e.g., they do not specify where the second red spin is located relative to the first red spin), but do break the symmetry within a single trajectory. The same analysis is performed on all trajectories in the ensemble to break the symmetry between spins in the same spin set. Once the order of each spin in the trajectory is determined so that spins in the same spin set are ordered, the mean magnetization of each spin type can be averaged over all trajectories.

Figure 5.4 shows the average trajectory during the MWP and TPE with broken symmetry for spins in the same spin set. In the MWP, each spin-flip event happens at relatively distinct times during the protocol: both green spins flip early, then all red and black spins flip, then finally both blue spins. Spins in the same spin set flip at approximately the same time, reflecting the symmetry of the driving protocol. In contrast, the TPE shows symmetry breaking for spins in the same spin set: a green and red spin flip first; followed by a second red spin; then the second green, first blue, and black spins; then the third red spin; and finally the fourth red and second blue spins. Such symmetry breaking cannot be achieved by a 4D protocol, therefore the mechanism observed in the 4D MWP cannot precisely match
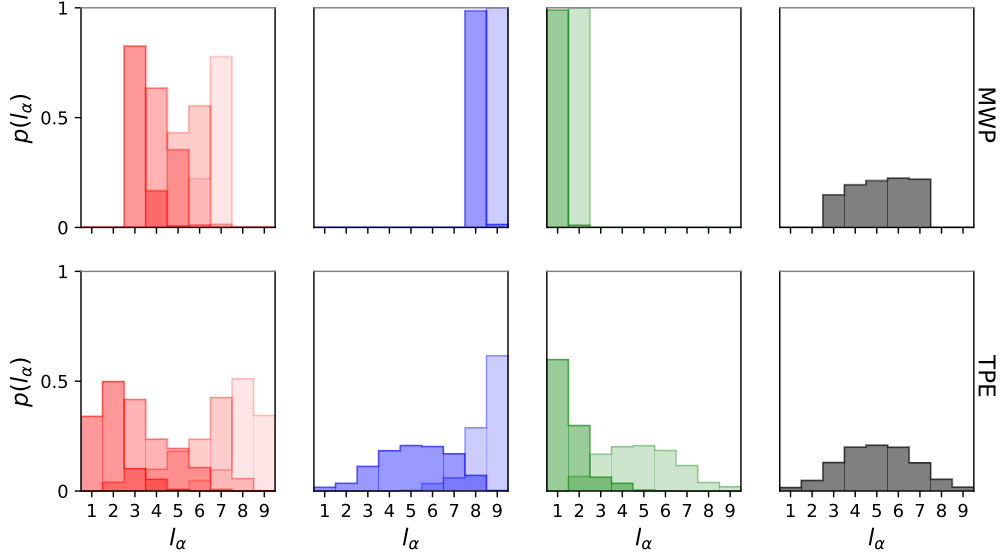
Figure 5.5: Distribution of spin-flip order l_α for each spin type during the MWP (top) and TPE (bottom). Each histogram shows the distribution of the spin-flip order for a given spin, with spins in the same set separated based on their flipping order in the trajectory. The earlier spin(s) of each spin type have darker shading in the histogram, while later spin(s) are lighter.

the mechanism observed in the TPE. However, the general feature of green spin(s) flipping early and blue spin(s) flipping late is preserved in both processes, reflecting the bias imposed by the fixed boundary spins that affects both the geometry of control-parameter space and the spontaneous transition mechanism.

While Fig. 5.4 separates spin flips in the same spin set, it does not show the overall order of spin flips during the trajectory. This is shown in Fig. 5.5 by plotting the distribution of the spin-flip order l_α of each spin type relative to all others. The MWP preserves the order across all trajectories, with both green (blue) spins flipping first (last), and red and black spins all flipping in between. The TPE spin-order distributions, however, show higher diversity. The first red and first green spins predominantly flip first, and the second blue and last red predominantly flip last, but the order of other spins is more variable. This indicates that there are many paths taken (with relatively high probability) by the system with spin flips occurring in different orders. It is possible that the TPE transition mechanism in this system cannot be characterized by a single ordering of spin flips, and instead some other physically intuitive collective variables that are able to capture the variety of transition paths would be better suited to describing the mechanism.

80

5.4 Energy flows during the minimum-work protocol and transition-path ensemble

Control protocols produce changes in the system configuration by doing work. In the quasi-static limit, a protocol drives the system over an internal energy barrier by first providing some of this energy as work to increase the system's internal energy and drive it over the barrier, then extracting work from the system as it decreases in energy. Transition paths overcome an energy barrier by waiting for a sequence of fluctuations from the environment that provide the necessary energy as heat to the system. It seems intuitive that if the system must overcome some internal energy barrier while changing between collective-variable endpoints, then a "good" control protocol would provide this energy to the system as work rather than waiting for the appropriate (rare) thermal fluctuations from the environment, essentially flattening the total energy landscape along the path the system takes through configuration space. Here, the energy flows into the system during each trajectory are determined and split between motion in each spin. The idea of analyzing energy flows in the transition-path ensemble has been discussed previously [116, 130] where energy flows are shown to be helpful for determining collective variables that may be relevant to the reaction.

During a transition path, no work is done on the system and therefore changes in internal energy during the trajectory is due to heat flow. For a trajectory $\{\phi(t_k)\}$ with $k = 1 \dots \tau_{\text{traj}}$, the total heat flow to the system is

$$Q(\tau_{\text{traj}}) = \sum_{k=1}^{\tau_{\text{traj}}-1} [E_{\text{int}}(\phi(t_{k+1})) - E_{\text{int}}(\phi(t_k))] = E_{\text{int}}(B) - E_{\text{int}}(A) , \qquad (5.5)$$

which is equal to the difference in internal energy of B relative to A. For the Ising system, $\Delta E_{AB} = E_{\text{int}}(B) - E_{\text{int}}(A) = 0$ due to the underlying symmetry of the metastable states. Since the single-spin-flip dynamics that are implemented are multipartite (a multidimensional extension of the bipartite dynamics condition in Section 3.3 where only one coordinate moves in a given time step while all others remain stationary), the total heat flow can be split between each spin to better understand the energetic costs associated with each spin's dynamics during the mechanism. The heat flow to spin σ_i during the trajectory is

$$Q^i(\tau_{\text{traj}}) = \sum_{k=1}^{\tau_{\text{traj}}-1} [E_{\text{int}}(\phi(t_{k+1})) - E_{\text{int}}(\phi(t_k))]\{1 - \delta[\sigma_i(\phi(t_{k+1})) - \sigma_i(\phi(t_k))]\} \qquad (5.6\text{a})$$

$$= \sum_{k=1}^{\tau_{\text{traj}}-1} \Delta E_{\text{int}}^i(k \to k+1) \qquad (5.6\text{b})$$

where the final factor in Eq. (5.6a) ensures that only transitions where σ_i changes contribute to the average. The mean heat flow to σ_i is binned to either the scaled transition-path time

or rescaled committor:

$$Q^i(t/\tau_{\text{traj}}) = \sum_{k=1}^{\tau_{\text{traj}}-1} \Delta E_{\text{int}}^i(k \to k+1)\, \delta\left[\frac{t}{\tau_{\text{traj}}} - \frac{k+0.5}{\tau_{\text{traj}}}\right] \tag{5.7a}$$

$$Q^i(f) = \sum_{k=1}^{\tau_{\text{traj}}-1} \Delta E_{\text{int}}^i(k \to k+1)\, \delta\left[f - \frac{f(\phi(t_{k+1})) + f(\phi(t_k))}{2}\right], \tag{5.7b}$$

where the condition $\delta[t/\tau_{\text{traj}} - (k+0.5)/\tau_{\text{traj}}]$ selects all transitions that fall in the discrete time bin centered at t/τ_{traj}, and $\delta[f - \frac{1}{2}(f(\phi(t_{k+1})) + f(\phi(t_k)))]$ selects all transitions that fall in the discrete rescaled committor bin centered at $f = \ln[q/(1-q)]$. The heat flow to spin σ_i can then be averaged over all trajectories to obtain the ensemble-average heat flow $\langle \dot{Q}_{\text{R}}^i \rangle$ to that spin during the trajectory. If heat flow is positive into a particular σ_i, i.e. $\langle \dot{Q}_{\text{R}}^i \rangle > 0$, this suggests that motion in σ_i primarily "activates" the system towards the energy barrier, while $\langle \dot{Q}_{\text{R}}^i \rangle < 0$ implies that motion in σ_i primarily relaxes the system to the product [116].

During the control protocol, both work and heat change the total energy of the system. The work and heat steps during the trajectory are considered separately. Work is identified as the change in total energy $V(\phi, \lambda)$ when control parameters change at fixed system position,

$$\mathcal{W}(\Delta t) = \sum_{k=1}^{\Delta t-1} [V(\phi(t_k), \lambda(t_{k+1})) - V(\phi(t_k), \lambda(t_k))] \tag{5.8a}$$

$$= -\sum_{k=1}^{\Delta t-1} \sum_i [\lambda_i(t_{k+1}) - \lambda_i(t_k)] X_i(t_k) \tag{5.8b}$$

$$= \sum_i \mathcal{W}^i(\Delta t), \tag{5.8c}$$

which can be partitioned among the control parameters λ_i. Heat is the change in total energy when the system state changes for fixed control parameters

$$Q(\Delta t) = \sum_{k=1}^{\Delta t-1} \left[V(\phi(t_{k+1}), \lambda(t_{k+1})) - V(\phi(t_k), \lambda(t_{k+1})) \right] \tag{5.9a}$$

$$= \sum_{k=1}^{\Delta t-1} \left([E_{\text{int}}(\phi(t_{k+1})) - E_{\text{int}}(\phi(t_k))] - \sum_i \lambda_i(t_{k+1}) [X_i(t_{k+1}) - X_i(t_k)] \right) \tag{5.9b}$$

$$= \sum_{k=1}^{\Delta t-1} \sum_i \Big([E_{\text{int}}(\phi(t_{k+1})) - E_{\text{int}}(\phi(t_k))]\{1 - \delta[\sigma_i(t_{k+1}) - \sigma_i(t_k)]\}$$
$$- \lambda_i(t_{k+1}) [X_i(t_{k+1}) - X_i(t_k)] \Big) \tag{5.9c}$$

$$= \sum_{k=1}^{\Delta t-1} \sum_i \left[\Delta E_{\text{int}}^i(k \to k+1) + \Delta E_{\text{ext}}^i(k \to k+1) \right], \tag{5.9d}$$

where in Eq. (5.9d) heat flow due to changes in σ_i is identified as making distinct changes to the internal and external energy. Together, the work and heat change the total energy,

$$\Delta V(\Delta t) = \mathcal{W}(\Delta t) + Q(\Delta t) \tag{5.10a}$$

$$= \sum_i \left[\mathcal{W}^i(\Delta t) + Q^i(\Delta t) \right] \tag{5.10b}$$

which is split into contributions from motion in each control parameter and associated spin magnetizations. Motion in control parameters is deterministic, while motion in spins is stochastic.

Control protocols and transition paths share the internal energy landscape E_{int}. Therefore changes in internal energy during both processes will be analyzed to better understand how the system overcomes the energy barrier. After identifying the order of spin flips in a trajectory, the energy flow associated with each spin in the trajectory is calculated. This allows identification of differences in energy flows to spins in the same spin set that may have different roles in the reaction mechanism.

Figure 5.6 shows the mean heat flow to the system during the minimum-work protocol and during a transition path. When the system moves during the MWP, some of the heat flow changes internal energy and some of it changes external energy. The change in internal energy (Fig. 5.6b) for the green spins is positive during the first half of the protocol, for the blue spins is negative during the second half of the protocol, and for red and black is zero throughout the protocol. The changes in internal energy for the blue and green spins are compensated by changes in the external energy (Fig. 5.6a): the external energy decreases on average when green spins flip (increasing the internal energy) and increases on average when blue spins flip (decreasing internal energy). This demonstrates how external energy provides a force to the system which allows it to access high-internal-energy configurations during the protocol. When control parameters do work they change only the external energy; therefore this biasing force that allows the system to access high-internal-energy configurations is provided by control parameters. Additionally, spins from the same spin set have nearly identical energy flows during the protocol, again reflecting the symmetry of the control parameters.

Figures 5.6c/d shows the heat flow (change in internal energy) during the transition-path ensemble for both scaled transition-path time and re-scaled committor. The symmetry-breaking observed in the reaction mechanism is also reflected in the changes in internal energy during the transition path. The first green and first red spins take in energy on average early in the transition path, bringing the system out of the all-down configuration into higher-internal-energy configurations. Similarly, at the end of the transition path, the second blue and last red spin release energy as the system reaches the low-internal-energy all-up configuration. Throughout the middle of the protocol, the other spins on average take
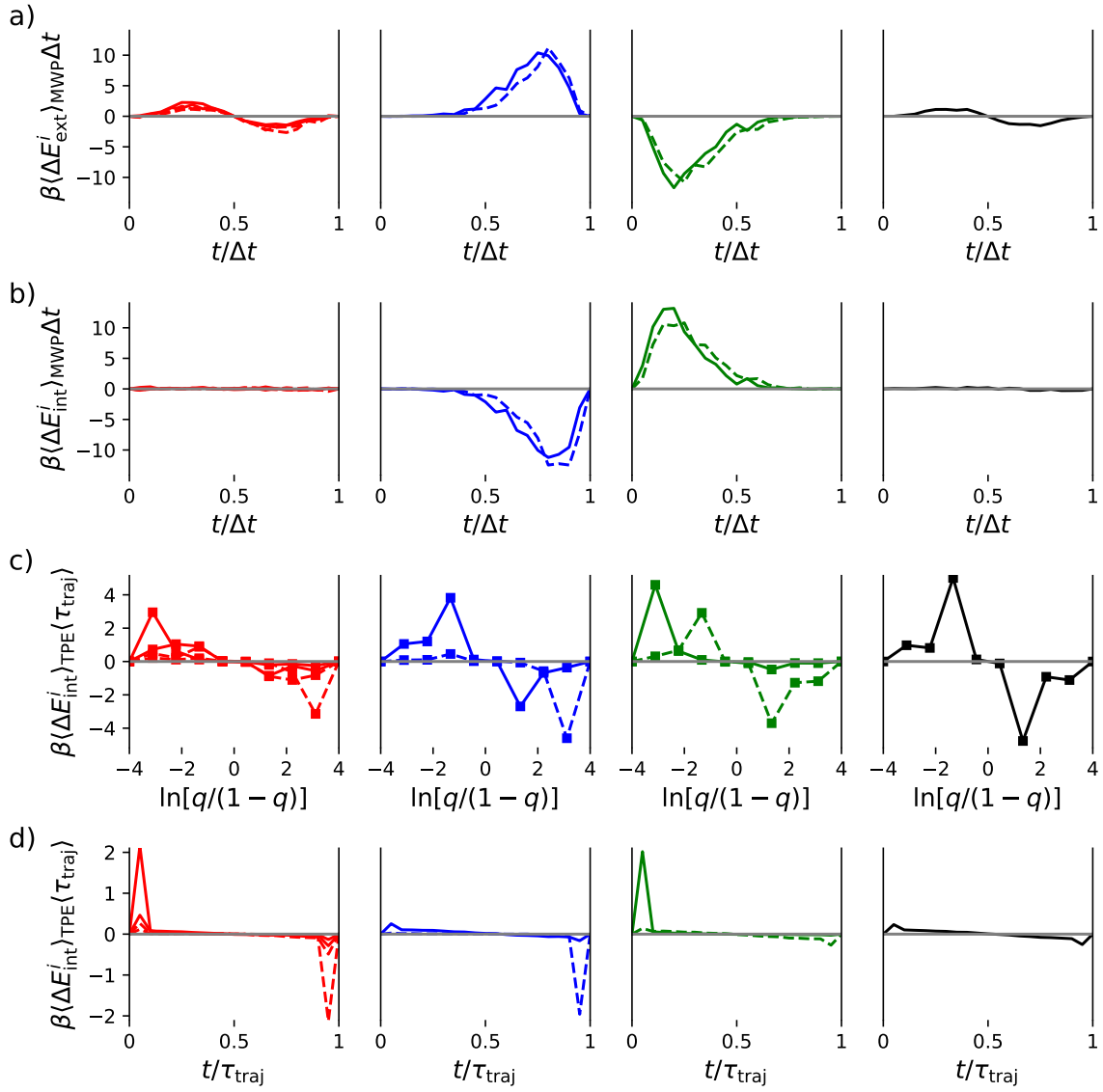
Figure 5.6: Heat flow to each spin during the MWP and TPE. Top two rows: Mean change in a) external energy $\beta\langle\Delta E^i_{\text{ext}}\rangle$ and b) internal energy $\beta\langle\Delta E^i_{\text{int}}\rangle$ due to magnetization change in σ_i during the MWP, scaled by protocol duration Δt. Bottom two rows: Mean change in internal energy $\beta\langle\Delta E^i_{\text{int}}\rangle$ due to magnetization change in σ_i during the TPE c) projected onto the scaled committor and d) as function of scaled trajectory time, scaled by the mean transition-path duration $\langle\tau_{\text{traj}}\rangle$. The first spin(s) to flip in each spin set are indicated by solid curves and the last spin(s) are indicated by dashed curves.

in or release heat. In general, if a spin flips in the first half of the trajectory, the heat flow is positive; if it flips in the second half of the trajectory, the heat flow is negative.

The heat flow during the transition-path ensemble also demonstrates differences in how the scaled transition-path time and rescaled committor provide information about the mechanism. While the mean state in these coordinates looks similar with either representation of transition paths (Fig. 5.4), it is much harder to visualize the heat flows using the scaled transition-path time than the rescaled committor. The heat flow into the system during the first step out of the stable all-down configuration and the heat flow out of the system during the final step into the all-up configuration inevitably occur in the first and last bins of the scaled transition-path time leading to highly peaked heat flow at these times. Throughout the middle of the transition path, heat flow is small for two reasons: first, some of the internal energy changes for transitions throughout the protocol are zero and thus do not contribute magnitude to the heat flow; second, linearly rescaling the transition-path time can result in spin flips occurring at a wide range of scaled transition-path times that are not necessarily aligned between different trajectories, which spreads out the heat flow associated with this transition over multiple scaled trajectory-time bins. For this reason, the rescaled committor is a more desirable reaction coordinate, since it provides a measure of the system's progress between reaction endpoints independent of the time that the system reaches a state in any given trajectory. However, qualitative aspects of the heat flows for the two coordinates are the same.

5.5   Summarizing the mechanisms

Finally, the reaction mechanisms identified for the minimum-work protocol and the transition-path ensemble are summarized.

The minimum-work protocol has a clear and intuitive reaction mechanism, shown in Figure 5.7. In the first two steps, both green spins are flipped. Each of the green spins breaks three energetically favourable interactions and forms one when it flips. The change in internal energy from each interaction has an absolute value of $2k_\mathrm{B}T$ for a total change of $4k_\mathrm{B}T$. This positive heat flow to green spins is also shown in Fig. 5.6. Once these spins have flipped, the four red spins and black spin can flip with no change in internal energy, since they each have two neighbouring spins in the up orientation and two in the down orientation. They each experience a driving force from the changing magnetic fields which is equivalent across all of these spins, leading to broad distributions over step number for all of these spins (Fig. 5.5). In the final two steps the blue spins flip, each of which releases $4k_\mathrm{B}T$ of internal energy as the system relaxes into the all-up configuration (Fig. 5.6).

Although the details are harder to pin down, Figure 5.8 shows a mechanism for spin-inversion during a spontaneous transition path. The first two steps flip a red and green spin (which can happen in either order) and take in $4k_\mathrm{B}T$ of heat in the first step. The
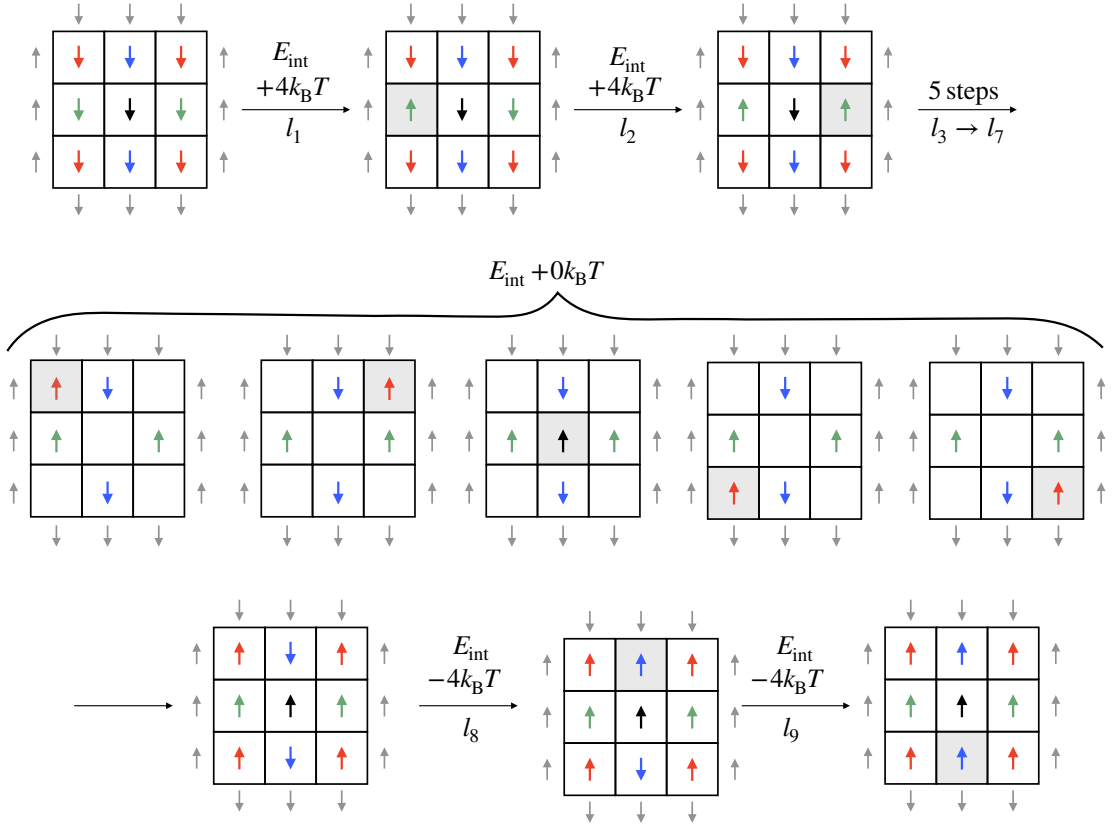
Figure 5.7: Transition mechanism during the designed protocol (MWP). The change in internal energy and step number l_α are indicated above and below the arrows. The five middle steps can happen in any order, but each has no change in internal energy.

next five steps can happen in many ways, and different spins will exchange heat with the environment depending on how the transition occurs. In general, one of these five steps will take in $4k_BT$ of heat and one will release $4k_BT$ of heat, and other steps can be accomplished with no change to the internal energy. The final two steps flip the last red and blue spins, decreasing system energy by $4k_BT$ as it reaches the all-down configuration.

5.6 Discussion

The transition mechanism for spin inversion in a 3×3 Ising model during the fully optimized 4D protocol and during a spontaneous transition path has been investigated. During the MWP, work is done on the system, which provides an energetic bias that drives the system up and over the energy barrier. In the transition-path ensemble, the system must wait for appropriate fluctuations of heat from the environment to overcome the same barrier.

The mechanisms contained in trajectories generated by each process were examined in detail, where a mechanism in this context is a sequence of spin-flip events. The MWP shows
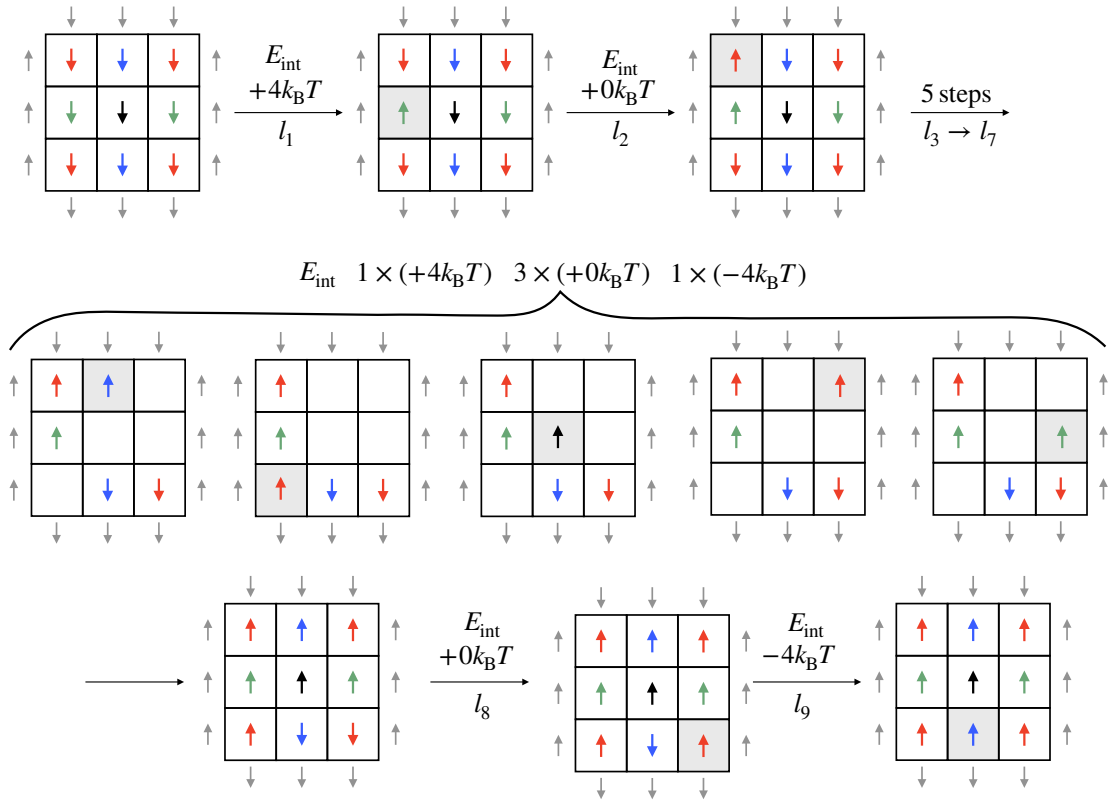
Figure 5.8: Transition mechanism during the transition-path ensemble (TPE). The change in internal energy and step number l_α are indicated above and below the arrows. The five middle steps can happen in many orders, and one step will increase the internal energy by $4k_\mathrm{B}T$ while another will decrease it by $4k_\mathrm{B}T$.

a clear transition mechanism, with both green spins taking in energy as they flip first, red and black spins flipping throughout the middle of the protocol with minimal energetic cost, and finally the blue spins flipping and releasing energy. This ordering preserves the symmetry of different spin types and is well-preserved across all trajectories.

The transition-path ensemble shows a wider diversity of transition paths that break the underlying symmetry of spin types. The first two spin flips are usually one green and one red spin, taking in $4k_\mathrm{B}T$ of energy in the first flip and $0k_\mathrm{B}T$ in the second. The next five steps have many orders of spin flips where at least one takes in $4k_\mathrm{B}T$ of heat and another releases $4k_\mathrm{B}T$. The final two steps involve flipping a red and blue spin, releasing $4k_\mathrm{B}T$ of heat in the final step.

The internal energy barrier of $8k_\mathrm{B}T$ is consistent between the two processes, but the overall manner in which the system overcomes the barrier differs. A significant factor affecting the comparison is the constrained symmetry of the 4D protocols, which prevents the controller from pushing on one spin differently than its symmetric counterpart(s). The

transition-path ensemble shows symmetry-breaking between spins in the same set, both in the order in which spin flips occur during the mechanism and the energetic cost. Relieving this constraint requires optimizing a higher-dimensional protocol, which may show the symmetry-breaking required to generate a mechanism consistent with the transition path. The minimum-work protocol in 9D space is too computationally expensive to compute using the same methods used to generate the 4D fully optimized protocol. Minimum-work protocols using the linear-response approximation in higher-dimensional spaces have been computed (e.g., a 100D protocol in Ref. [103]), but under the assumption that relaxation time is constant throughout control-parameter space so that in the optimization the friction matrix can be approximated with the force covariance matrix, significantly reducing computational cost.

More generally, what can be said about connections between minimum-work protocols and spontaneous transition mechanisms? The choice of control parameters and their ability to affect change in collective variables that are relevant to describing the reaction mechanism is an important factor in the comparison. The 4D protocol studied here does not have the ability to reproduce the dynamics in the transition-path ensemble because some of the relevant information is coarse-grained. The Ising model has $2^9 = 512$ configurations, but only 168 are unique with respect to the symmetry operations. There are also 168 unique committor values, indicating that the specific geometry of each state is relevant to parameterizing the committor. The 4D conjugate-force vector, on the other hand, only distinguishes $5 \times 3 \times 3 \times 2 = 90$ unique configurations. Thus there is loss of information about the transition mechanism in the 4D collective-variable space. Choosing control parameters that drive all collective variables that are relevant to the reaction would allow a closer comparison with the transition-path ensemble.

The underlying transition mechanisms and heat flows differ between the processes. The MWP seems to drive the system through a specific set of configurations, providing a clear and interpretable transition mechanism and analysis of energy flow to the system. In contrast, for this system the transition-path ensemble has a variety of transition mechanisms which makes it difficult to determine the details of any one dominant mechanism. If the transition-path ensemble were such that reactive trajectories tend to follow the same path (i.e., lie in a "transition tube" [92]) and control parameters are chosen to push on the relevant collective variables to describe the transition, a stronger correspondence between the processes may be observed.

Chapter 6

Conclusions

The study of reaction mechanisms is fundamental to the understanding of chemical and biological processes. Determination of physically intuitive reaction coordinates and the overall characterization of the reaction pathway from trajectories with microscopic detail is a central challenge in chemical physics today. As the ability to study larger systems with finer spatial and longer temporal resolution grows, there is greater need for theoretical frameworks that provide a basis for extraction and interpretation of reaction mechanisms in arbitrary systems. The development of transition-path theory [29, 30] over the last couple of decades provides this versatile theoretical description of a reaction through definition of an "ideal" reaction coordinate, the committor.

In Chapter 3, an information-theoretic interpretation of transition-path theory and the committor's ability to identify relevant coordinates is described. Information theory provides a powerful and versatile framework for interpreting many forms of data and the information it contains. In the context of reaction mechanisms, the committor is the ideal reaction coordinate because it preserves the mutual information (a statistical measure of dependency between random variables) between the system's microstate and the outcome (A or B) of the current trajectory segment under elimination of all other degrees of freedom. This satisfies the intuition about what a reaction coordinate is (a one-dimensional complete representation of a reactive event [23, 46, 40]) while providing a concrete understanding of the information preserved under projection of dynamics onto the reaction coordinate, in this case, information about the trajectory variables that define the transition-path ensemble and dictate its thermodynamics. Furthermore, incorporating information theory into the definition of the reaction coordinate provides a thermodynamic interpretation of the coarse-graining of system microstates that is done when low-dimensional collective variables are found to describe the reaction. In stochastic thermodynamics, a particular coarse-grained representation of system state is said to be thermodynamically consistent when the (physical) entropy production of the process is the same when calculated in the full state space and in the coarse-grained space [51, 108]. Low-dimensional collective variables that describe reaction dynamics should be thermodynamically consistent with the high-dimensional dy-

namics, but a definition of what this thermodynamic consistency entails has been lacking. The transition-path ensemble entropy production introduced in Section 3.2—that characterizes the time-asymmetry of the reaction dynamics—provides a definition of thermodynamic consistency for reaction dynamics. Additionally, a thermodynamic definition is provided for the as-yet vague concept of relevance of a collective variable to the reaction, which provides a criterion for optimization. This contribution provides a theoretical basis for the development of computational methods for extracting and interpreting information about the reaction mechanism using tools from information theory and stochastic thermodynamics.

In Chapter 4, minimum-work protocols for spin-inversion in a 2D Ising model are calculated and characterized, focusing on how multidimensional protocols interact with specific system coordinates and the underlying energy landscape to speed system relaxation and reduce excess work for the protocol. Designed multidimensional protocols take advantage of their flexibility to avoid control-parameter regions of high friction (Fig. 4.2a), decorrelate spins (Fig. 4.7) and flatten the energy landscape, thereby boosting the system population in configurations with high internal energy (Fig. 4.8), and heterogeneously input and extract work in concert with system relaxation timescales (Fig. 4.9). This drives the system along a fast-relaxing path connecting the configuration endpoints, keeping the system closer to equilibrium and reducing resistance and hence work relative to naive protocols. This work highlights the advantages of using several collective variables to drive the system since there is greater flexibility in choosing regions of configuration space to drive the system through. The designed protocols highlight some relevant features of the system's energy landscape, in particular the asymmetric boundary conditions on the system and (approximately) the size of the energy barrier the system must overcome during the reaction.

Finally, in Chapter 5 the transition-path ensemble for spin-inversion in the same 2D Ising model is studied, with methods developed to extract the reaction mechanism from trajectories and compare it to the mechanism during the 4D fully optimized protocol calculated in Chapter 4. The mechanisms differ in their finer details but the general features are preserved given the constrained symmetry of the 4D protocol. This highlights the importance of identifying relevant collective variables for the reaction: since the spontaneous transition mechanism depends on energetic and structural relationships between collective variables, control parameters that cannot affect change in the relevant collective variables may not be able to drive the system through regions of configuration space that are typically visited during a spontaneous transition. During a spontaneous transition in the 2D Ising model, spins of the same symmetry type (e.g. the two green spins) have different heat flow, and therefore the 4D control parameters that treat all spins of the same type equally cannot do work on individual spins in a way that matches their energetic requirements in the transition mechanism. However, given the constraint on the control parameters, the 4D fully optimized protocol drives the system along a path with a minimal internal energy barrier. It would be interesting to see how minimum-work protocols for this system designed in a 9D

control parameter space compare to the 4D protocol and the spontaneous transition mechanism, but clever methods for optimizing the protocol in the high-dimensional space need to be developed. Nevertheless, the results suggest that minimum-work protocols are able to uncover aspects of the spontaneous transition mechanism, potentially providing another optimization criterion for finding relevant collective variables and spontaneous transition mechanisms from nonequilibrium driving protocols.

6.1   Future directions

There are several avenues of investigation opened up by work in this thesis.

The information thermodynamics of transition paths derived in Chapter 3 opens several opportunities for future work. First, it would be helpful to better understand the meaning of the transition-path ensemble entropy production rate and how it depends on underlying details of the system, such as the magnitude of the energy barrier separating metastable states, position-dependent diffusion throughout configuration space, the presence of multiple distinct transition pathways, and the precise choice of reaction endpoints. To this end, it would be helpful to derive these results in a continuous state space to complement the Master-equation description presented in Chapter 3 and expand the applicability of the theory. Further, developing transition-path thermodynamics under coarse-graining of the state space [108] would be beneficial to more precisely understand the reaction information available from observation of a one-dimensional coordinate. Additionally, the Fisher information metric in configuration space defined in Section 3.2.3 is intriguing and may be worth further investigation using ideas from information geometry and its relationship to stochastic thermodynamics [109, 136]. With a better understanding of the intuition that can be gained about a reaction from its entropy production rate, single-molecule experiments examining transition-path dynamics could be performed and analyzed to yield insight into the reaction [3, 38].

The connection established between transition-path theory and stochastic thermodynamics suggests that relevant coordinates are identified by their time-asymmetry in the transition-path ensemble. A promising avenue for future work is the development of algorithms to find maximally informative coordinates from a sample of transition paths using established methodology for inferring entropy production in nonequilibrium systems [89, 117, 118, 119]. Additionally, incorporating stochastic-thermodynamic ideas into the interpretation of reaction mechanisms could be insightful, by using the same concepts of work, heat, and information flows used to analyze the operation of molecular machines operating in a nonequilibrium steady state [137, 138]. In particular, analyzing energy flows into system coordinates was useful in interpreting the spin-inversion mechanism in Chapter 5, and similar ideas have also been used to study conformational changes in alanine dipeptide [116, 130]. This analysis of energy flows goes beyond a one-dimensional reac-

tion free-energy profile (Fig. 1.1) and provides more detailed understanding of the energy landscape and how it depends on collective variables of interest.

Another direction is to extend the information thermodynamics for a transition between A and B to describe a network of transition paths between N distinct subsets of configurations I. This has strong ties to Markov-state modeling [25] and milestoning [28], which use transition dynamics between nearby subsets of (metastable) configurations to describe dynamics on a global scale between distant configurations of interest. In this case, the trajectory outcome and origin would be multinomial random variables, with conditional probabilities given by some multinomial generalization of the committor. Each node would have its own $I \to I$ stationary subensemble with no entropy production, as well as a set of reactive subensembles going to each other node, each with its own entropy production. This gives a total of $\binom{N}{2}$ reactive subensembles for the entire system, in principle requiring $\binom{N}{2}$ reaction coordinates that generalize the traditional binomial committor. Nevertheless, the information generated in this multinomial scenario about the multivariate trajectory outcome would be equal to the average time-irreversibility throughout all these reactive subensembles. Relevant coordinates that maximize entropy production in a set of reactive subensembles of interest could be found, gaining insight into the overall reaction network and collective variables describing each transition.

Transition-path thermodynamics were derived for a system in equilibrium, but in principle this theory could be extended to nonequilibrium steady states, where the system is connected to one or more reservoirs that maintain it out of equilibrium, breaking the detailed balance condition which is used to derive several relationships in Chapter 3. In a nonequilibrium context, the mechanism for a system transition from a reactant to a product can also be considered, and the time-asymmetry of dynamics within a transition-path ensemble and the information generated by particular degrees of freedom could be derived separately from the thermodynamic entropy production of the underlying dynamics. This may have connections with recent work by Ref. [139], which showed that milestoning (a transition-path sampling technique [28] that essentially uses a multinomial version of the trajectory outcome and origin variables [91]) is a thermodynamically consistent method of coarse-graining nonequilibrium dynamics without losing information about entropy production. There are strong similarities to the ideas discussed in Chapter 3, and a more concrete connection between these ideas is another avenue for cross-pollination of ideas from the fields of stochastic thermodynamics and transition-path theory.

Regarding minimum-work protocols, there is a need for better computational methods to design minimum-work protocols, especially in multidimensional control-parameter spaces. There has been recent progress in machine-learning techniques aimed at finding minimum-work protocols [105] which do not use the generalized friction and can therefore provide minimum-work protocols for any protocol duration. However, the generalized friction and its role in defining thermodynamic geometry in control-parameter space is physically ap-

pealing and suggests deeper connections with the geometry of the configuration space it has access to, warranting further investigation into its properties. Therefore better methods to calculate the generalized friction and its geodesics in a multidimensional control-parameter space would be helpful in expanding the applications of this approach to designing efficient protocols.

Finally, methods to optimize control parameters for efficiently driving the system would be a useful avenue for further investigation. The work in this thesis suggests that small coordinate sets identified to be relevant to a reaction could be driven with minimal energetic cost by appropriately choosing control parameters to drive them. This could plug into the immense progress in using machine learning to learn small sets of collective variables to drive a system across energy barriers [49, 120, 140, 47, 141, 142, 143, 48]. The correspondence between "good" collective variables and "good" control parameters (and particular forms of the energetic coupling between control parameters and collective variables) that do minimal work could be studied in a more diverse set of model systems to further elucidate connections between minimum-work control and spontaneous transitions.

6.2   Final remarks

Reaction mechanisms are a fundamental concept in chemistry and biology, and yet their theoretical description is challenging. Concepts such as the transition state and a one-dimensional reaction coordinate provide a starting point for development of reaction theories, but lack *a priori* definition without first having some intuition about the reaction of interest. The development of transition-path theory [29, 30] has alleviated some of the ambiguity associated with these concepts through definition of the committor, which provides the one-dimensional reaction coordinate and identifies the transition-state ensemble as the $q = 0.5$ surface. Yet challenges remain in interpreting the committor and gaining physical intuition about the mechanism in terms of small sets of system coordinates.

In this thesis, it is shown that stochastic thermodynamics provides some of the remaining pieces of the puzzle, with concepts of entropy production, information generation, and energy flows allowing identification of relevant dynamics and interpretation of the reaction mechanism. Given the rapid theoretical developments in the field in the last few decades, it should not be surprising that this powerful theoretical framework, which combines kinetics and thermodynamics by defining thermodynamic quantities for individual trajectories, is well-suited for describing the transient dynamics and energetics of a system as it undergoes a reaction. Further, the collective motions of the reaction mechanism intricately depend on interactions between collective variables. Therefore the response to dynamic manipulation of these collective variables during a protocol reflects their interactions. Protocols that minimize work take advantage of the interactions between collective variables that the control parameters affect, providing work to collective variables in accordance with the energetic

requirements of their motion and providing fast-relaxing paths through configuration space. While these concepts were studied in a particular model system, the concepts and resulting intuition are general enough to be applied to other systems. The work in this thesis provides an initial road map for further investigations of the connections between efficient control and spontaneous transitions.

Bibliography

[1] A. I. Brown and D. A. Sivak, "Theory of nonequilibrium free energy transduction by molecular machines," *Chem. Rev.*, vol. 120, no. 1, pp. 434–459, 2020.

[2] T. Yanagida, "Fluctuation as a tool of biological molecular machines," *BioSystems*, vol. 93, no. 1-2, pp. 3–7, 2008.

[3] K. Neupane, N. Q. Hoffer, and M. T. Woodside, "Measuring the Local Velocity along Transition Paths during the Folding of Single Biological Molecules," *Phys. Rev. Lett.*, vol. 121, p. 018102, 2018.

[4] M. Karplus and J. A. McCammon, "Molecular Dynamics Simulations of Biomolecules," *Nat. Struct. Biol.*, vol. 9, no. 9, pp. 646–652, 2002.

[5] C. Hartmann, R. Banisch, M. Sarich, T. Badowski, and C. Scáz§utte, "Characterization of rare events in molecular dynamics," *Entropy*, vol. 16, no. 1, pp. 350–376, 2014.

[6] G. E. Crooks, "Path-ensemble averages in systems driven far from equilibrium," *Phys. Rev. E*, vol. 61, no. 3, pp. 2361–2366, 2000.

[7] D. A. Sivak and G. E. Crooks, "Thermodynamic metrics and optimal paths," *Phys. Rev. Lett.*, vol. 108, no. 19, p. 190602, 2012.

[8] J. Liphardt, S. Dumont, S. B. Smith, I. Tinoco, and C. Bustamante, "Equilibrium information from nonequilibrium measurements in an experimental test of Jarzynski's equality," *Science*, vol. 296, pp. 1832–1835, 2002.

[9] S. Tafoya, S. J. Large, S. Liu, C. Bustamante, and D. A. Sivak, "Using a system's equilibrium behavior to reduce its energy dissipation in nonequilibrium processes," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 116, no. 13, pp. 5920–5924, 2019.

[10] M. C. Engel, D. B. Ritchie, D. A. Foster, K. S. Beach, and M. T. Woodside, "Reconstructing folding energy landscape profiles from nonequilibrium pulling curves with an inverse Weierstrass integral transform," *Phys. Rev. Lett.*, vol. 113, no. 23, pp. 1–5, 2014.

[11] M. Moradi and E. Tajkhorshid, "Computational recipe for efficient description of large-scale conformational changes in biomolecular systems," *J. Chem. Theory Comput.*, vol. 10, no. 7, pp. 2866–2880, 2014.

[12] P. Atkins, J. de Paula, and J. Keeler, *Physical Chemistry*. Oxford: Oxford University Press, 11 ed., 2018.

[13] R. Zwanzig, *Nonequilibrium Statistical Mechanics.* New York: Oxford University Press, 2001.

[14] H. A. Kramers, "Brownian motion in a field of force and the diffusion model of chemical reactions," *Physica*, vol. 7, no. 4, pp. 284–304, 1940.

[15] H. Eyring, "The activated complex in chemical reactions," *J. Chem. Phys.*, vol. 3, pp. 107–115, 1935.

[16] M. G. Evans and M. Polanyi, "Some applications of the transition state method to the calculation of reaction velocities, especially in solution," *Trans. Faraday Soc.*, vol. 31, p. 875, 1935.

[17] E. Wigner, "Calculation of the Rate of Elementary Association Reactions," *J. Chem. Phys.*, vol. 5, no. 0, pp. 720–725, 1937.

[18] J. L. Bao and D. G. Truhlar, "Variational transition state theory: Theoretical framework and recent developments," *Chem. Soc. Rev.*, vol. 46, no. 24, pp. 7548–7596, 2017.

[19] B. Peters, "Chapter 10 - transition state theory," in *Reaction Rate Theory and Rare Events Simulations* (B. Peters, ed.), pp. 227–271, Amsterdam: Elsevier, 2017.

[20] B. Peters, "Chapter 16 - kramers theory," in *Reaction Rate Theory and Rare Events Simulations* (B. Peters, ed.), pp. 435–450, Amsterdam: Elsevier, 2017.

[21] O. K. Dudko, "Decoding the mechanical fingerprints of biomolecules," *Q. Rev. Biophys.*, vol. 49, no. e3, pp. 1–14, 2016.

[22] O. K. Dudko, G. Hummer, and A. Szabo, "Theory, analysis, and interpretation of single-molecule force spectroscopy experiments," *PNAS*, vol. 105, no. 41, pp. 15755–15760, 2008.

[23] B. Peters, "Reaction coordinates and mechanistic hypothesis tests," *Annu. Rev. Phys. Chem*, vol. 67, pp. 669–690, 2016.

[24] J. Langer, "Statistical theory of the decay of metastable states," *Ann. Phys. (N. Y).*, vol. 54, no. 2, pp. 258–275, 1969.

[25] J. D. Chodera and F. Noé, "Markov state models of biomolecular conformational dynamics," *Curr. Opin. Struct. Biol.*, vol. 25, pp. 135–144, 2014.

[26] C. Dellago, P. G. Bolhuis, F. S. Csajka, and D. Chandler, "Transition path sampling and the calculation of rate constants," *J. Chem. Phys.*, vol. 108, p. 1964, 1998.

[27] 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*, vol. 53, pp. 291–318, 2002.

[28] A. K. Faradjian and R. Elber, "Computing time scales from reaction coordinates by milestoning," *J. Chem. Phys.*, vol. 120, pp. 10880–10889, 2004.

[29] W. E and E. Vanden-Eijnden, "Towards a theory of transition paths," *J. Stat. Phys.*, vol. 123, no. 3, pp. 503–523, 2006.

[30] P. Metzner, C. Schutte, and E. Vanden-Eijnden, "Transition path theory for Markov jump processes," *Multiscale Model. Simul.*, vol. 7, no. 3, pp. 1192–1219, 2009.

[31] T. S. Van Erp, D. Moroni, and P. G. Bolhuis, "A novel path sampling method for the calculation of rate constants," *J. Chem. Phys.*, vol. 118, p. 6617, 2003.

[32] R. J. Allen, P. B. Warren, and P. R. Ten Wolde, "Sampling rare switching events in biochemical networks," *Phys. Rev. Lett.*, vol. 94, p. 018104, 2005.

[33] B. Peters, "Chapter 13 - reactive flux," in *Reaction Rate Theory and Rare Events Simulations* (B. Peters, ed.), pp. 335–362, Amsterdam: Elsevier, 2017.

[34] L. Onsager, "Initial recombination of ions," *Phys. Rev.*, vol. 54, no. 8, pp. 554–557, 1938.

[35] R. Du, V. S. Pande, A. Y. Grosberg, T. Tanaka, and E. Shakhnovich, "On the transition coordinate for protein folding," *J. Chem. Phys*, vol. 108, pp. 334–350, 1998.

[36] A. Berezhkovskii and A. Szabo, "One-dimensional reaction coordinates for diffusive activated rate processes in many dimensions," *J. Chem. Phys.*, vol. 122, p. 014503, 2005.

[37] A. M. Berezhkovskii and A. Szabo, "Diffusion along the splitting/commitment probability reaction coordinate," *J. Phys. Chem. B*, vol. 117, pp. 13115–13119, 2013.

[38] A. G. T. Pyo, N. Q. Hoffer, K. Neupane, and M. T. Woodside, "Transition-path properties for folding reactions in the limit of small barriers," *J. Chem. Phys.*, vol. 149, p. 115101, 2018.

[39] W. E and E. Vanden-Eijnden, "Transition-path theory and path-finding algorithms for the study of rare events," *Annu. Rev. Phys. Chem*, vol. 61, pp. 391–420, 2010.

[40] W. Li and A. Ma, "Recent developments in methods for identifying reaction coordinates," *Mol Simul.*, vol. 40, no. 10-11, pp. 784–793, 2014.

[41] 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.*, vol. 138, p. 054106, 2013.

[42] P. V. Banushkina and S. V. Krivov, "Optimal reaction coordinates," *WIREs Comput Mol Sci*, vol. 6, pp. 748–763, 2016.

[43] Y. Khoo, J. Lu, and L. Ying, "Solving for high-dimensional committor functions using artificial neural networks," *Res. Math. Sci.*, vol. 6, p. 1, 2019.

[44] Q. Li, B. Lin, and W. Ren, "Computing committor functions for the study of rare events using deep learning," *J. Chem. Phys.*, vol. 151, p. 54112, 2019.

[45] G. M. Rotskoff, A. R. Mitchell, and E. Vanden-Eijnden, "Active importance sampling for variational objectives dominated by rare events: Consequences for optimization and generalization." arXiv:2008.06334v2.

[46] P. G. Bolhuis and C. Dellago, "Practical and conceptual path sampling issues," *Eur. Phys. J. Spec. Top.*, vol. 224, pp. 2409–2427, 2015.

[47] G. Pérez-Hernández, F. Paul, T. Giorgino, G. De Fabritiis, and F. Noé, "Identification of slow molecular order parameters for Markov model construction," *J. Chem. Phys.*, vol. 139, no. 1, 2013.

[48] C. X. Hernández, H. K. Wayment-Steele, M. M. Sultan, B. E. Husic, and V. S. Pande, "Variational encoding of complex dynamics," *Phys. Rev. E*, vol. 97, no. 6, pp. 1–11, 2018.

[49] Y. Wang and P. Tiwary, "State predictive information bottleneck," *J. Chem. Phys.*, vol. 154, p. 134111, 2021.

[50] C. Jarzynski, "Nonequilibrium equality for free energy differences," *Phys. Rev. Lett.*, vol. 78, no. 14, pp. 2690–2693, 1997.

[51] U. Seifert, "Stochastic thermodynamics, fluctuation theorems and molecular machines," *Rep. Prog. Phys*, vol. 75, p. 126001, 2012.

[52] A. C. Barato and U. Seifert, "Thermodynamic Uncertainty Relation for Biomolecular Processes," *Phys. Rev. Lett.*, vol. 114, no. 15, pp. 1–5, 2015.

[53] T. R. Gingrich, J. M. Horowitz, N. Perunov, and J. L. England, "Dissipation Bounds All Steady-State Current Fluctuations," *Phys. Rev. Lett.*, vol. 116, no. 12, pp. 1–5, 2016.

[54] A. N. Gupta, A. Vincent, K. Neupane, H. Yu, F. Wang, and M. T. Woodside, "Experimental validation of free-energy-landscape reconstruction from non-equilibrium single-molecule force spectroscopy measurements," *Nat. Phys.*, vol. 7, no. 8, pp. 631–634, 2011.

[55] C. Dellago and G. Hummer, "Computing equilibrium free energies using nonequilibrium molecular dynamics," *Entropy*, vol. 16, no. 1, pp. 41–61, 2014.

[56] J. Morfill, J. Neumann, K. Blank, U. Steinbach, E. M. Puchner, K. E. Gottschalk, and H. E. Gaub, "Force-based analysis of multidimensional energy landscapes: Application of dynamic force spectroscopy and steered molecular dynamics simulations to an antibody fragment-peptide complex," *J. Mol. Biol.*, vol. 381, no. 5, pp. 1253–1266, 2008.

[57] G. Hummer and A. Szabo, "Free energy surfaces from single-molecule force spectroscopy," *Acc. Chem. Res.*, vol. 38, pp. 504–513, 2005.

[58] Y. Suzuki and O. K. Dudko, "Biomolecules under mechanical stress: A simple mechanism of complex behavior," *J. Chem. Phys.*, vol. 134, p. 65102, 2011.

[59] C. H. Bennett, "Efficient estimation of free energy differences in Monte Carlo data," *J. Comp. Phys.*, vol. 22, no. 2, pp. 245–268, 1976.

[60] 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*, vol. 80, p. 046705, 2009.

[61] S. Kim, Y. W. Kim, P. Talkner, and J. Yi, "Comparison of free-energy estimators and their dependence on dissipated work," *Phys. Rev. E*, vol. 86, no. 4, pp. 1–10, 2012.

[62] S. Blaber and D. A. Sivak, "Skewed thermodynamic geometry and optimal free energy estimation," *J. Chem. Phys.*, vol. 153, no. 24, 2020.

[63] G. Hummer and A. Szabo, "Free energy reconstruction from nonequilibrium single-molecule pulling experiments," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 98, no. 7, pp. 3658–3661, 2001.

[64] B. Andresen and P. Salamon, "Thermodynamic geometry determines optimal temperature profile in distillation column," in *Methods Appl. Invers.* (P. Hansen, B. Jacobsen, and K. Mosegaard, eds.), pp. 15–30, Berlin: Springer, 2000.

[65] P. S. Huang, S. E. Boyken, and D. Baker, "The coming of age of de novo protein design," *Nature*, vol. 537, no. 7620, pp. 320–327, 2016.

[66] T. Schmiedl and U. Seifert, "Optimal finite-time processes in stochastic thermodynamics," *Phys. Rev. Lett.*, vol. 98, no. 10, pp. 1–4, 2007.

[67] G. E. Crooks, "Measuring thermodynamic length," *Phys. Rev. Lett.*, vol. 99, p. 100602, 2007.

[68] M. V. S. Bonança and S. Deffner, "Optimal driving of isothermal processes close to equilibrium," *J. Chem. Phys.*, vol. 140, p. 244119, 2014.

[69] S. J. Large and D. A. Sivak, "Optimal discrete control: Minimizing dissipation in discretely driven nonequilibrium systems," *J. Stat. Mech. Theory Exp.*, vol. 2019, no. 8, 2019.

[70] S. Blaber, M. D. Louwerse, and D. A. Sivak, "Steps minimize dissipation in rapidly driven stochastic systems," *Phys. Rev. E*, vol. 104, no. 2, pp. 1–6, 2021.

[71] D. Mandal and C. Jarzynski, "Analysis of slow transitions between nonequilibrium steady states," *J. Stat. Mech.*, vol. 063204, 2016.

[72] M. Berkowitz, J. D. Morgan, J. A. Mccammon, and S. H. Northrup, "Diffusion-controlled reactions: A variational formula for the optimum reaction coordinate," *J. Chem. Phys.*, vol. 79, no. 10, pp. 5563–2325, 1983.

[73] 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.*, vol. 125, no. 2, p. 024106, 2006.

[74] R. Zhao, J. Shen, and R. D. Skeel, "Maximum flux transition paths of conformational change," *J. Chem. Theory Comput.*, vol. 6, no. 8, pp. 2411–2423, 2010.

[75] M. E. Johnson and G. Hummer, "Characterization of a Dynamic String Method for the Construction of Transition Pathways in Molecular Reactions," *J. Phys. Chem. B*, vol. 116, no. 29, pp. 8573–8583, 2012.

[76] M. D. Louwerse and D. A. Sivak, "Information thermodynamics of the transition-path ensemble," *Phys. Rev. Lett*, vol. 128, p. 170602, 2022.

[77] M. D. Louwerse and D. A. Sivak, "Multidimensional minimum-work control of a 2D Ising model," *J. Chem. Phys.*, vol. 156, p. 194108, 2022.

[78] D. Morin, *Introduction to Classical Mechanics.* Cambridge University Press, 2008.

[79] L. Peliti and S. Pigolotti, *Stochastic Thermodynamics: An Introduction.* New Jersey: Princeton University Press, 2021.

[80] D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications.* San Diego: Academic Press, second ed., 2002.

[81] W. Press, S. A. Teukolksy, W. T. Vetterline, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing.* Cambridge University Press, 2007.

[82] R. J. Glauber, "Time-dependent statistics of the Ising model," *J. Math. Phys.*, vol. 4, no. 2, pp. 294–307, 1963.

[83] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* Hoboken, New Jersey: John Wiley & Sons, Inc., 2nd ed., 2006.

[84] R. Landauer, "Irreversibility and heat generation in the computing process," *IBM J. Res. Dev.*, vol. 5, pp. 183–191, 1961.

[85] J. M. Horowitz, "Multipartite information flow for multiple Maxwell demons," *J. Stat. Mech. Theory Exp.*, vol. 2015, no. 3, pp. 1–13, 2015.

[86] D. Hartich, A. Barato, and U. Seifert, "Stochastic thermodynamics of bipartite systems: transfer entropy inequalities and a Maxwell's demon interpretation," *J. Stat. Mech Theory Exp.*, vol. 2014, p. 02016, 2014.

[87] A. C. Barato, D. Hartich, and U. Seifert, "Rate of mutual information between coarsegrained non-markovian variables," *J. Stat. Phys.*, vol. 153, 2013.

[88] C. Van Den Broeck and M. Esposito, "Ensemble and trajectory thermodynamics: A brief introduction," *Physica A*, vol. 418, pp. 6–16, 2015.

[89] U. Seifert, "From stochastic thermodynamics to thermodynamic inference," *Annu. Rev. Condens. Matter Phys.*, vol. 10, pp. 171–192, 2019.

[90] E. Vanden-Eijnden, "Transition path theory," in *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation* (G. R. Bowman, V. S. Pande, and F. Noe, eds.), ch. 7, pp. 91–100, Springer, 2014.

[91] A. M. Berezhkovskii and A. Szabo, "Committors, first-passage times, fluxes, Markov states, milestones, and all that," *J. Chem. Phys*, vol. 150, p. 54106, 2019.

[92] E. Vanden-Eijnden and M. Venturoli, "Revisiting the finite temperature string method for the calculation of reaction tubes and free energies," *J. Chem. Phys.*, vol. 130, no. 19, p. 194103, 2009.

[93] W. E, W. Ren, and E. Vanden-Eijnden, "String method for the study of rare events," *Phys. Rev. B*, vol. 6, 2002.

[94] 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.*, vol. 45, pp. 188–222, 2009.

[95] D. A. Sivak and G. E. Crooks, "Thermodynamic geometry of minimum-dissipation driven barrier crossing," *Phys. Rev. E*, vol. 94, p. 052106, 2016.

[96] P. Tiwary and A. van de Walle, "A review of enhanced sampling approaches for accelerated molecular dynamics," in *Multiscale Materials Modeling for Nanomechanics* (C. R. Weinberger and G. J. Tucker, eds.), ch. 6, pp. 195–221, Springer International Publishing, 2016.

[97] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman, "The weighted histogram analysis method for free-energy calculations on biomolecules. i. the method," *J. Comput. Chem.*, vol. 13, 1992.

[98] P. R. Zulkowski and M. R. DeWeese, "Optimal control of overdamped systems," *Phys. Rev. E*, vol. 92, p. 032117, 2015.

[99] J. N. E. Lucero, A. Mehdizadeh, and D. A. Sivak, "Optimal control of rotary motors," *Phys. Rev. E*, vol. 99, p. 012119, 2019.

[100] S. Blaber and D. A. Sivak, "Optimal control of protein copy number," *Phys. Rev. E*, vol. 101, p. 22118, 2020.

[101] P. R. Zulkowski, D. A. Sivak, G. E. Crooks, and M. R. Deweese, "Geometry of thermodynamic control," *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.*, vol. 86, no. 4, pp. 1–8, 2012.

[102] G. M. Rotskoff and G. E. Crooks, "Optimal control in nonequilibrium systems: Dynamic Riemannian geometry of the Ising model," *Phys. Rev. E*, vol. 92, p. 060102, 2015.

[103] G. M. Rotskoff, G. E. Crooks, and E. Vanden-Eijnden, "A geometric approach to optimal nonequilibrium control: Minimizing dissipation in nanomagnetic spin systems," *Phys. Rev. E*, vol. 95, p. 012148, 2017.

[104] T. Gingrich, G. Rotskoff, G. Crooks, and P. Geissler, "Near-optimal protocols in complex nonequilibrium transformations," *PNAS*, vol. 113, pp. 10263–10268, 2016.

[105] M. C. Engel, J. A. Smith, and M. P. Brenner, "Optimal control of nonequilibrium systems through automatic differentiation." arXiv:2201.00098.

[106] D. M. Busiello, D. Gupta, and A. Maritan, "Entropy production in systems with unidirectional transitions," *Phys. Rev. Res.*, vol. 2, no. 2, pp. 1–15, 2020.

[107] J. M. Horowitz and M. Esposito, "Thermodynamics with continuous information flow," *Phys. Rev. X*, vol. 4, p. 031015, 2014.

[108] M. Esposito, "Stochastic thermodynamics under coarse-graining," *Phys. Rev. E*, vol. 85, p. 041125, 2012.

[109] S.-I. Amari, *Information Geometry and its Applications.* Springer, Tokyo, 1 ed., 2016.

[110] F. Nielsen, "An elementary introduction to information geometry," *Entropy*, vol. 22, p. 1100, 2020.

[111] S. Ito, "Stochastic Thermodynamic Interpretation of Information Geometry," *Phys. Rev. Lett.*, vol. 121, p. 30605, 2018.

[112] G. Ruppeiner, "Thermodynamics: A Riemannian geometric model," *Phys. Rev. A*, vol. 20, no. 4, pp. 1608–1613, 1979.

[113] S.-T. Tsai and P. Tiwary, "On the distance between A and B in molecular configuration space," *Mol. Simul.*, vol. 47, pp. 449–456, 2021.

[114] R. Chetrite, M. L. Rosinberg, T. Sagawa, and G. Tarjus, "Information thermodynamics for interacting systems without bipartite structure," *J. Stat. Mech.*, vol. 21, p. 114002, 2019.

[115] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of state calculations by fast computing machines," *J. Chem. Phys.*, vol. 21, p. 1087, 1953.

[116] W. Li and A. Ma, "Reaction mechanism and reaction coordinates from the viewpoint of energy flow," *J. Chem. Phys.*, vol. 144, p. 114103, 2016.

[117] J. Li, J. M. Horowitz, T. R. Gingrich, and N. Fakhri, "Quantifying dissipation using fluctuating currents," *Nature Communications*, vol. 10, 2019.

[118] D. J. Skinner and J. Dunkel, "Improved bounds on entropy production in living systems," *PNAS*, vol. 118, 2021.

[119] F. S. Gnesotto, G. Gradziuk, P. Ronceray, and C. P. Broedersz, "Learning the non-equilibrium dynamics of Brownian movies," *Nat. Commun.*, vol. 11, p. 5378, 2020.

[120] A. Ma and A. R. Dinner, "Automatic method for identifying reaction coordinates in complex systems," *J. Phys. Chem. B*, vol. 109, pp. 6769–6779, 2005.

[121] Y. Wang, J. M. L. Ribeiro, and P. Tiwary, "Machine learning approaches for analyzing and enhancing molecular dynamics simulations," *Curr. Opin. Struct. Biol.*, vol. 61, pp. 139–145, 2020.

[122] Y. I. Yang, Q. Shao, J. Zhang, L. Yang, and Y. Q. Gao, "Enhanced sampling in molecular dynamics," *J. Chem. Phys.*, vol. 151, no. 7, 2019.

[123] L. Maragliano and E. Vanden-Eijnden, "A temperature accelerated method for sampling free energy and determining reaction pathways in rare events simulations," *Chem. Phys. Lett.*, vol. 426, no. 1-3, pp. 168–175, 2006.

[124] J. Pfaendtner and M. Bonomi, "Efficient sampling of high-dimensional free-energy landscapes with parallel bias metadynamics," *J. Chem. Theory Comput.*, vol. 11, no. 11, pp. 5062–5067, 2015.

[125] W. Jiang, Y. Luo, L. Maragliano, and B. Roux, "Calculation of free energy landscape in multi-dimensions with Hamiltonian-exchange umbrella sampling on petascale supercomputer," *J. Chem. Theory Comput.*, vol. 8, no. 11, pp. 4672–4680, 2012.

[126] T. Zhao, H. Fu, T. Lelièvre, X. Shao, C. Chipot, and W. Cai, "The extended generalized adaptive biasing force algorithm for multidimensional free-energy calculations," *J. Chem. Theory Comput.*, vol. 13, no. 4, pp. 1566–1576, 2017.

[127] C. Chipot and T. Lelièvre, "Enhanced sampling of multidimensional free-energy landscapes using adaptive biasing forces," *SIAM J. Appl. Math*, vol. 81, no. 5, pp. 1673–1695, 2011.

[128] M. Plischke and B. Bergersen, *Equilibrium Statistical Physics.* Toh Tuck Link, Singapore: World Scientific Publishing Co., 3nd ed. ed., 2006.

[129] J. Jiang, D. A. Sivak, and M. Thomson, "Active learning of spin network models." arXiv:1903.10474.

[130] S. Wu, H. Li, and A. Ma, "A Rigorous Method for Identifying a One-Dimensional Reaction Coordinate in Complex Molecules," *J. Chem. Theory Comput.*, 2022.

[131] N. Q. Hoffer, K. Neupane, A. G. T. Pyo, and M. T. Woodside, "Measuring the average shape of transition paths during the folding of a single biological molecule," *Proc. Natl. Acad. Sci.*, vol. 116, no. 17, pp. 8125–8130, 2019.

[132] W. K. Kim and R. R. Netz, "The mean shape of transition and first-passage paths," *J. Chem. Phys.*, vol. 143, no. 22, 2015.

[133] D. E. Makarov, "Shapes of dominant transition paths from single-molecule force spectroscopy," *J. Chem. Phys.*, vol. 143, no. 19, pp. 2–6, 2015.

[134] P. Cossio, G. Hummer, and A. Szabo, "Transition paths in single-molecule force spectroscopy," *J. Chem. Phys.*, vol. 148, p. 123309, 2018.

[135] S. Park, M. K. Sener, D. Lu, and K. Schulten, "Reaction paths based on mean first-passage times," *J. Chem. Phys.*, vol. 119, no. 10, p. 1313, 2003.

[136] S. Ito, M. Oizumi, and S.-I. Amari, "Unified framework for the entropy production and the stochastic interaction based on information geometry," *Phys. Rev. Res.*, vol. 2, p. 33048, 2020.

[137] E. Lathouwers, J. N. Lucero, and D. A. Sivak, "Nonequilibrium Energy Transduction in Stochastic Strongly Coupled Rotary Motors," *J. Phys. Chem. Lett.*, vol. 11, no. 13, pp. 5273–5278, 2020.

[138] M. P. Leighton and D. A. Sivak, "Performance scaling and trade-offs for collective motor-driven transport," *New J. Phys.*, vol. 24, no. 1, 2022.

[139] D. Hartich and A. Godec, "Emergent memory and kinetic hysteresis in strongly driven networks," *Phys. Rev. X*, vol. 11, p. 041047, 2021.

[140] F. Noé and F. Nüske, "A variational approach to modeling slow processes in stochastic dynamical systems," *Multiscale Model. Simul.*, vol. 11, no. 2, pp. 635–655, 2013.

[141] A. Mardt, L. Pasquali, H. Wu, and F. Noé, "VAMPnets for deep learning of molecular kinetics," *Nat. Commun.*, vol. 9, no. 1, pp. 1–11, 2018.

[142] P. Tiwary and B. J. Berne, "Spectral gap optimization of order parameters for sampling complex molecular systems," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 113, no. 11, 2016.

[143] Y. Wang, J. Marcelo, L. Ribeiro, and P. Tiwary, "Past-future information bottleneck for sampling molecular reaction coordinate simultaneously with thermodynamics and kinetics," *Nat. Commun.*, vol. 10, p. 3573, 2019.

[144] M. D. Louwerse, "Thesis code for 2d ising model." `https://github.com/mdlouwerse/MDL_thesis_code`, 2022.

[145] P. A. Walker, "Quadcubic interpolation: a four-dimensional spline method." arXiv:1904.09869.

[146] P. A. Walker, U. Krohn, and D. Carty, "ARBTools: A tricubic spline interpolator for three-dimensional scalar or vector fields," *J. Open Res. Softw.*, vol. 7, no. 1, p. 12, 2019.

Appendix A

Code

Select code supporting the work in this thesis is publicly available in a Github reposi-
tory [144]. Further code and data are available upon reasonable request.

Appendix B

Computational details for Ising model

B.1 Calculation of generalized friction matrix

The friction matrix for control parameter $\boldsymbol{\lambda}$ was estimated from time-correlation functions of pairs of conjugate forces (2.62). Single-spin-flip Glauber dynamics [82] are used to simulate four equilibrium trajectories, each of length $10^7 \mathrm{d}t$, where $\mathrm{d}t$ is a constant simulation time interval after which a random spin flip is attempted. Time-correlation functions were calculated between pairs of conjugate-force trajectories using fast Fourier transforms [81] and integrated to lag time $2.5 \times 10^4 \,\mathrm{d}t$. The four independent estimates of friction at each control-parameter value were then averaged.

The 4×4 friction matrix is calculated at grid points with discrete spacing $\mathrm{d}\lambda_i = 0.2$ within the control-parameter space bounded by: $\lambda_{\mathrm{red}} \in [-2.1, 2.1]$, $\lambda_{\mathrm{blue}} \in [-3.0, 1.0]$, $\lambda_{\mathrm{green}} \in [-1.0, 3.0]$, $\lambda_{\mathrm{black}} \in [-2.1, 2.1]$. Since this control-parameter space has spatial symmetry, statistical power is doubled by averaging friction values for symmetrically equivalent control-parameter values: $(\lambda_{\mathrm{red}}, \lambda_{\mathrm{blue}}, \lambda_{\mathrm{green}}, \lambda_{\mathrm{black}}) \rightarrow (-\lambda_{\mathrm{red}}, -\lambda_{\mathrm{green}}, -\lambda_{\mathrm{blue}}, -\lambda_{\mathrm{black}})$.

B.2 Design of control protocols

Here, the methods used to design minimum-work protocols are described in more detail.

The time-optimized protocols are effectively one-dimensional with a single field applied equally to all spins; these protocols have optimal velocity inversely proportional to the square root of the 1D friction coefficient $\zeta^{\mathrm{1D}}(\boldsymbol{\lambda}) \equiv \sum_{i,j} \zeta_{ij}(\boldsymbol{\lambda})$.

The fully optimized protocols are calculated using the string method (Ref. [103] and Section 2.3.5).

Calculating both time- and fully optimized protocols requires knowledge of the 4×4 friction matrix in relevant regions of control-parameter space. Each component of the friction

matrix and its gradient (which is required for fully optimized protocols) with respect to each control parameter are interpolated using four-dimensional cubic splines [145, 146], which solve piecewise cubic polynomials over the data array to ensure smooth first and second derivatives. Since components of the friction matrix vary several orders of magnitude throughout control-parameter space, cubic splines are fit to the natural logarithm of each matrix component and re-exponentiated interpolated values from this surface to recover the friction coefficient. The natural logarithm of the friction is defined for this Ising model since each friction matrix component is positive. In low-dissipation regions at relatively large field values (top and left edges of the 2D control-parameter space in Fig. 4.2), noise causes the friction estimate at some control-parameter values to become negative; these were set to a constant value of $10^{-5}\,\tau_{\mathrm{rel}}/k_{\mathrm{B}}T$.

B.2.1 Time-optimized protocols

To calculate the time-optimized protocols (Fig. 4.3, middle), the friction is interpolated at 121 evenly spaced points along the fully naive protocol, then these points are reparameterized to be evenly spaced in terms of the 1D friction $\zeta^{1\mathrm{D}}(\boldsymbol{\lambda}) \equiv \sum_{i,j} \zeta_{ij}(\boldsymbol{\lambda})$. A reparameterization scheme commonly used with the string method was used to keep string points evenly spaced [73], with the modification that distances between discrete points are measured using the friction metric [102], to ensure constant excess power along this protocol (Fig. 4.11).

The string method divides a protocol of duration Δt into M discrete control-parameter points, with fixed time $\delta t = \Delta t/(M-1)$ between each string point and boundary conditions $\lambda(0) = \lambda_{\mathrm{initial}}$ and $\lambda(\Delta t) = \lambda_{\mathrm{final}}$. Since the minimum-work protocols are independent of duration (simply differing by a rescaling of δt), $\Delta t = 1$ without loss of generality. Let $\mathcal{L}^2(m)$ denote the length of the string up to discrete point $m \in [0, M-1]$. The generalized friction is used to define the distance between adjacent string points

$$\mathrm{d}\ell^2_{\boldsymbol{\lambda}(m),\boldsymbol{\lambda}(m-1)} = \sum_{ij} [\boldsymbol{\lambda}_i(m) - \boldsymbol{\lambda}_i(m-1)] \frac{[\zeta_{ij}(\boldsymbol{\lambda}(m)) + \zeta_{ij}(\boldsymbol{\lambda}(m-1))]}{2} [\boldsymbol{\lambda}_j(m) - \boldsymbol{\lambda}_j(m-1)],$$
(B.1)

so that the length of the string up to point m is

$$\mathcal{L}^2(m) = \sum_{m'=1}^{m} \mathrm{d}\ell^2_{\boldsymbol{\lambda}(m'),\boldsymbol{\lambda}(m'-1)},$$
(B.2)

and estimates (up to a factor of protocol duration) the excess power on the system up to time $m\delta t$ in the protocol, with $\mathcal{L}^2(M-1)$ estimating the squared metric distance (and excess power) between protocol endpoints.

The string points $\boldsymbol{\lambda}^*(m)$ are then updated by linear interpolation of the current string so that the mth point is situated in proportion to the total distance $m/(M-1)$. Define $s(m) = [m/(M-1)]\mathcal{L}^2(M-1)$ as the distance along the string where point m should sit, then find the string point k in the original string such that $\mathcal{L}^2(k-1) < s(m) \leq \mathcal{L}^2(k)$. The

reparameterized point $\boldsymbol{\lambda}^*(m)$ is then linearly interpolated between $\boldsymbol{\lambda}(k-1)$ and $\boldsymbol{\lambda}(k)$ by

$$\boldsymbol{\lambda}^*(m) = \boldsymbol{\lambda}(k-1) + \frac{s(m) - \mathcal{L}^2(k-1)}{\mathrm{d}\ell^2_{\boldsymbol{\lambda}(k),\boldsymbol{\lambda}(k-1)}} \left[\boldsymbol{\lambda}(k) - \boldsymbol{\lambda}(k-1)\right] . \tag{B.3}$$

The new string has points that are evenly spaced in terms of square distance (B.1), and therefore has constant excess power.

B.2.2 Fully optimized protocols

To calculate the fully optimized protocols (Fig. 4.3, bottom), the multidimensional Euler-Lagrange equation [78] for excess work (2.64) is solved numerically using the string method, as developed in Ref. [103]. This involves numerically solving the Euler-Lagrange equation for the protocol minimizing the excess work, by dividing the protocol into a discrete set of points along a "string" that runs between fixed protocol endpoints. The string is initialized as the naive protocol, where each point is evenly separated in Euclidean space. The string is updated from the nth to $(n+1)$th iteration by solving the set of linear equations for each component λ_α at each discrete point m along the string:

$$h_\alpha^{n+1}(m) - h_\alpha^n(m) = \Delta r \Bigg(\mathrm{D}^2 \lambda_\alpha^{n+1}(m)$$

$$+ \sum_{ijk} [\zeta^{-1}]_{\alpha k}^n(m) \, \mathrm{D}h_i^n(m) \, \mathrm{D}h_j^n(m) \left[\partial_i \zeta_{kj}(m) - \frac{1}{2}\partial_k \zeta_{ij}(m) \right] \Bigg) , \tag{B.4}$$

where

$$\mathrm{D}\lambda_\alpha(m) \equiv \frac{\lambda_\alpha(m+1) - \lambda_\alpha(m-1)}{2\delta t} \tag{B.5a}$$

$$\mathrm{D}^2\lambda_\alpha(m) \equiv \frac{\lambda_\alpha(m+1) + \lambda_\alpha(m-1) - 2\lambda_\alpha(m)}{\delta t^2} \tag{B.5b}$$

are finite-difference estimators of time derivatives for the string and δt is the difference in scaled protocol time between adjacent string points. Δr is a parameter controlling the size of the string update; $\Delta r = 10^{-4}$ for the 4D protocol and $\Delta r = 10^{-5}$ for the 2D protocol allowed convergence of each string to a constant-excess-power protocol (required for the time-optimized and fully optimized protocols). Unlike Ref. [103], the string was not reparameterized to keep points equally spaced; omitting this step allows simultaneously optimization of the spatial and temporal aspects of the fully optimized protocols.

The algorithm requires evaluation of the friction matrix and its spatial gradients with respect to control parameters at arbitrary control-parameter values, which are obtained from the spline fit. Ref. [103] assumes the integral relaxation time is constant and therefore the friction matrix can be replaced by the force covariance matrix; its derivative is simply the third cumulant of conjugate forces, $\langle \delta X_i \delta X_j \delta X_k \rangle$. This assumption greatly simplifies the calculation, but is not a good approximation for this system, as seen in Fig. 4.2 where the variation in relaxation time throughout the 2D control-parameter space is on the same order of magnitude as the variation in force covariance, and therefore variation of each

element contributes significantly to the variation of the total friction. Optimization was carried out using 121 discrete string points while the symmetry of the protocol was constrained, $(\lambda_{\text{red}}(m), \lambda_{\text{blue}}(m), \lambda_{\text{green}}(m), \lambda_{\text{black}}(m)) = (-\lambda_{\text{red}}(M-1-m), -\lambda_{\text{green}}(M-1-m), -\lambda_{\text{blue}}(M-1-m), -\lambda_{\text{black}}(M-1-m))$. The 4D fully optimized protocol was passed through a Gaussian filter [81] with $\sigma = 0.03\Delta t$ to smooth noisy fields near protocol endpoints resulting from noisy friction estimates in low-dissipation regions. The smoothing affects the corresponding linear-response approximation to excess power in Fig. 4.11, which visibly dips for green and blue fields close to the start and end of the protocol, respectively. However, the total excess power throughout the vast majority of the protocol is constant as expected.

B.3   Protocol simulation and analysis

To simulate the control protocol, a trajectory is initialized from the equilibrium distribution at $\boldsymbol{\lambda}(t=0)$, then its spin-flip dynamics are calculated for duration Δt using Glauber transition rates 2.6, while changing the control parameters according to the protocol. $N = 5000$ repetitions of each driving protocol were used to compute nonequilibrium averages of power, work, spin magnetization, total energy, entropy, and free energy. Quasistatic properties are calculated from Boltzmann-weighted equilibrium distributions along the protocol.