ABSTRACT

Title of dissertation: STOCHASTIC PROCESSES IN PHYSICS:
DETERMINISTIC ORIGINS AND CONTROL

Jeffery Demers, Doctor of Philosophy, 2016

Dissertation directed by: Professor Christopher Jarzynski
Department of Physics,
Department of Chemistry and Biochemistry,
and Institute for Physical Sciences and Technology

Stochastic processes are ubiquitous in the physical sciences and engineering. While often used to model imperfections and experimental uncertainties in the macroscopic world, stochastic processes can attain deeper physical significance when used to model the seemingly random and chaotic nature of the underlying microscopic world. Nowhere more prevalent is this notion than in the field of stochastic thermodynamics - a modern systematic framework used describe mesoscale systems in strongly fluctuating thermal environments which has revolutionized our understanding of, for example, molecular motors, DNA replication, far-from equilibrium systems, and the laws of macroscopic thermodynamics as they apply to the mesoscopic world. With progress, however, come further challenges and deeper questions, most notably in the thermodynamics of information processing and feedback control. Here it is becoming increasingly apparent that, due to divergences and subtleties of interpretation, the deterministic foundations of the stochastic processes themselves must be explored and understood.

This thesis presents a survey of stochastic processes in physical systems, the deterministic origins of their emergence, and the subtleties associated with controlling them. First, we study time-dependent billiards in the quivering limit - a limit where a billiard system is indistinguishable from a stochastic system, and where the simplified stochastic system

allows us to view issues associated with deterministic time-dependent billiards in a new light and address some long-standing problems. Then, we embark on an exploration of the deterministic microscopic Hamiltonian foundations of non-equilibrium thermodynamics, and we find that important results from mesoscopic stochastic thermodynamics have simple microscopic origins which would not be apparent without the benefit of both the micro and meso perspectives. Finally, we study the problem of stabilizing a stochastic Brownian particle with feedback control, and we find that in order to avoid paradoxes involving the first law of thermodynamics, we need a model for the fine details of the thermal driving noise. The underlying theme of this thesis is the argument that the deterministic microscopic perspective and stochastic mesoscopic perspective are both important and useful, and when used together, we can more deeply and satisfyingly understand the physics occurring over either scale.

STOCHASTIC PROCESSES IN PHYSICS:
DETERMINISTIC ORIGINS AND CONTROL

by

Jeffery Demers

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy
2016

Advisory Committee:
Professor Christopher Jarzysnki, Chair/Advisor
Professor P.S. Krishnaprasad
Professor Charles David Levermore
Professor Dionisios Margetis
Professor Edward Ott

Acknowledgments

First and foremost, I need to thank my academic advisor Chris Jarzynski for his uncountably infinite supplies of patience and tolerance. Chris is an incredibly insightful scientist, a gifted teacher, and an all around wonderful human being. From him, I've learned the benefits of clear presentation and the satisfaction that comes with it, and I will forever strive to reach the standard he has set for formulating clear, precise, and eloquent ideas.

Next, I must acknowledge the professors and postdocs at the University of Maryland from which I've enjoyed many helpful discussions. Even the brief informal chats at lunches and in the hallways have been immeasurably helpful. Sebastain, Marcus, Yigit, Oren, and Krisha are the names that come to mind, but I'm sure there are many more that I'm forgetting. In that regard, thanks to current, former, and visiting students in Chris's group: Ayoti, Andrew, Zhiyue, and Victor - it's been a long, strange, and enjoyable experience.

A thank-you goes out to all of my friends and family. Extra thanks to my parents, especially for seeing us through the teenage years while minimizing the long-term psychological damage to all parties involved. Extra special thanks to Kerri for being a friend, for sharing in the joy of the absurd, and for always being ready to honestly yet constructively and lovingly let me know when I haven't been at my best.

Finally, a sincere heartfelt thanks goes out to Drew Baden, Tom Cohen, and Nick Hammer of (or formally of) the physics department at the University of Maryland, College Park. In September of 2011, just a month or two after moving to College Park and beginning my graduate studies, my residence caught fire and burnt down while my dog and I were out, the fire killing my cat and destroying the last vestiges of a former life. The morning after

the fire, I walked into the graduate office in the physics department, lost, dirty, defeated, and looking to quit the graduate program in order to spend some time sorting out life. Fortunately, Drew, Tom, and Nick refused to acknowledge this as on option - they focused me, encouraged me to keep going, and purchased for me the computer on which I type these words. I would not be here today, scrambling to finish my thesis and preparing to graduate, were it not for their support and kindness. It's a debt which cannot be repaid, and I can only hope to one day pay it forward to others in need.

Contents

List of Figures

List of Abbreviations

L (Chap 2)	1-D billiard average length
u_c	Billiard wall characteristic oscillation speed
$\Psi.$	Wall phase
a (Chap 2)	Billiard wall oscillation amplitude
τ (Chap 2)	Billiard wall oscillation period
m	Point particle mass
v_b	Incoming particle speed at b^{th} collision
v (Chap 2)	Point particle speed
t_b	Time of bth collision
$P(u\|v)$	Probability density for particle with speed v to collide with 1-D billiard wall quivering with velocity u
$P(u\|v, \mathbf{q}, \theta)$	Probability density for particle with speed v to collide with billiard wall quivering with normal velocity u at location $\mathbf{q}$ and collision angle θ
ΔE_b	Change in particle energy during b^{th} bounce
ε	Small parameter
$\{...\}_b$	Conditional expectation given previous $b-1$ collisions
μ_b	Conditional expectation of ΔE_b given previous $b-1$ collisions
$M_n(\mathbf{q}_b)$	n^{th} moment of quivering wall velocity at location $\mathbf{q}_b$
σ_b^2	Conditional variance of ΔE_b given previous $b-1$ collisions
$\text{Cov}_{b,b+1}$	Conditional covariance of ΔE between adjacent bounces
$L_b\|_F$	Distance between the bth and $b+1^{th}$ collision locations in the frozen dynamics
$\langle...\rangle$ (Chap 2)	Average over ensemble of independent particles
$\eta(E,t)$	Time-dependent energy distribution
$g_1(E,t)$ (Chap 2)	Drift coefficient
$g_2(E,t)$ (Chap 2)	Diffusion coefficient
$\overline{u^2}(t';B)$	Coarse grained ensemble averaged squared wall speed at collisions
$\bar{l}(t';B)$	Coarse grained ensemble averaged free flight distance
$\Sigma(E)$ (Chap 2)	Microcanonical partition function of a single particle with energy E in frozen billiard

Ω_d (Chap 2)	d-dimensional solid angle
s (Chap 2)	Rescaled time
J_{d-1}	Ordinary Bessel function of order $d-1$
I_{d-1}	Modified Bessel function of order $d-1$
$\Gamma(d)$	Gamma function
$\mathbf{Q}$	Slow system position
$\mathbf{P}$	Slow system momentum
$\mathbf{X}$	$(\mathbf{Q}, \mathbf{P})$
M	Diagonal matrix of slow system inertias
$\mathbf{q}^\alpha$ (Chaps. 3 and 4)	Bath $\alpha's$ position
$\mathbf{p}^\alpha$	Bath α's momentum
$\mathbf{x}^\alpha$	$(\mathbf{q}^\alpha, \mathbf{p}^\alpha)$
$\mathbb{X}$	Joint bath phase space coordinates
^{T}	Matrix transpose
E^α	Bath α's energy
$\mathbf{E}$	Vector of bath energies
$\mathcal{H}$	Hamiltonian of universe
H	Hamiltonian of system of interest
V	Potential energy
$\boldsymbol{\lambda}$	Control parameter
h^α	Bath Hamiltonian
Σ^α	Microcanonical partition function of bath α
Ω^α	Ergodic adiabatic invariant of bath α
$\mathbf{u}^\alpha$	Adiabatic force
$\hat{\mathbf{D}}$	Constrained momentum gradient
ϕ	Microscopic phase space distribution
W_t or $W(\mathbf{X}, \mathbf{E}, t)$	Mesoscopic time-dependent distribution
L^α	Integrated correlation matrix
$L^{\mathrm{sy}\alpha}$	Symmetric part of integrated correlation matrix
$L^{\mathrm{ay}\alpha}$	Anti-symmetric part of integrated correlation matrix
W_s	Mesoscopic stationary distribution
P_s	Mesoscopic stationary distribution with coarse grained bath states
E^U	Total universe energy as a function
$\mathcal{E}$	Value of total universe energy
$\mathbb{D}$	Relative entropy
Σ^U	Microcanonical partition function of universe
Σ^B	Joint bath microcanonical partition function
s^α	Bath α microcanonical entropy
S^B	Joint bath microcanonical entropy
S^U	Microcanonical entropy of universe
$\mathcal{F}$	Free Energy
T^α	Microcaonincal temperature of bath α
T^B	Coarse grained joint bath microcanonical temperature

T^U	Microcanonical temperature of universe
$\boldsymbol{\mathcal{W}}^\alpha$	Vector of standard Wiener processes
$\boldsymbol{\xi}^\alpha$	Vector of Gaussian white noises
$\mathbf{f}$	Deterministic mesoscopic force
$\mathcal{A}$	Stochastic action
$\mathcal{L}$	Stochastic Lagrangian
$\mathrm{d}q$	heat increment flowing from surroundings
$\mathrm{d}w$	work increment
$P_{\boldsymbol{\lambda}(\tau)}(\mathbf{X}_f, \mathcal{E}_f, t_f \,;\, \mathbf{X}_0, \mathcal{E}_0, t_0)$	Endpoint transition probability
$P_{\boldsymbol{\lambda}(\tau)}[\mathbf{X}(\tau), \mathcal{E}(\tau)]$	Path transition probability
$\dagger$	Conjugation operator
$x(t)$ (Chap 5)	Overdamped particle position
k	Spring constant
T (Chap 5)	Temperature of environment
b (Chap 5)	Damping constant
τ (Chap 5)	b/k Natural overdamped diffusion times-scale
ξ	Gaussian white noise
η (Chap 5)	Gaussian white noise
σ	Measurement noise strength
$y(t)$	Noisy observation
$\hat{x}(t)$	Estimated position
$\epsilon(t)$	Estimate error
$\mathfrak{E}^2(t)$	Mean square estimate error
$K(t)$	Kalman gain
τ_ϵ	Natural error time-scale
$z(t)$	Ornstein Uhlenbeck process
τ_z	Ornstein Uhlenbeck correlation time
F_{fb}^o	Perfect observation feedback force
F_{fb}	Noisy observation feedback force

Chapter 1

Introduction

1.1 Background

Stochastic processes have been prevalent in the physical sciences for over a century now, with origins dating back to Einstein's seminal 1905 and work on Brownian motion [1]. It was in these papers that Einstein employed the then somewhat controversial atomic kinetic theory of matter to resolve an obscure eighty year-old mystery posed by botanist Robert Brown's observations of pollen grains from the plant *Clarkia Pulchella* suspended in water [2]. Brown noted that the microscopic grains appeared to engage in a continuously erratic, seemingly random motion which could not be attributed to the presence of any kind of "vital life force" [3], and Einstein's insight was to attribute this movement to the collective force exerted by individual colliding water molecules (to which the grains were giants by comparison) which were in continual thermal motion, yielding a diffusion equation for the motion of the pollen grains. Shortly thereafter, Smoluchowski, Langevin, Fokker, and Planck, working independently and building on each others results, systematically incorporated probability theory into Einstein's insights, in the process forming a generalized diffusion equation (the Fokker-Planck equation) and a rudimentary notion of stochastic calculus (the Langevin equation) [4, 5].

Following these initial beginnings, in 1923, Wiener formalized the notion of Brownian

motion with mathematical rigor, defining the paradigmatic continuous time stochastic process now known as the Wiener process [6], which, after Kolmogorov's axiomatization of probability theory in 1933 [7], facilitated the birth of Ito's stochastic calculus [8], a now indispensable tool in fields as diverse as mathematical finance, quantitative ecology, and control engineering [9]. In the physical sciences, both continuous and discrete time and/or space stochastic processes are found wherever there is a need model the behavior of a few relevant degrees of freedom (i.e. those accessible to direct observation and experimentation) in the presence of large numbers of complicated irrelevant degrees of degrees of freedom: stellar dynamics, chemical kinetics, plasma physics, biophysics, the molecular machinery of life, and the thermodynamics of small systems and information processing are just a few examples [5, 10, 11, 12, 13]. The mathematics of stochastic processes and their important applications are deeply rooted in the need to understand the world of the large and slow in terms of the underlying world of the small and fast, and in this thesis, we argue for the reverse viewpoint. Stochastic processes in physics are equally useful and important tools for understanding the world of the small and fast in terms of the world of the large and slow, and when working together, both views compliment each other and can yield new insights into old problems.

Throughout this thesis, we will frequently refer to mesoscopic and microscopic scales. In making this somewhat blurred distinction, we will follow the conventions of Altaner [14]. Microscopic scales will refer to distance and time scales associated with a few individual degrees of freedom whose evolution is reversible and deterministic, where the environmental degrees of freedom, if present, are explicitly considered. A typical example of a microscale system is an isolated collection of a few charged classical particles interacting through electromagnetic potentials. Mesoscopic scales correspond to larger, possibly collective degrees of freedom which evolve stochastically and irreversibly over distance and time-scales which are much larger than microscopic scales, with environmental effects accounted for by the stochastic driving noise. Mesoscale systems typically live in highly dissipative environments where the energy associated with thermal fluctuations is comparable to the energy scales associated with the systems themselves. A prototypical example of a mesoscale system is the molecular kinesin motor, which converts chemical energy from the surrounding environment

into mechanical energy in order to stochastically transport cellular cargo along microtubules found inside the cells of living creatures [12]. Both the microscale and mesoscale stand in stark contrast to the macroscale of everyday experience, where the energy associated with thermal fluctuations is irrelevant, and systems evolve deterministically but irreversibly due to the second law of thermodynamics. The macroscale is the world of such familiar phenomena as the heating of a tire when pumped with air, or the shattering of a ceramic mug when dropped by a clumsy graduate student.

In this thesis, it will be important to recognize that the same physical system can be effectively described at both the microscale and mesoscale, with each description having a domain of applicability and usefulness. For example, a Brownian particle described on the mesoscale evolves diffusively, with the force of the surrounding water molecules modelled by a stochastic differential equation. This description is expected to be accurate over time scales much, much greater than the average time between individual collisions with the water molecules, and accurate down to distance scales whose square is given by the product of a characteristic diffusion time scale and a diffusion constant. The microscopic description of a Brownian particle is deterministic and explicitly models each individual collision with the water molecules using Hamilton's equations. The accuracy of this description is limited only by quantum effects at very fine distance and time-scales. A recurring theme throughout our work will be the emergence of partial differential equations (diffusion equations in particular) at the mesoscale from the deterministic ordinary differential equations which govern the microscale.

1.2 Motivation and outline

Our motivation for the framing of this thesis stems from recent advances in the thermodynamics of control and information processing made possible by stochastic thermodynamics. Stochastic thermodynamics is a systematic framework which describes the effective mesoscale evolution of small systems which exist is dissipative fluctuating thermal environments. In this framework, mesoscale systems evolve along individual stochastic trajectories, and the stochastic driving noise is modelled in terms of the bulk physical properties of the

surrounding medium. Stochastic thermodynamics endows macroscopic thermodynamic notions such has work, heat, entropy production, and the first law with physically meaningful definitions at the level of individual mesoscale trajectories, and provides a setting to derive and unify the fluctuation theorems, a collection of important results from non-equilibrium thermodynamics which are related to the second law and are distinguished by their validity for systems driven arbitrarily far from equilibrium [13, 15, 16, 17, 18, 19, 20]. Like the second law, fluctuation theorems place constraints on work and entropy production when controlling and manipulating physical systems in thermal environments. In addition to revolutionizing our understanding of and systems driven far from equilibrium [13], stochastic thermodynamics and fluctuation theorems have together helped to elucidate the precarious relationship between information entropy and thermodynamic entropy, a relationship first exposed by the introduction of Maxwell's demon [21]. A Maxwell's demon functions as a feedback controller, using observations of a thermodynamic system to enact apparent violations of the second law [21, 22, 23, 24]. In recent times, Maxwell's demon has been effectively exorcised; it is now well-understood that these apparent second law violations can be quantified in terms of various measures of mutual information flow between the feedback controller and thermodynamic system, and that if the feedback controller is itself treated as a physical system, the thermodynamic cost of information erasure compensates for the demon's second law violations [13, 23, 24, 25, 26, 27, 28, 29, 30].

Although the results exorcising Maxwell's demon are well-established and have been experimentally verified [31], their still exist some subtleties of interpretation. Feedback controllers are often modelled assuming the existence of some observation error, and in many cases, in the limit of perfect observations, the mutual information quantities relevant to Maxwell's demon can be seen to diverge to infinity [26, 29, 32]. To the best of our knowledge, these divergences are either ignored entirely, or are addressed with the explanation that it takes an infinite amount of information to precisely specify the trajectory of a stochastically evolving system [26]. We find this explanation physically unsatisfying, and we hypothesize that these divergences actually signal a breakdown of the validity of stochastic thermodynamics. In short, if a Maxwell's demon observes the trajectory of a thermodynamic system with infinite precision, then it will have acquired detailed *micro-*

scopic information regarding the system's evolution, so a stochastic *mesoscopic* description is not a priori valid.

Some hints supporting the veracity of our hypothesis already exist in the literature. The issue seems to be related to the fact that, as shown by Gaspard and Wang [33], when the sample paths of certain stochastic processes (such as the white noise and Poisson processes commonly used in stochastic thermodynamics) are observed with infinite precision, information is generated at an infinitely fast rate. The underlying physical world, however, is not stochastic - classical physics is a deterministic theory, so seemingly random processes must be deterministic in origin and may actually be chaotic. When the sample path of a chaotic process is observed in the limit of infinite precision, the information rate does not diverge, but instead saturates to a value known as the Kolmogorov-Sinai entropy rate [33]. This entropy rate is determined by the process's underlying dynamical properties, and for Hamiltonian systems, the entropy rate is related to the classical action (in the sense of Lagrange and Hamilton) [34]. Thus, there appear to be deep connections between feedback control in thermodynamics and the underlying Hamiltonian nature of the universe waiting to be discovered, but in order to make such connections, we need to understand thermodynamic systems at both the microscopic and mesoscopic level, and, most importantly we need to understand the nature of the transition between the two levels of description. To the best if our knowledge, such needs have not been readily acknowledged in the literature, and there exists no established framework in which to address related issues. The purpose of this thesis is to present a selection of results which demonstrate that establishing the micro-meso connection is feasible, useful, and in some cases necessary in physical problems, both inside and outside the context of thermodynamics.

This thesis is structured as follows. In Chap. 2, we present our work on the quivering limit - a time-dependent billiard limit in which a billiard's wall motion becomes effectively stochastic. The evolution of a particle's energy in such a billiard is described on the mesoscopic level, and the results of our work shed light on some long-standing problems in the time-dependent billiard literature. This work is presented as given in the publication of J. Demers and C. Jarzynski in Physical Review E [35]. Chapters 3 and 4 together present a study of the microscopic Hamiltonian foundations of thermodynamics. Here, we demon-

strate the precise manner in which fast Hamiltonian chaos at the microscale generically leads to effectively stochastic and irreversible thermodynamic behavior at the mesoscale. After establishing our framework, we derive important results from stochastic thermodynamics, and we then employ the established micro-meso connection to show that these important results follow somewhat trivially from the underlying generic properties of Hamilton's equations. In Chap. 5 we study the problem of stabilizing a diffusing Brownian particle with feedback control on the mesoscopic level using stochastic thermodynamics. Here, we find that in order to avoid contradictions with the first law of thermodynamics, we need to model the stochastic driving noise with more detail than is commonly used in stochastic thermodynamics. In Chap. 6, we summarize our work and present potential avenues for future research.

Chapter 2

Universal Energy Diffusion in a Quivering Billiard

2.1 Introduction

Billiards are remarkably useful physical models; they allow a diverse range of classical dynamics to be understood intuitively through easy-to-visualize particle trajectories and are a natural setting for quantum and wave chaos [36], while the discrete time nature of particle-billiard boundary interactions make classical billiards especially amenable to numerical study. Time-dependent billiards (billiards with boundaries in motion) in particular can be found in a wide range of applications: KAM theory [37, 38, 39], one-body dissipation in nuclear dynamics [40], Fermi acceleration [41, 42, 43, 44, 38, 37, 45, 46, 47], and adiabatic energy diffusion [48, 49], for example.

This work was originally motivated by the desire to study and simulate classical particle trajectories in time-dependent billiard systems. The task is complicated by the boundary's displacement, which produces implicit equations for the time between particle-boundary collisions. We propose a fixed wall simplification by considering the limit of infinitesimally small boundary displacements. Our limit will be called the quivering limit, and the resulting billiard system will be called a quivering billiard. The purpose of this paper is to show that, although simple, quivering billiards are accurate descriptions of time-dependent billiards in

the limit of small boundary displacements, and to support our conjecture that any physically consistent, non-trivial, fixed wall simplification of a time-dependent billiard must be physically equivalent to a quivering billiard. Using physical reasoning, we will argue that in the quivering limit, deterministic billiard dynamics become inherently stochastic. Then, by utilizing the simplifications allowed by stochastic methods and fixed billiard walls, we will derive analytic expressions to describe energy evolution in a quivering billiard. Our investigations will uncover universal behavior in time-dependent billiards when billiard motion is close to the quivering limit, and our results will enable us to addresses several issues that have been raised in previous Fermi acceleration and time-dependent billiard literature.

The outline of this paper is as follows. In Sec. 2.2, we first define a quivering billiard and determine its behavior in one dimension, and then generalize to quivering billiards in arbitrary dimensions. The energy statistics of a single particle and a particle ensemble are examined in Sec. 2.3, and the results are discussed in the context previous literature in Sec. 2.4. In Sec. 2.5, we give examples of quivering billiards and present numerical analyses, and we conclude in Sec. 2.6.

2.2 The Quivering Limit

In this section, we define quivering as a particular limit of time-dependent billiard motion. Because the dynamics are so poorly behaved in this limit, billiard systems can only be described stochastically. For simplicity, we first work with a one-dimensional billiard with a single moving wall, and then extend to arbitrary billiard motion in arbitrary dimensions.

2.2.1 The 1-D Fermi-Ulam Model

Consider a particle in one dimension bouncing between two infinitely massive walls. One wall is fixed at $x = 0$, and the other oscillates about its mean position at $x = L$, where we take $L > 0$. The particle's energy fluctuates due to collisions with the moving wall, and the dynamical system corresponding to the particle's motion defines the well-known Fermi-Ulam model [43, 44, 42, 37, 38]. Suppose that the moving wall oscillates periodically

with period τ, characteristic oscillation amplitude a, and characteristic speed $u_c = a/\tau$. The moving wall's position $x(t)$ and velocity $u(t)$ at time t can be written as

$$
\begin{aligned}
x(t) &= L + g(t) \\
u(t) &= \frac{\mathrm{d}g}{\mathrm{d}t},
\end{aligned}
\tag{2.1}
$$

where $g(t)$ is some piecewise smooth τ-periodic function with mean zero. The wall velocity scales like u_c, and $g(t)$ scales like a. To make the scaling obvious, we note that $g(t)$ depends on t only through the value of $t \bmod \tau$, and we make the following substitutions:

$$
\begin{aligned}
\Psi(t) &= \frac{t}{\tau} \bmod 1 \\
g(t) &= a\, h(\Psi(t)).
\end{aligned}
\tag{2.2}
$$

The quantity $\Psi(t)$ will be referred to as the wall's phase. Here, h is regarded as a function of Ψ, and $h(\Psi(t))$ means $h(\Psi)$ evaluated for $\Psi = \Psi(t)$. The quantity $h(\Psi(t))$ is just $g(t)$ rescaled to have a characteristic oscillation amplitude of unity. The state of the wall at time t is thus

$$
\begin{aligned}
x(t) &= L + a\, h(\Psi(t)) \\
u(t) &= u_c\, h'(\Psi(t)),
\end{aligned}
\tag{2.3}
$$

where the h' denotes the derivative of h with respect to its argument Ψ.

We define the quivering limit of the Fermi-Ulam model by taking $a, \tau \to 0$ while holding u_c constant and leaving the dependence of h on Ψ fixed. In the quivering limit, the moving wall's position reduces to $x(t) = L$, so no implicit equations for the time between collisions arise from the dynamics. This simplification comes at a price; when $\tau \to 0$, Ψ oscillates infinitely fast in time, and $u(t)$ does not converge to any value for any given t. That is, in the quivering limit, $u(t)$ becomes ambiguous to evaluate. Our task now is to physically interpret and resolve this ambiguity.

Note that in the quivering limit, the wall makes infinitely erratic motions at finite speeds;

the nth derivative of $g(t)$, scaling like a/τ^n, diverges for all $n \geq 2$. An infinitesimal change in the state of a particle results in a finite and essentially unpredictable change in the wall's velocity at the time of the next bounce. We assert that one could never, even in principle, specify the state of the particle with enough precision to reliably predict the velocity of the moving wall, and thus the change in particle energy, during the next collision. We therefore claim that in the quivering limit, the dynamics of the Fermi-Ulam model become inherently stochastic; deterministic particle trajectories defined on phase space transition to stochastic processes defined on a probability space. Given any initial condition, the resulting particle trajectory actually represents one possible realization drawn from an ensemble of initial conditions infinitesimally displaced from one another. The wall's velocity during a collision will be treated as a random variable, and we now find the corresponding probability distribution.

Consider again the moving wall with non-zero a and τ. Let $P(u|0)$ be the probability density for a stationary observer to measure the velocity u during a randomly timed snapshot of the wall:

$$\begin{aligned} P(u|0) &= \frac{1}{\tau} \int_0^\tau \mathrm{d}t \, \delta\big(u - u(t)\big) \\ &= \int_0^1 \mathrm{d}\Psi \delta\big(u - u_c \, h'(\Psi)\big). \end{aligned} \tag{2.4}$$

The reason for placing the conditional $|0$ in the argument of P will become apparent shortly. We note that $P(u|0)$ is normalized, so it is indeed a well-defined probability density. In the quivering limit, u_c and the dependence of h on Ψ remain constant, so $P(u|0)$ remains well-defined and unchanged. If the stationary observer were to measure the wall velocity in the quivering limit, any observation, no matter how well-timed, would be an essentially random snapshot due to the wall's infinitely erratic motion. We thus take $P(u|0)$ to be the probability for a stationary observer to measure the wall with velocity u when the wall is quivering.

The particle bouncing between the walls effectively measures the wall's velocity during collisions, but the particle is not a stationary observer. Collisions with large relative speeds of approach occur more frequently than collisions with small relative speeds of approach,
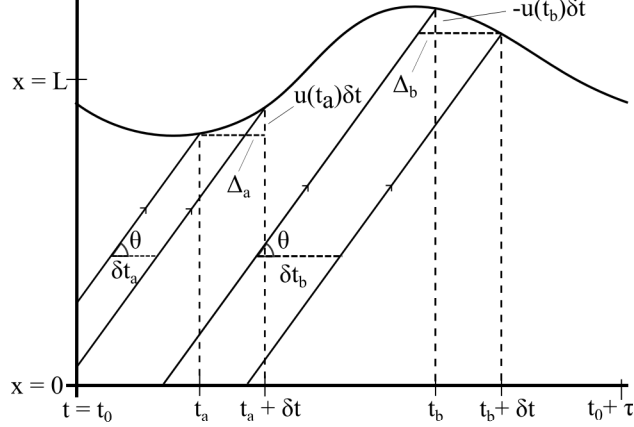
Figure 2.1: Spacetime diagram over one period of a moving wall's motion. The smooth curve represents the wall's position, and particles approach the wall along the diagonal arrows to collide at times $t_a, t_a + \delta t, t_b$ and $t_b + \delta t$

so there exists a statistical bias that favors collisions for which the wall moves towards the particle. If the quivering dynamics are to be physically consistent with the Fermi-Ulam dynamics, this statistical bias must be incorporated into the probability distribution used to determine the wall's velocity during collisions. The mathematical realization of the statistical bias can be found with the aid of Fig. 2.1, a construction first employed by Hammersley [44] and Brahic [37].

In Fig. 2.1, the position of the moving wall in the Fermi-Ulam model is plotted over one period of motion in the interval $(t_0, t_0 + \tau)$. Consider an ensemble of particles approaching the moving wall with speed v. For the moment, we assume that v is larger than the maximum wall velocity u_{max}. The particles are launched from $x = 0$ at a uniform rate over a period of duration τ such that they all collide with the wall during the interval $(t_0, t_0 + \tau)$. We concern ourselves only with the first collision each particle makes with the moving wall. Four trajectories from the ensemble are shown in Fig. 2.1, representing collisions with the wall at times t_a, $t_a + \delta t$, t_b, and $t_b + \delta t$. Because the launch times are uniformly distributed, the fraction of particles that collide with the wall between t_a and $t_a + \delta t$ will be proportional to the interval $\delta t_a = \delta t - \Delta_a$. Likewise, the fraction that collide between t_b and $t_b + \delta t$ will be proportional to $\delta t_b = \delta t + \Delta_b$. Using the geometry of Fig. 2.1 and the fact that $\tan(\theta) = v$, we find the probability density for randomly selected ensemble member collide

11

with the moving wall at a time t within the interval $(t_0, t_0 + \tau)$ to be

$$P(u(t)|v) = \frac{1}{\tau}\left(1 - \frac{u(t)}{v}\right). \tag{2.5}$$

Multiplying by a delta function and integrating Eq. (2.5) over a period of the wall's motion gives $P(u|v)$, the probability density for a randomly selected ensemble member's collision to occur when the wall moves with velocity u:

$$\begin{aligned}
P(u|v) &= \frac{1}{\tau}\int_0^\tau dt\, \delta(u - u(t))\left(1 - \frac{u(t)}{v}\right) \tag{2.6}\\
&= \int_0^1 d\Psi\, \delta(u - u_c\, h'(\Psi))\left(1 - \frac{u_c\, h'(\Psi)}{v}\right)\\
&= P(u|0)\left(1 - \frac{u}{v}\right).
\end{aligned}$$

Because the wall's average displacement over one period of motion is zero, the product $uP(u|0)$ integrated over all wall velocities must also give zero, and $P(u|v)$ is therefore normalized and a well-defined probability density. The distribution $P(u|v)$ has a statistical bias towards larger negative u due to the flux factor $1 - u/v$. We will henceforth refer to $P(u|0)$ as the unbiased distribution and $P(u|v)$ as the biased distribution. In the quivering limit, $P(u|v)$ remains well-defined and unchanged. As $\tau \to 0$, an ensemble of particles launched over a period of wall motion from a fixed x is essentially equivalent to an ensemble of infinitesimally displaced initial conditions. We therefore take $P(u|v)$ to be the conditional probability density to observe a quivering wall with velocity u during a collision, given that the particle approaches the wall with speed $v > u_{max}$.

If a particle approaches the moving wall with speed $v < u_{max}$, then $P(u|v)$ will become negative for some values of u, and Eq. (2.6) will make no sense as a probability density. These u values correspond to impossible collisions for which the wall moves with positive velocity away from the particle faster than the particle moves toward the wall. Such collisions occur with probability zero, and we can account for this by simply attaching a step-function to

12

the biased distribution, yielding

$$P(u|v) = \begin{cases} P(u|0)\left(1 - \frac{u}{v}\right), & v \geq u_{max} \\ N(v)P(u|0)\left(1 - \frac{u}{v}\right)\Theta(v - u), & v < u_{max}, \end{cases} \tag{2.7}$$

where $\Theta(x)$ is the unit step function (equal to 0 for $x < 0$ and 1 for $x \geq 0$) and $N(v)$ is a v dependent normalization.

Equation (2.7) determines the statistics of a particle's energy evolution in a quivering Fermi-Ulam system. As with any billiard system, the particle's energy is simply the kinetic energy $\frac{1}{2}mv^2$, where m is the particle's mass and v is its speed. The particle bounces between the two walls as if the system were time-independent, but when colliding with the quivering wall at an incoming speed v_i (the particle moves in the positive x direction to collide with the moving wall, so v_i is also the incoming velocity), a value for the wall velocity u is selected using the biased distribution $P(u|v_i)$. The particle's velocity just after the collision, v_f, is given by

$$v_f = 2u - v_i, \tag{2.8}$$

and the corresponding energy change, ΔE, is given by

$$\Delta E = 2mu^2 - 2mu\,v_i. \tag{2.9}$$

Equations (2.8) and (2.9) are determined using the standard collision kinematics for a particle in one-dimension colliding elastically with an infinitely massive moving object.

Before moving on to higher dimensions, we must address the possibility of particles escaping the billiard interior. This issue will plague any fixed wall simplification of time-dependent billiards, and is discussed in detail in Ref. [50]. From Eq. (2.8), we see that if $0 < u < v_i \leq 2u$, the particle does not turn around after a collision with the moving wall, but instead slows down and continues forward. We refer to these types of collisions as glancing collisions. For non-zero a and τ, just after a glancing collision, the particle continues forward slower than the wall moves outward, so the particle will remain within the billiard interior. With a fixed wall simplification, however, the wall does not actually

move outward after a glancing collision, so the particle will continue forward and escape the billiard interior. A particle escaping through a hard wall is a non-physical by-product of setting $a = 0$, so in order to make a physically reasonable fixed wall simplification, one must always devise a method to handle glancing collisions. Our method for a quivering Fermi-Ulam system is devised as follows.

For non-zero a and τ, after a glancing collision occurs, the wall continues to evolve through its period, and one of two possibilities will occur. The wall may slow down sometime after the glancing collision and allow the particle to catch up and collide again, or the wall may reverse its direction and move inward sometime after the glancing collision, also allowing the particle to collide again. In either case, a second collision occurs after the first collision, and as a and τ approach zero, the second occurs essentially instantaneously after the first. Therefore, we treat a glancing collision in a quivering Fermi-Ulam billiard as a double collision. When a particle with speed v_i (also the particle's velocity) collides with the quivering wall, we draw a u value from the distribution $P(u|v_i)$. If the selected value of u is such that $0 < u < v_i \leq 2u$, the particle's new speed v_f (also velocity) is given by $v_f = 2u - v_i$, and we draw a new u value from the distribution $P(u|v_f)$. If the second u value gives another glancing collision, we again update the particle's speed and then draw a third u value. The process is repeated until a non-glancing collision occurs, and the whole event (which occurs instantaneously) is treated as a single collision.

2.2.2 Arbitrary Time-Dependent Billiards

We now generalize to arbitrary billiards in arbitrary dimensions. Consider a time-dependent billiard in d dimensions moving periodically through some continuous sequence of shapes with period τ, characteristic oscillation amplitude a, and characteristic speed $u_c = a/\tau$. The evolution of any one point on the boundary will be denoted by the path $\mathbf{q}(t)$, where $\mathbf{q}(t + \tau) = \mathbf{q}(t)$. For every t, the set of all boundary points $\{\mathbf{q}(t)\}$ is assumed to define a collection of unbroken $d - 1$ dimensional surfaces, which we refer to as the boundary components, enclosing some d dimensional bounded connected volume. The outward unit normal to the billiard boundary at the point $\mathbf{q}(t)$ is denoted by $\hat{\mathbf{n}}(\mathbf{q}(t))$, and

the velocity of the boundary point $\mathbf{q}(t)$ is denoted by $\mathbf{u}(\mathbf{q}(t)) = d\mathbf{q}(t)/dt$. The billiard shape evolves continuously in time, and we assume that the boundary components remain unbroken throughout their evolution, so $\mathbf{u}(\mathbf{q}(t))$ forms a smooth vector field with domain on the boundary $\{\mathbf{q}(t)\}$ for any fixed t. Likewise, $\hat{\mathbf{n}}(\mathbf{q}(t))$ forms a smooth field on $\{\mathbf{q}(t)\}$ for any fixed t, except possibly at corners, where $\hat{\mathbf{n}}(\mathbf{q}(t))$ is ill-defined and discontinuous. We denote the outward normal velocity of the point $\mathbf{q}(t)$ by $u(\mathbf{q}(t)) = \mathbf{u}(\mathbf{q}(t)) \cdot \hat{\mathbf{n}}(\mathbf{q}(t))$.

Denote by $\mathbf{q}$ the average of $\mathbf{q}(t)$ over one period:

$$\mathbf{q} = \frac{1}{\tau} \int_0^\tau dt\, \mathbf{q}(t). \tag{2.10}$$

Noting that the boundary components remain unbroken throughout the period of motion, it is straightforward to show that set of average boundary points $\{\mathbf{q}\}$ forms a collection of unbroken $d-1$ dimensional surfaces. The trajectory $\mathbf{q}(t)$ and normal velocity $u(\mathbf{q}(t))$ of any given boundary point can be written as functions of the corresponding average location $\mathbf{q}$ and the time t:

$$
\begin{aligned}
\mathbf{q}(t) &= \mathbf{q} + \mathbf{g}(\mathbf{q}, t) \\
u(\mathbf{q}, t) &= \partial_t \mathbf{g}(\mathbf{q}, t) \cdot \hat{\mathbf{n}}(\mathbf{q}(t)),
\end{aligned}
\tag{2.11}
$$

where $\mathbf{g}(\mathbf{q}, t)$ is a piecewise smooth in time τ periodic function with a time average of zero. $\mathbf{g}(\mathbf{q}, t)$ scales like a and $u(\mathbf{q}(t))$ scales like u_c. Equation (2.11) depends on t only through the value of $\Psi(t) = t/\tau \bmod 1$, so we write

$$
\begin{aligned}
\mathbf{q}(t) &= \mathbf{q} + a\,\mathbf{h}(\mathbf{q}, \Psi(t)) \\
u(\mathbf{q}, t) &= u_c\, \partial_\Psi \mathbf{h}(\mathbf{q}, \Psi(t)) \cdot \hat{\mathbf{n}}(\mathbf{q}(t)).
\end{aligned}
\tag{2.12}
$$

where $a\,\mathbf{h}(\mathbf{q}, \Psi(t)) = \mathbf{g}(\mathbf{q}, t)$. Analogously to the one dimensional case, $\mathbf{h}$ is regarded as a function of $\mathbf{q}$ and Ψ, and $\mathbf{h}(\mathbf{q}, \Psi(t))$ means $\mathbf{h}(\mathbf{q}, \Psi)$ evaluated for $\Psi = \Psi(t)$. The quivering limit of an arbitrary dimensional billiard is defined by taking $a, \tau \to 0$ while holding u_c and the dependence of $\mathbf{h}$ on Ψ and $\mathbf{q}$ constant. In this limit, the billiard's boundary points

15
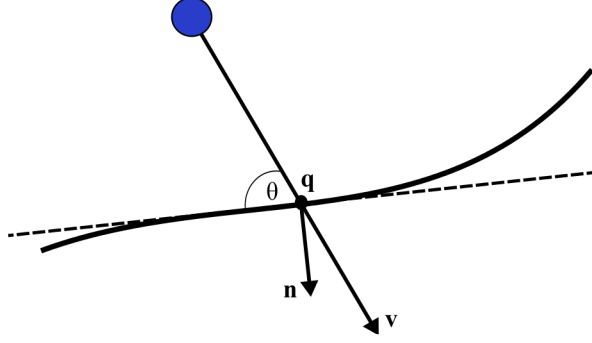
Figure 2.2: Collision geometry in a two-dimensional billiard. A particle with velocity $\mathbf{v}$ approaches the point $\mathbf{q}$ on the billiard boundary, where the outward unit normal vector is $\mathbf{n}$. The dotted line represents the tangent line to the boundary at $\mathbf{q}$

become fixed in time at the average locations $\{\mathbf{q}\}$, so the outward normal vectors become fixed in time as well. Thus, in the quivering limit, we have

$$
\begin{aligned}
\mathbf{q}(t) &= \mathbf{q} & (2.13)\\
u(\mathbf{q}, t) &= u_c\, \partial_\Psi \mathbf{h}(\mathbf{q}, \Psi(t)) \cdot \hat{\mathbf{n}}(\mathbf{q}) \\
&= u_c\, h'(\mathbf{q}, \Psi(t)),
\end{aligned}
$$

where we write $h'(\mathbf{q}, \Psi(t)) = \partial_\Psi \mathbf{h}(\mathbf{q}, \Psi(t)) \cdot \hat{\mathbf{n}}(\mathbf{q})$ for brevity. Any time-dependent billiard taken to the quivering limit will be called a quivering billiard.

Analogously to the one dimensional case, we define the unbiased distribution for each $\mathbf{q}$:

$$
P(u|0, \mathbf{q}) = \int_0^1 \mathrm{d}\Psi\, \delta\big(u - u_c\, h'(\mathbf{q}, \Psi)\big). \tag{2.14}
$$

The biased distribution for each $\mathbf{q}$ can also be defined analogously to the one dimensional case, but we must also consider the collision angle θ, depicted for two-dimensional billiard in Fig. 2.2. For a particle approaching the boundary point $\mathbf{q}$ with speed v, θ is the angle between the particle's velocity vector and the $d-1$ dimensional tangent surface to the wall at $\mathbf{q}$, and $v\sin(\theta)$ thus gives the component of the particle's velocity in the $\hat{\mathbf{n}}(\mathbf{q})$ direction. If the particle collides when the wall has normal velocity u, then the relative speed of approach just before the collision is determined by $v\sin(\theta)$ and u, so $v\sin(\theta)$ determines the statistical bias towards collisions with large negative u. We account for this by simply replacing v

16

with $v\sin(\theta)$ in Eq. (2.7), yielding

$$
P(u|v,\mathbf{q},\theta) = \begin{cases} P(u|0,\mathbf{q})\left(1 - \frac{u}{v\sin(\theta)}\right), & v\sin(\theta) \geq u_{max}(\mathbf{q}) \\ N(v,\theta)P(u|0,\mathbf{q})\left(1 - \frac{u}{v\sin(\theta)}\right)\Theta(v\sin(\theta) - u), & v\sin(\theta) < u_{max}(\mathbf{q}), \end{cases}
$$
(2.15)

Equation (2.15) determines the statistics of a particle's energy evolution in a quivering billiard.

To summarize, we describe how one may construct a quivering billiard and determine a particle's trajectory, without the need to define a real, fully time-dependent billiard and take the quivering limit. First, one must select a billiard shape by defining a surface $\{\mathbf{q}\}$, then set boundary quivering by giving a value to u_c and defining a scalar field $h'(\mathbf{q}, \Psi)$ on $\{\mathbf{q}\}$. If the constructed quivering billiard is to honestly represent some deterministic billiard's motion in the quivering limit, then $h'(\mathbf{q}, \Psi)$ should be chosen to be a smooth function of $\mathbf{q}$ for any Ψ wherever $\hat{\mathbf{n}}(\mathbf{q})$ in continuous. Using the field h' and the value of u_c, one may then calculate the unbiased distribution $P(u|0,\mathbf{q})$ from Eq. (2.14) for any $\mathbf{q}$ on the billiard boundary. For a particle in free flight inside the quivering billiard, the next collision location is found deterministically using the geometry of the billiard boundary, just as with a time-independent billiard. When a particle with velocity $\mathbf{v}_i$ and speed v_i collides with the boundary at $\mathbf{q}$ with a collision angle θ_i, we draw a value of u from the distribution $P(u|v_i,\mathbf{q},\theta_i)$. The particle's velocity component tangent to the boundary remains constant, and the component normal to the boundary just after the collision, $\mathbf{v}_f \cdot \hat{\mathbf{n}}(\mathbf{q})$, is given by

$$
\begin{aligned}
\mathbf{v}_f \cdot \hat{\mathbf{n}}(\mathbf{q}) &= 2u - \mathbf{v}_i \cdot \hat{\mathbf{n}}(\mathbf{q}) \\
&= 2u - v_i \sin(\theta_i).
\end{aligned}
$$
(2.16)

The corresponding change in energy, ΔE, is given by

$$
\Delta E = 2mu^2 - 2mu\,v_i \sin(\theta_i).
$$
(2.17)

Analogously to the one dimensional case, if the selected value of u is such that $0 < u < v_i \sin(\theta_i) \leq 2u$, then a glancing collision occurs, and we draw a second value of u using the same collision located and updated particle speed and collision angle, determined from Eqs. (2.16) and (2.17).

2.3   Energy Statistics

In this section, we study in detail the statistical behavior of particles and ensembles in a d-dimensional quivering billiard, with the aim of describing energy evolution of a ensemble of initial conditions as a diffusion process. Our notation will be as follows: $\mathbf{q}_b$ is the location of a particle's b^{th} collision with the billiard boundary, θ_b is the bth collision angle, u_b is the selected value of the wall velocity during the bth collision (sampled using Eq. (2.15)), v_{b-1} is the particle's speed just before the bth collision, and ΔE_b is the change in particle energy due to the bth collision, given by

$$\Delta E_b = 2mu_b^2 - 2mu_b\, v_{b-1} \sin\left(\theta_b\right). \tag{2.18}$$

In order to derive analytic results, we will assume that the initial particle speeds v_0 are much larger than u_c, and we will solve to leading order in the small parameter $\varepsilon = u_c/v_0$. We regard u_c as an $O(1)$ quantity, and v_0 as an $O(\varepsilon^{-1})$ quantity. This approximation allows us to ignore glancing collisions in our analysis, and also allows us ignore the possibility of $v_{b-1} \sin(\theta_b) \leq u_{max}(\mathbf{q}_b)$, so that the biased distributions at the time of collision always take the form $P(u_b|v_{b-1}, \mathbf{q}_b, \theta_b) = P(u_b|0, \mathbf{q}_b)\left(1 - u_b/v_{b-1} \sin(\theta_b)\right)$ (as opposed to the more complicated Eq. (2.15)). The assumption $\varepsilon \ll 1$ is not particularly restrictive; even if particles begin with an initial speed comparable to or less than u_c, energy gaining collisions are more likely than energy losing collisions due to the flux factor in the biased distribution, and a slow particle will gain roughly mu_c^2 of energy during a collision according to Eq. (2.18). Therefore, a slow particle will more than likely gain speed $u_c \sim O(1)$ during a single bounce, and after $1/\delta$ bounces, where $\delta \ll 1$ is some small number, the particle will more than likely have a speed v such that $u_c/v \lesssim \delta \ll 1$. Thus, slow particles are very likely to eventually

18

become fast particles, and the assumption $u_c/v \ll 1$ will give a better and better approximation over time.

In the analysis, it will prove useful to consider both the *full dynamics* and *frozen dynamics*, as is done in Refs. [51, 49]. If the frozen dynamics are used at the bth collision, the energy change ΔE_b is calculated, but the particle's energy remains constant, and the angle of reflection is equal to the collision angle θ_b. In other words, the frozen dynamics are identical to those of a time-independent billiard, but we calculate and keep track of the ΔE_b's that would have occurred had the billiard walls been quivering. In the full dynamics, the particle's energy is actually incremented by the calculated value of ΔE_b, and the angle of reflection is consequently altered.

2.3.1 Expectations

Consider single a particle with energy E_0 released at time t_0 in a d-dimensional quivering billiard. The resulting particle trajectory generates a sequence of energy increments $\{\Delta E_1, \Delta E_2, ..., \Delta E_{b-1}, \Delta E_b, \Delta E_{b+1}, ...\}$. Let the operator $\{...\}_b$ denote the conditional expectation value of the quantity ..., given the outcomes of the previous $b-1$ bounces. The first $b-1$ bounces determine v_{b-1}, $\mathbf{q}_b$, and θ_b, so the bth conditional expected energy change, $\mu_b \equiv \{\Delta E_b\}_b$, can be calculated using the biased distribution $P(u_b|v_{b-1}, \mathbf{q}_b, \theta_b)$ and the expression for ΔE_b in Eq. (2.18):

$$
\begin{aligned}
\mu_b &\equiv \{\Delta E_b\}_b \\
&= \int \mathrm{d}u_b P(u_b|v_{b-1}, \mathbf{q}_b, \theta_b) \Delta E_b \\
&= \int \mathrm{d}u_b P(u_b|0, \mathbf{q}_b) \left(4mu_b^2 - \frac{2mu_b^3}{v_{b-1}\sin(\theta_b)} - 2mu_b\, v_{b-1}\sin(\theta_b) \right).
\end{aligned}
\tag{2.19}
$$

The integral in Eq. (2.19) is taken over all possible values of u_b at $\mathbf{q}_b$.

Let $M_n(\mathbf{q}_b)$ denote the nth moment of the wall velocity at $\mathbf{q}_b$ as measured by a stationary observer:

$$
M_n(\mathbf{q}) = \int \mathrm{d}u P(u|0, \mathbf{q}) u^n.
\tag{2.20}
$$

By construction, $M_1(\mathbf{q}) = 0$ for all $\mathbf{q}$. Otherwise, $M_n(\mathbf{q}_b)$ generally scales like u_c^n. The conditional mean thus simplifies to

$$
\begin{aligned}
\mu_b &= \int \mathrm{d}u_b P(u_b|0, \mathbf{q}_b) 4 m u_b^2 \left(1 - \frac{u_b}{2 v_{b-1} \sin(\theta_b)} \right) & (2.21) \\
&= 4 m M_2(\mathbf{q}_b) \left(1 - \frac{M_3(\mathbf{q}_b)/M_2(\mathbf{q}_b)}{2 v_{b-1} \sin(\theta_b)} \right).
\end{aligned}
$$

Similarly, the conditional variance σ_b^2 is given by

$$
\begin{aligned}
\sigma_b^2 &\equiv \{(\Delta E_b)^2\}_b - \{\Delta E_b\}_b^2 & (2.22) \\
&= \int \mathrm{d}u_b P(u_b|v_{b-1}, \mathbf{q}_b, \theta_b) \left((\Delta E_b)^2 - \{\Delta E_b\}_b^2 \right) \\
&= 4 m^2 [M_2(\mathbf{q}_b)]^2 \left(\frac{v_{b-1}^2 \sin^2(\theta_b)}{M_2(\mathbf{q}_b)} - 3 \frac{v_{b-1} \sin(\theta_b)}{[M_2(\mathbf{q}_b)]^2/M_3(\mathbf{q}_b)} \right. \\
&\quad + 3 \frac{M_4(\mathbf{q}_b)}{[M_2(\mathbf{q}_b)]^2} - 4 + \frac{4 M_3(\mathbf{q}_b)/M_2(\mathbf{q}_b) - M_5(\mathbf{q}_b)/[M_2(\mathbf{q}_b)]^2}{v_{b-1} \sin(\theta_b)} \\
&\quad \left. - \frac{[M_3(\mathbf{q}_b)]^2/[M_2(\mathbf{q}_b)]^2}{v_{b-1}^2 \sin^2(\theta_b)} \right).
\end{aligned}
$$

The terms enclosed in the parentheses of Eqs. (2.21) and (2.22) are ordered in increasing powers of ε. To leading order, we have

$$
\mu_b = 4 m M_2(\mathbf{q}_b) \tag{2.23}
$$
$$
\sigma_b^2 = 4 m^2 M_2(\mathbf{q}_b) v_{b-1}^2 \sin^2(\theta_b)
$$

The quantities μ_b and σ_b^2 are $O(1)$ and $O\left(\varepsilon^{-2}\right)$, respectively; average energy gain is moderate, and fluctuations are huge.

2.3.2   Correlations

The conditional covariance between adjacent bounces, $\mathrm{Cov}_{b,b+1}$, is defined by

$$
\begin{aligned}
\mathrm{Cov}_{b,b+1} &\equiv \{(\Delta E_b - \{\Delta E_b\}_b)(\Delta E_{b+1} - \{\Delta E_{b+1}\}_b)\}_b & (2.24) \\
&= \{\Delta E_b \Delta E_{b+1}\}_b - \{\Delta E_b\}_b \{\Delta E_{b+1}\}_b.
\end{aligned}
$$

The conditional expectations in Eq. (2.24) are taken given the outcomes of the previous $b - 1$ collisions, with the outcome of the bth collision yet to be determined. That is, we must average over all possible realizations of the stochastic process $E_{b-1} \to E_{b-1} + \Delta E_b \to E_{b-1} + \Delta E_b + \Delta E_{b+1}$, given the first $b-1$ collisions. Denote $\{\Delta E_{b+1}|u_b\}_{b+1}$ as the conditional expectation of E_{b+1}, given the first $b - 1$ collision outcomes and supposing that u_b is the wall velocity during the bth collision. The expression for $\{\Delta E_{b+1}\}_b$ is then

$$\{\Delta E_{b+1}\}_b = \int \mathrm{d}u_b P(u_b|v_{b-1}, \mathbf{q}_b, \theta_b)\{\Delta E_{b+1}|u_b\}_{b+1}. \tag{2.25}$$

The expression for $\{\Delta E_b \Delta E_{b+1}\}_b$ can be written similarly:

$$\begin{aligned}
\{\Delta E_b \Delta E_{b+1}\}_b &= \int \mathrm{d}u_b \mathrm{d}u_{b+1} P(u_b|v_{b-1}, \mathbf{q}_b, \theta_b) P(u_{b+1}|v_b, \mathbf{q}_{b+1}, \theta_{b+1}|u_b)\Delta E_b \Delta E_{b+1} \\
&= \int \mathrm{d}u_b P(u_b|v_{b-1}, \mathbf{q}_b, \theta_b)\Delta E_b\{\Delta E_{b+1}|u_b\}_{b+1}.
\end{aligned} \tag{2.26}$$

The term $P(u_{b+1}|v_b, \mathbf{q}_{b+1}, \theta_{b+1}|u_b)$ denotes the value of $P(u_{b+1}|v_b, \mathbf{q}_{b+1}, \theta_{b+1})$ when v_b, θ_{b+1}, and $\mathbf{q}_{b+1}$ are determined given the first $b - 1$ collision outcomes while supposing that u_b is the wall velocity upon the bth collision. Equation (2.24) can thus be expressed as

$$\mathrm{Cov}_{b,b+1} = \int \mathrm{d}u_b P(u_b|v_{b-1}, \mathbf{q}_b, \theta_b)\{\Delta E_{b+1}|u_b\}_{b+1} \left(\Delta E_b - \{\Delta E_b\}_b\right). \tag{2.27}$$

If the frozen dynamics are used at the bth collision, then v_b, θ_{b+1}, and $\mathbf{q}_{b+1}$ are independent of u_b, so we have

$$\{\Delta E_{b+1}|u_b\}_{b+1}|_F = \{\Delta E_{b+1}\}_{b+1}|_F = \mu_{b+1}|_F, \tag{2.28}$$

where $...|_F$ denotes the quantity ... evaluated using the frozen dynamics. $\mu_{b+1}|_F$ carries no u_b dependence, so it can be brought outside of the integral in Eq. (2.27), giving

$$\mathrm{Cov}_{b,b+1}|_F = 0. \tag{2.29}$$

Adjacent energy increments are thus statistically uncorrelated in the frozen dynamics.

Under the assumption $\varepsilon \ll 1$, the frozen dynamics closely resemble the full dynamics over the time scale of a few bounces [49]. Over such a time scale, we can regard the full dynamics trajectory as a stochastic perturbation of the deterministic frozen dynamics trajectory. Let $\mathbf{q}_{b+1}|u_b = \mathbf{q}_{b+1}|_F + \delta\mathbf{q}_{b+1}|u_b$ be the $(b+1)^{th}$ collision location when the full dynamics are used at the bth bounce, given the first $b-1$ collisions and supposing that u_b is the observed wall velocity upon the bth collision. Equation (2.23) then gives, to leading order in ε

$$
\begin{aligned}
\{\Delta E_{b+1}|u_b\}_{b+1} &= 4mM_2(\mathbf{q}_{b+1}|u_b) \\
&= 4mM_2(\mathbf{q}_{b+1}|_F) + 4m\,\nabla M_2(\mathbf{q}_{b+1}|_F) \cdot \delta\mathbf{q}_{b+1}|u_b.
\end{aligned}
\tag{2.30}
$$

where the gradient ∇M_2 is constrained to act along directions tangent to the billiard boundary at $\mathbf{q}_{b+1}|_F$. In Appendix A, we solve for $\|\delta\mathbf{q}_{b+1}|u_b\|$ to leading order in ε and find

$$
\|\delta\mathbf{q}_{b+1}|u_b\| = 2L_b|_F \frac{\cos(\theta_b)}{\sin(\theta_{b+1}|_F)} \frac{|u_b|}{v_{b-1}},
\tag{2.31}
$$

where $L_b|_F$ is the distance between the bth and $b+1^{th}$ collision locations in the frozen dynamics. Combining Eqs. (2.27), (2.30), and (2.31), gives to leading order in ε

$$
\begin{aligned}
\mathrm{Cov}_{b,b+1} &= \int \mathrm{d}u_b P(u_b|v_{b-1}, \mathbf{q}_b, \theta_b)\left(4m\,\nabla M_2(\mathbf{q}_{b+1}|_F) \cdot \delta\mathbf{q}_{b+1}|u_b\right)(\Delta E_b - \{\Delta E_b\}_b) \\
&= \int \mathrm{d}u_b P(u_b|0, \mathbf{q}_b)\left(4m\nabla M_2(\mathbf{q}_{b+1}|_F) \cdot \frac{\delta\mathbf{q}_{b+1}|u_b}{\|\delta\mathbf{q}_{b+1}|u_b\|}\right. \\
&\quad \times\ 2L_b|_F \frac{\cos(\theta_b)}{\sin(\theta_{b+1}|_F)}\frac{|u_b|}{v_{b-1}}\right)\left(4mu_b^2 - 2m\frac{u_b^3}{v_{b-1}\sin(\theta_b)} - 2mu_b v_{b-1}\sin(\theta_b)\right. \\
&\quad \left.\left.-4mM_2(\mathbf{q}_b) + 4mM_2(\mathbf{q}_b)\frac{u_b}{v_{b-1}\sin(\theta_b)}\right)\right. \\
&= -16m^2 L_b|_F \frac{\cos(\theta_b)\sin(\theta_b)}{\sin(\theta_{b+1}|_F)} \\
&\quad \times \nabla M_2(\mathbf{q}_{b+1}|_F) \cdot \int \mathrm{d}u_b P(u_b|0, \mathbf{q}_b)\frac{\delta\mathbf{q}_{b+1}|u_b}{\|\delta\mathbf{q}_{b+1}|u_b\|}u_b|u_b|.
\end{aligned}
\tag{2.32}
$$

All but the leading order terms are dropped in the last line of Eq. (2.32). With exception to the one-dimensional case, $\mathrm{Cov}_{b,b+1}$ is thus an $O(1)$ quantity. In a one-dimensional billiard,

the frozen and full dynamics always give the same collision location, so $\{\Delta E_{b+1}|u_b\}_{b+1} = \{\Delta E_{b+1}\}_{b+1}|_F$, and consequently, $\mathrm{Cov}_{b,b+1}$ is identically zero.

The conditional correlation $\rho_{b,b+1}$ is defined as the normalized conditional covariance, and is given by

$$\rho_{b,b+1} = \frac{\mathrm{Cov}_{b,b+1}}{\sigma_b\{\sigma_{b+1}\}_b}. \tag{2.33}$$

To leading order in ε, the conditional expectation $\{\sigma_{b+1}\}_b$ can be taken as the frozen dynamics value in Eq. (2.33). Therefore, the conditional correlation $\rho_{b,b+1}$ is $O(\varepsilon^2)$ (with exception to the one-dimensional case, where $\rho_{b,b+1} = 0$). This quantity is very small, and correlations between more distant collisions will further diminish due to the mixing of particle trajectories induced by the stochastic wall motion. We thus conclude that, in any dimension, correlations between energy increments effectively decay over the time scale of a single collision.

2.3.3 Ensemble averages

Consider now a microcanonical ensemble of independent particles with energy E_0 released at time t_0. The resulting trajectories will generate an ensemble of statistically independent energy increment sequences, and we denote $\Delta E_{i,b}$ as the bth recorded energy increment of the ith particle. Define the ensemble averaged b^{th} energy increment $\langle \Delta E_b \rangle$ as

$$\langle \Delta E_b \rangle \underset{N\to\infty}{=} \sum_{i=1}^{N} \frac{\Delta E_{i,b}}{N}, \tag{2.34}$$

and the ensemble averaged b^{th} conditional mean $\langle \mu_b \rangle$ as

$$\langle \mu_b \rangle \underset{N\to\infty}{=} \sum_{i=1}^{N} \frac{\mu_{i,b}}{N}, \tag{2.35}$$

where $\mu_{i,b} = \{\Delta E_{i,b}\}_b$. Equation (2.22) shows that the bth conditional variances $\sigma_{i,b}^2 \equiv \{(\Delta E_{i,b})^2\}_b - \{\Delta E_{i,b}\}^2$ are finite and bounded from above. Noting this, and the fact that

23

the series $\sum\limits_{k=1}^{\infty} k^{-2}$ converges, we deduce

$$\lim_{N\to\infty} \sum_{k=1}^{N} \frac{\sigma_{k,b}^2}{k^2} < \infty. \tag{2.36}$$

By Kolmogorov's strong law of large numbers [52], Eq. (2.36) assures that, with probability unity,

$$\langle \Delta E_b \rangle = \langle \mu_b \rangle. \tag{2.37}$$

Combining Eqs. (2.37),(2.35), and (2.23) gives, to leading order in ε,

$$\begin{aligned} \langle \Delta E_b \rangle \quad &\underset{N\to\infty}{=} \quad \sum_{i=1}^{N} \frac{4mM_2(\mathbf{q}_{i,b})}{N} \\ &= \quad 4m\langle M_2(\mathbf{q}_b)\rangle, \end{aligned} \tag{2.38}$$

where $\mathbf{q}_{i,b}$ denotes the bth collision location of the ith particle. By similar law of large number arguments, we also have, to leading order in ε,

$$\langle u_b^2 \rangle = \langle M_2(\mathbf{q}_b)\rangle, \tag{2.39}$$

where

$$\langle u_b^2 \rangle \underset{N\to\infty}{=} \sum_{i=1}^{N} \frac{u_{i,b}^2}{N}, \tag{2.40}$$

and $u_{i,b}$ is the wall velocity during the bth collision of the ith particle. To leading order, we thus have

$$\langle \Delta E_b \rangle = 4m\langle u_b^2 \rangle. \tag{2.41}$$

2.3.4 Energy diffusion

We now consider the normalized energy distribution of an ensemble of independent particles, denoted by $\eta(E,t)$. We have thus far shown that energy of any one ensemble member evolves stochastically, in small increments, with correlations in energy changes effectively decaying over a characteristic time scale given by time between collisions. A

particle's energy evolution is therefore effectively a Markov process describing a random walk along an energy axis, so following Refs. [48, 49], we assert that $\eta(E, t)$ evolves like a diffusion process and obeys a Fokker-Planck equation:

$$\partial_t \eta(E, t) = -\partial_E \left[g_1(E, t) \eta(E, t) \right] + \frac{1}{2} \partial_E^2 \left[g_2(E, t) \eta(E, t) \right]. \tag{2.42}$$

The functions $g_1(E, t)$ and $g_2(E, t)$, the drift and diffusion terms, respectively, are to be determined in this section. The energy of any one particle in a quivering billiard evolves discretely in time, so the continuous time evolution implied by Eq. (2.42) will be an accurate description of the ensemble only down to a coarse-grained time scale. The time scale must be large enough to ensure that most particles in the ensemble experience at least a few bounces off the billiard wall, but small enough to ensure the energy change experienced by most particles is small compared to their total energy. Generally speaking, a diffusive description of a stochastic process is only accurate over time scales larger than the process's typical correlation time [51, 49]. We have established that energy correlations for any one particle effectively decay over the time scale of a single collision, thus, the diffusion approach to energy evolution in a quivering billiard is justified on any time scale over which $\eta(E, t)$ can be described by a continuous evolution.

The drift term $g_1(E', t')$ is defined as the rate of ensemble averaged energy change for an ensemble of particles all with energy E' at time t'. Specifically,

$$g_1(E', t') \underset{\Delta t \to 0}{=} \frac{\langle E(t' + \Delta t) - E(t') \rangle}{\Delta t}, \tag{2.43}$$

where $E(t') = E'$ for all particles in the ensemble, and $\langle E(t' + \Delta t) \rangle$ is the ensemble averaged particle energy at time $t' + \Delta t$. We can not actually take the limit $\Delta t \to 0$ because g_1 has no meaning over time scales for which the evolution of η appears discontinuous. Instead, we will let Δt be the average time for which the ensemble members make B bounces after time t', and we will find corresponding ensemble averaged change in energy. We assume that B is small enough so that the particle energies change very little relative to E' over the time Δt, so that Δt is the smallest coarse-grained time scale for which Eq. (2.42) is valid

for an ensemble with common energy E'. We let E_B be a particle's energy B bounces after t', and find from Eq. (2.37)

$$\langle E_B - E' \rangle \;=\; \left\langle \sum^B \Delta E_b \right\rangle \tag{2.44}$$

$$=\; \sum^B 4m \langle u_b^2 \rangle.$$

We denote the coarse grained squared wall speed by $\overline{u^2}(t'; B)$, defined as the time average of $\langle u_b^2 \rangle$ over the first B bounces after t':

$$\overline{u^2}(t'; B) = \sum^B \frac{\langle u_b^2 \rangle}{B}. \tag{2.45}$$

We thus have

$$\langle E_B - E' \rangle = 4m\, \overline{u^2}(t'; B)B \tag{2.46}$$

The time scale Δt corresponding to the B bounces after t' is the ensemble averaged total free flight time over which the B bounces occur. If we denote by Δt_b a particle's b^{th} free flight time after t', we have

$$\Delta t = \sum^B \langle \Delta t_b \rangle. \tag{2.47}$$

We are assuming small wall velocities, so the particles' speeds change very little relative to their initial speed $\sqrt{2E'/m}$ over the B bounces. Therefore, to leading order in ε, we have

$$\Delta t_b = \sqrt{\frac{m}{2E'}} l_b, \tag{2.48}$$

where l_b denotes a particles b^{th} free flight distance after t'. We now define the coarse grained free flight distance, $\bar{l}(t', B)$ by time averaging the ensemble average of l_b over the first B bounces after t':

$$\bar{l}(t'; B) = \sum^B \frac{\langle l_b \rangle}{B} \tag{2.49}$$

26

Substituting Eqs. (2.49) and (2.48) into Eq. (2.47) gives

$$\Delta t = B \, \bar{l}(t'; B) \sqrt{\frac{m}{2E'}}, \tag{2.50}$$

and substituting for B in Eq. (2.46) gives

$$\left\langle E_B - E' \right\rangle = \Delta t \, \frac{4\sqrt{2m} \, \overline{u^2}(t'; B)}{\bar{l}(t'; B)} E'^{\frac{1}{2}}. \tag{2.51}$$

Equation (2.51) gives the ensemble averaged change in energy over the time Δt after t' for an ensemble of particles with energy E'. Comparing to Eq. (2.43), we see that dividing both sides of Eq. (2.51) by Δt gives us $g_1(E', t')$. We thus have,

$$g_1(E, t) = \frac{4\sqrt{2m} \, \overline{u^2}(t)}{\bar{l}(t)} E^{\frac{1}{2}}, \tag{2.52}$$

where we have switched from primed to unprimed variables, and the dependence on B has been suppressed.

The diffusion term $g_2(E', t')$ is defined as

$$g_2(E', t') \underset{\Delta t \to 0}{=} \frac{\left\langle (E(t' + \Delta t) - E(t'))^2 \right\rangle}{\Delta t}, \tag{2.53}$$

where $E(t') = E'$ for all particles in the ensemble, and $\langle E(t' + \Delta t) \rangle$ is the ensemble averaged particle energy at time $t' + \Delta t$. An expression for the diffusion term can be found by employing similar methods used to find the drift term. Alternatively, $g_2(E, t)$ can be found by invoking Liouville's theorem, as in Ref. [51]. Combing Liouville's theorem and the Fokker-Planck equation allows one to deduce a fluctuation-dissipation relation:

$$g_1(E, t) = \frac{1}{2\Sigma(E)} \partial_E \left[\Sigma(E) g_2(E, t) \right], \tag{2.54}$$

where $\Sigma(E)$ is the microcanonical partition function of a single particle with energy E in the corresponding frozen billiard. In a d dimensional billiard, the microcanonical partition

function is given by [49]

$$\Sigma(E) = \frac{1}{2}V_d\,\Omega_d\,(2m)^{\frac{d}{2}}\,E^{\frac{d}{2}-1},\tag{2.55}$$

where Ω_d is the d-dimensional solid angle, and V_d is the d-dimensional billiard's volume. Combining Eqs. (2.52),(2.54), and (2.55), we find

$$g_2(E,t) = \frac{4}{d+1}\frac{4\sqrt{2m}\,\overline{u^2}(t)}{\bar{l}(t)}E^{\frac{3}{2}}.\tag{2.56}$$

This method of determining g_2 allows for an additive constant, but this constant must be identically zero; when $E = 0$, the particles are motionless and there can be no drift or diffusion of energies, so we must have $g_1(0,t) = g_2(0,t) = 0$.

With our expressions for g_1 and g_2, we may rewrite the Fokker-Planck equation:

$$\partial_t\eta(E,t) = \frac{2\alpha(t)}{d+1}\partial_E\left[E^{\frac{1+d}{2}}\partial_E\left(E^{\frac{2-d}{2}}\eta(E,t)\right)\right]\tag{2.57}$$

where we define $\alpha(t)$ as

$$\alpha(t) \equiv \frac{4\sqrt{2m}\,\overline{u^2}(t)}{\bar{l}(t)}.\tag{2.58}$$

Equation (2.57) can be simplified by defining a rescaled time s:

$$s = \int_{t_0}^{t}\mathrm{d}t'\alpha(t'),\tag{2.59}$$

which gives

$$\partial_s\eta(E,s) = \frac{2}{d+1}\partial_E\left[E^{\frac{1+d}{2}}\partial_E\left(E^{\frac{2-d}{2}}\eta(E,s)\right)\right]\tag{2.60}$$

Equation (2.60) can be solved by separation of variables. We assume a solution of the form $\phi(s)f(E)$, and upon making the substitutions $F(E) = E^{\frac{3-d}{4}}f(E)$ and $z = E^{\frac{1}{4}}$ one finds a first order homogeneous linear differential equation for $\phi(s)$ and a Bessel equation of order $d-1$ for $F(z)$. The details of the separation of variables, including existence, uniqueness, and boundary conditions, are given in Ref. [53] and will be omitted here. We also acknowledge a similar, much older, one-dimensional solution given in Ref. [42]. The

separation of variables solution is

$$\eta(E, s) = E^{\frac{d-3}{4}} \int_0^\infty \mathrm{d}k\, A(k) J_{d-1}(kE^{\frac{1}{4}}) e^{-\frac{sk^2}{8(d+1)}}, \tag{2.61}$$

where J_{d-1} is an ordinary Bessel function of order $d-1$, and the amplitudes $A(k)$ are found by taking a Hankel transform of the initial ensemble $\eta(E, 0)$. When the ensemble begins in the microcanonical distribution with energy E_0, we have $\eta(E, 0) = \delta(E - E_0)$, and a closed form expression for $A(k)$ results. The energy distribution $\eta(E, s)$, subject to $\eta(E, 0) = \delta(E - E_0)$, is then

$$\eta(E, s) = \frac{1}{4E_0^{\frac{1}{2}}} \left(\frac{E}{E_0}\right)^{\frac{d-3}{4}} \int_0^\infty \mathrm{d}k\, k J_{d-1}(kE_0^{\frac{1}{4}}) J_{d-1}(kE^{\frac{1}{4}}) e^{-\frac{sk^2}{8(d+1)}}. \tag{2.62}$$

Making use of an identity of Bessel integrals utilized in Eq. (22) of Ref. [53], we can solve the integral in Eq. (2.62) and simplify the expression to

$$\eta(E, s) = \frac{d+1}{sE_0^{\frac{1}{2}}} \left(\frac{E}{E_0}\right)^{\frac{d-3}{4}} I_{d-1} \left[\frac{4(d+1)}{s} E_0^{\frac{1}{4}} E^{\frac{1}{4}}\right] e^{-\frac{2(d+1)}{s}\left(E_0^{\frac{1}{2}} + E^{\frac{1}{2}}\right)}, \tag{2.63}$$

where I_{d-1} is a modified Bessel function of order $d-1$. Using this energy distribution, we can find the ensemble averaged energy as a function of time:

$$\langle E(s) \rangle = \frac{d}{d+1} \frac{s^2}{4} + \sqrt{E_0}\, s + E_0. \tag{2.64}$$

Equation (2.63) is only valid under the assumption $\varepsilon \ll 1$. If we begin with an ensemble where ε is order unity or larger, over sufficiently long time, the slow particles inevitably gain so much energy that the fast particle assumption holds and Eq. (2.63) becomes valid asymptotically. We can thus find a universally valid asymptotic energy distribution by considering Eq. (2.62) or Eq. (2.63) in the limit of very large s. Specifically, if $k \ll d/\sqrt{E_0}$ for all $k^2 \gg 8(d+1)/s$, which implies that $s \gg 8\sqrt{E_0}(d+1)/d$, one can approximate $J_{d-1}(kE_0^{\frac{1}{4}})$ by the lowest order term in its Taylor expansion over the non-negligible contributions to

the integral in Eq. (2.62), and the solution reduces to

$$\eta_a(E, s) = \frac{1}{2E\Gamma(d)} \left[\frac{2(d+1)}{s} E^{\frac{1}{2}} \right]^d e^{-\frac{2(d+1)}{s} E^{\frac{1}{2}}}, \tag{2.65}$$

where Γ is the gamma function. One can easily verify that $\eta_a(E, s)$ is normalized and obeys the Fokker-Planck equation. Using the asymptotic energy distribution Eq. (2.65), we find the ensemble averaged energy at a large times to be

$$\langle E(s) \rangle_a = \frac{d}{1+d} \frac{s^2}{4}. \tag{2.66}$$

The results of this section are summarized as follows. In the quivering limit, correlations in particle energy decay over the time scale of a single collision, and as a result, the energy distribution of an ensemble evolves diffusively, regardless of the shape and dimensionality of the billiard boundary. Ensembles universally evolve to the asymptotic energy distribution given in Eq. (2.65), and ensemble averaged energy asymptotically grows quadratically in time. Before discussing the implications and broader context of these results, we comment on the interpretations of the coarse grained quantities $\bar{l}$ and $\overline{u^2}$.

If the particular billiard shape is ergodic, then their exists a characteristic ergodic time scale over which ensembles uniformly explore the entire billiard boundary. Invoking ergodicity and replacing time averages with phase space averages, we deduce that, over time scales greater than the ergodic time scale, $\bar{l}$ will be the billiard's mean free path, and $\overline{u^2}$ will be the second wall moment $M_2(\mathbf{q})$ uniformly averaged over the billiard boundary. This implies that, over time scales greater than the ergodic time scale, g_1 and g_2 are time-independent and that α is merely a constant. In this case, the expression for g_1 in Eq. (2.52) is equivalent to the wall formula, which was originally used to model energy dissipation from collective to microscopic degrees of freedom in nuclear dynamics [40]. In non-ergodic billiards, or over time scales shorter than the ergodic time scale in ergodic billiards, $\bar{l}$ and $\overline{u^2}$ will generally be time-dependent and can not be interpreted in terms of properties of the billiard shape alone. Nevertheless, they are still well-defined properties of the ensemble; $\bar{l}$ is simply the ensemble's average free flight distance over the coarse grained time scale, and $\overline{u^2}$ is the

average squared wall velocity for the collisions taking place over the coarse grained time scale.

2.4 Discussion

2.4.1 Approximate Quivering

The quivering limit is most certainly an idealization of time-dependent billiard motion; no real billiard boundary can actually move with zero amplitude and period. However, if the idealized system is defined in a physically consistent manner, then we expect that for smaller and smaller a and τ, real time-dependent billiards will be better and better approximated by quivering billiards. We now clarify how small a and τ must actually be for a time-dependent billiard to be well-approximated by a quivering billiard.

In Refs. [38] and [39], Lieberman, Lichtenberg, and Cohen studied the Fermi-Ulam model numerically and analytically using dynamical systems theory. It was shown that the energy evolution of a particle in the Fermi-Ulam model is generically diffusive and can be described by a Fokker-Planck equation for particle speeds such that, using our notation from Sec. 2.2.1, $v \ll u_c\sqrt{L/a}$. The value $u_c\sqrt{L/a}$ is associated with the stability of periodic orbits in v-Ψ space, where v and Ψ are the particle velocity and wall phase during collisions, respectively. At particle speeds much below $u_c\sqrt{L/a}$, Refs. [38] and [39] show that periodic orbits in v-Ψ space are unstable, dynamical correlations are small, and trajectories in v-Ψ space are generally chaotic (the language of the day labelled such trajectories stochastic as opposed to chaotic). At particle speeds above $u_c\sqrt{L/a}$, periodic orbits begin to stabilize, correlations become important, and the presence of elliptic islands and invariant spanning curves inhibit energy growth [38] [39]. In a one-dimensional quivering billiard, correlations vanish, trajectories are stochastic, and particle energy evolves diffusively, so, based on Lieberman, Lichtenberg, and Cohen's work, we see that a quivering billiard is a good description of the Fermi-Ulam model when $v \ll u_c\sqrt{L/a}$. As a becomes smaller and smaller with u_c held fixed, elliptic islands and invariant spanning curves move away to regions of larger and larger particle speeds, correlations become smaller and smaller due

to the more and more erratic wall motion, and quivering becomes a valid approximation for wider and wider ranges of particle speeds. As a approaches zero in the idealized limit, the infinitely erratic wall motion destroys correlations, elliptic islands and spanning curves occur only at infinite energy, and quivering becomes an exact description for all particle speeds. The same reasoning can be applied to higher dimensional time-dependent billiards; as a becomes smaller and smaller with u_c held constant, correlations become smaller and smaller and non-diffusive dynamics occur at higher and higher energies. We thus claim that when $v \ll u_c \sqrt{l_c/a}$ for all possible particle speeds v that could be observed in a simulation or experiment, where l_c is a characteristic free-flight distance, an arbitrary-dimensional time-dependent billiard will be approximately a quivering billiard.

Due to the inevitable increase in particle energy, the speed bound inequality $v \ll u_c \sqrt{l_c/a}$ implies that quivering will closely approximate a real billiard simulation or experiment only up to some maximum time t_{max}. The value of t_{max} depends on the particles' initial energy distribution, but we can estimate its scaling behavior in situations where the actual energy distribution is able to evolve the asymptotic distribution given in Eq. (2.65). In such cases, the average particle speed at large times can be estimated from the asymptotic ensemble averaged energy given by Eq. (2.66), and we find $v \sim t \, u_c^2/l_c$. Substituting this estimate for v into the speed bound inequality yields $t \ll (l_c/a)^{1/2} \, (l_c/u_c) = (l_c/a)^{3/2}\tau$. We thus have $t_{max} \sim (l_c/a)^{1/2} \, (l_c/u_c) = (l_c/a)^{3/2}\tau$. As expected, in the quivering limit, t_{max} diverges.

2.4.2 Consistency

Quivering wall motion corresponds to volume preserving billiard motion with negligible correlations in particles' energy changes. Therefore, if the quivering limit is actually physically meaningful, then the results obtained in Sec. 2.3 should agree with previous time-dependent billiard literature for the special case of volume preserving billiard motion with negligible correlations in energy changes. We now highlight three such examples.

In Ref. [39], $\langle \Delta E \rangle$ and $\langle (\Delta E)^2 \rangle$ are calculated for a single collision in the Fermi-Ulam model, assuming periodic wall motion (which corresponds to volume preserving billiard mo-

tion on average) and no correlations in the wall velocity between collisions. The authors also assume, without explicitly stating, that the wall velocity is an even function of time. The expressions obtained in Ref. [39] are in fact identical to our expressions for $\{\Delta E_b\}_b$ in Eq. (2.21) and $\{(\Delta E_b)^2\}_b$, which can be found by adding $\{\Delta E_b\}_b^2$ to Eq. (2.22), under the assumption that all odd moments of the wall velocity M_{2n+1} vanish. The odd moments vanish in a quivering billiard when we take the quivering limit of wall motion defined by an even function of time, so our results agree perfectly with those of Ref. [39].

Reference [49] studies the energy evolution of ensembles of independent particles in chaotic adiabatic billiards in two and three dimensions. A Fokker-Planck equation to describe the evolution of the energy distribution is proposed, and expressions for the corresponding drift and diffusion coefficients are derived. These results are obtained for general adiabatic billiard motion, under the assumption that correlations in a particle's energy changes decay over the mixing time scales corresponding to the frozen chaotic billiard shapes. The expressions for g_1 and g_2 are given in terms of a diffusion constant D, and an explicit expression for D is given using the *quasilinear approximation* - the assumption that energy changes between bounces are completely uncorrelated. Under the quasilinear approximation, assuming volume preserving billiard motion, the expressions for g_1 and g_2 in Ref. [49] are identical to our two and three-dimensional expressions for g_1 and g_2 in Eqs. (2.52) and (2.56), respectively, for ergodic billiards, over time scales greater than the ergodic time scale. Our results are thus consistent with those of Ref. [49]. It is remarked in Ref. [49] that it is not precisely clear under what conditions the quasilinear approximation will be valid for time-dependent billiards in general, but roughly speaking, the approximation requires the billiard shapes and motion to be "sufficiently irregular." Our results help clarify this issue; the quasilinear approximation is justified when a time-dependent billiard is approximately quivering, and the quasilinear approximation is in fact exact, not an approximation, in the quivering limit.

In Ref. [40], it is shown that the velocity distribution for independent particles in a time-dependent irregular container is asymptotically universally an exponential. This work assumes an isotropic velocity distribution, volume preserving billiard motion, and a three-dimensional billiard. If we assume an isotropic velocity distribution in a quivering billiard,

we can change variables from energy to velocity in Eq. (2.65), and we find the asymptotic velocity distribution $f_a(\mathbf{v}, s)$ in arbitrary dimensions

$$f_a(\mathbf{v}, s) = \frac{1}{\Omega_d \Gamma(d)} \left(\frac{2(d+1)}{s} \sqrt{\frac{m}{2}} \right)^d e^{-\frac{2(d+1)}{s}||\mathbf{v}||}. \tag{2.67}$$

In agreement with Ref. [40], the isotropic velocity distribution in a quivering billiard is universally an exponential in all dimensions. For a three-dimensional chaotic quivering billiard, where $s = \alpha t$ and chaotic mixing ensures an isotropic velocity distribution, Eq. (2.67) is identical to the velocity distribution obtained in Ref. [40].

2.4.3 Fermi acceleration

Equation (2.66) shows that the ensemble averaged growth is unbounded, increasing quadratically in time. Unbounded average energy growth in time-dependent billiards is known as Fermi acceleration. Fermi acceleration was originally proposed by Fermi as the mechanism by which cosmic rays gain enormous energies through reflections off of moving magnetic fields [41], and since become an active field of research in its own right. The current research generally seeks to determine under what conditions time-dependent billiards allow for Fermi acceleration, and to understand how the dynamics of sequence of frozen billiard shapes affects the energy growth rate. In Refs. [42, 37, 38, 39], it was established that sufficiently smooth wall motion in the one-dimensional Fermi-Ulam model prohibits Fermi acceleration, and that non-smooth wall motion allows for Fermi acceleration that may be much slower than quadratic in time. While the one-dimensional billiard is always integrable, higher dimensional billiards allow for integrable, pseudo-integrable, chaotic, or mixed dynamics. In Ref. [54], it was conjectured that fully chaotic frozen billiard shapes are a sufficient condition for Fermi acceleration in multi-dimensional time-dependent billiards, and the energy growth rate in such billiards was thought to be quadratic in time [54, 45]. It has since been shown that the problem is a bit more subtle; certain symmetries in the sequence frozen billiard shapes can prohibit or stunt the quadratic energy growth in chaotic billiards [46]. The problem is complicated for non-chaotic multi-dimensional billiards as well. Integrable billiards may prohibit [55] or allow [56] quadratic or slower Fermi accel-

eration, while exponential Fermi acceleration is possible for pseudo-integrable billiards [57] and billiards with multiple ergodic components [58, 59, 45, 60, 47] with possibly mixed or pseudo-integrable dynamics.

Given the complexities observed in the previous literature, our result in Eq. (2.66) is surprising; in the quivering limit, regardless of the dimensionality or underlying frozen dynamics, time-dependent billiards universally show quadratic Fermi acceleration. The apparent contradiction between our work and previous work is due to a difference in the limits studied. Both our work and the previous literature, because of the inevitable speed up of particles, analyze time-dependent billiards in the adiabatic limit, where the wall speed is much slower than the particle speed. In the previous literature, however, the period of billiard oscillations is typically fixed and non-zero (with numerical results often presented as a function of the oscillation amplitude), so in the adiabatic limit, the typical time between collisions is always much shorter than the billiard's oscillation period. In our work, the oscillation period approaches zero, so the time between collisions is always much larger than the oscillation period, even in the adiabatic limit where particles move much faster than walls.

2.4.4 Fixed wall simplifications

An alternative simplification similar to the quivering billiard has been frequently employed in the literature. The so-called static wall approximation (sometimes called the simplified Fermi-Ulam model) was originally introduced in [38] in order to ease the analytical and numerical study of the Fermi-Ulam model, and through the years has become a standard approximation assumed valid for small oscillation amplitudes, often studied entirely in lieu of the exact dynamics. See Ref. [38, 39, 61, 54, 62, 63, 64, 50] for example. Using the notation of Sec. 2.2, assuming $v \gg u_c$ so that we may ignore glancing collisions for the sake of simplicity, the dynamics of the one-dimensional Fermi-Ulam model can be

described by the deterministic map,

$$v_b \;=\; v_{b-1} - 2u(t_b), \tag{2.68a}$$

$$t_b \;=\; t_{b-1} + \frac{2L}{v_{b-1}} + \frac{g(t_b) + g(t_{b-1})}{v_{b-1}}, \tag{2.68b}$$

while the corresponding static wall approximation is given by the deterministic map,

$$v_b \;=\; v_{b-1} - 2u(t_b), \tag{2.69a}$$

$$t_b \;=\; t_{b-1} + \frac{2L}{v_{b-1}}. \tag{2.69b}$$

In the above maps, v_{b-1} is the particle's velocity just before the bth collision, and t_b is the time of the bth collision. An analogous static wall approximation can be constructed for higher dimensional billiards [61, 54, 64]. Like the quivering billiard, the static wall approximation eliminates the implicit equations for the time between collisions by holding the billiard boundary fixed. The two models differ because the static wall approximation assumes $u(t_b)$ to be a well behaved function. It is common practice to consider stochastic versions of the maps (2.68) and (2.69), where $u(t_b)$ is replaced by $u(t_b + \zeta)$ for some random variable ζ [38, 61, 54, 63, 64, 50]. The stochastic case simulates the effects of external noise on the system and allows one to average over ζ when determining ensemble averages, which often facilitates analytical calculations.

In Refs. [63, 64], Karlis et al. show that the stochastic static wall map and its analogue for the two-dimensional Lorentz gas give one half the asymptotic energy growth rate of the stochastic Fermi-Ulam map. This inconsistency exists even for small a, so Karlis et al. conclude that (2.69) is not a valid approximation of (2.68). We add that the same factor of two discrepancy can be observed between our quivering billiard expression for g_1 and the corresponding expressions obtained from the deterministic static wall maps given in [38, 39, 61, 54]. In an early study of the Fermi-Ulam model, Ref. [42] obtains a drift term that is actually in agreement with the static wall approximation value, but a careful reading reveals that the authors make a series of simplifications that inadvertently reduce their Fermi-Ulam model to the static wall approximation. Ref. [63] corrects for the energy

inconsistency to a high degree of accuracy in the stochastic case by introducing the hopping wall approximation. The hopping wall approximation assumes wall motion slow enough such that the moving wall's position at the bth bounce can be approximated by its position at the $(b-1)^{th}$ bounce, or by its position at the time of the particle's collision with the fixed wall just after the $(b-1)^{th}$ bounce. This approximation allows $g(t_b)$ in Eq. (2.68b) to replaced by either $g(t_{b-1})$ or $g(t_{b-1} + L/v_{b-1})$. An analogous hopping wall approximation for two dimensions is presented in [64]. Like the static wall approximation, the hopping wall approximation eliminates the implicit equations for the time between collisions, which eases numerical and analytical study. Based on the hopping wall approximation's more accurate asymptotic energy growth rate, Karlis et al. conclude in Refs. [63, 64] that the energy discrepancy between the Fermi-Ulam model and the static wall approximation is due to dynamical correlations induced by small changes in the free flight time between collisions which are neglected in the static wall approximation.

Based on the results of this paper, we propose an alternative explanation of the energy discrepancy. The energy discrepancy is observed because the static wall approximation is simply unphysical, and it can not accommodate for the fact that, due to the relative motion between the particles and walls, collisions with inward moving walls are more likely than collisions with outward moving walls. In fact, defining the quivering billiard *without* the flux factor in the biased distribution (so that the biased and unbiased distribution are equal) reproduces the asymptotic energy growth rate predicted by the stochastic static wall approximation. Evidently, the last term in Eq. (2.68b) is responsible for the bias towards inward moving wall collisions in the exact Fermi-Ulam model, and hopping wall approximation's estimate of this term is responsible for its more accurate energy growth rate. Although the static wall approximation is a mathematically well-defined dynamical system, it is an ill-posed physical system for the following reasons. If a billiard boundary is truly static such that (2.68b) somehow reduces to (2.69b), then we must have $a \to 0$. But if $a \to 0$, then $u_c \to 0$ and the billiard becomes trivially time-independent unless $\tau \to 0$ as well. However, if both a and $\tau \to 0$, then $u(t)$ can not be a well-behaved function as required by the definition of the static wall map, and, as argued in Sec. 2.2, the wall velocity becomes stochastic. This logic seems to be unavoidable; if the walls are to be genuinely fixed,

then physical consistency demands that the wall motion must be non-existent or stochastic. Based on this reasoning, we propose the following conjecture: any physically consistent, non-trivial, fixed wall limit of a time-dependent billiard must be physically equivalent to the quivering limit, and the corresponding quivering billiard as defined in this paper yields the correct dynamics and energy growth rate (by physically equivalent, we mean equivalent energy and velocity statistics). Of particular note, corrections to the free flight time between collisions are not needed to achieve the correct energy growth rate.

2.5 Examples and Numerics

We now give explicit examples of quivering billiards in one and two dimensions and support the previous sections' analyses with numerical work. Consider first a one dimensional Fermi-Ulam model with one wall oscillating at a constant speed. Following the notation of Sec. 2.2, the position of the moving wall about its mean position is given by

$$
g(t) = \begin{cases} a[-1 + 4\Psi(t)], & 0 \leq \Psi(t) < \frac{1}{2} \\ a[1 - 4(\Psi(t) - \frac{1}{2})], & \frac{1}{2} \leq \Psi(t) < 1, \end{cases} \tag{2.70}
$$

and the corresponding wall velocity is given by

$$
u(t) = \begin{cases} 4u_c, & 0 \leq \Psi(t) < \frac{1}{2} \\ -4u_c, & \frac{1}{2} \leq \Psi(t) < 1. \end{cases} \tag{2.71}
$$

The numerical analyses of this Fermi-Ulam model are presented in Figs. 2.3 and 2.4. The histograms in Fig. 2.3 show of the evolution of the energy distribution of 10^5 particles of mass $m = 1$ in a microcanonical ensemble with initial speed $v_0 = 1$ at time $t = 0$, and the curves show the analytical solution for this system in the quivering limit as predicted by Eq. (2.63). For this simulation, we set $L = 1.0$, $a = 10^{-5}$, and $\tau = 10^{-2}$, which gives $u_c = 10^{-3}$. We see good agreement, with some small deviation apparent beginning at $t = 5000$. We suspect that the deviation is due to the faster particles interacting with the elliptic islands in phase space, which is not accounted for in the quivering billiard. By
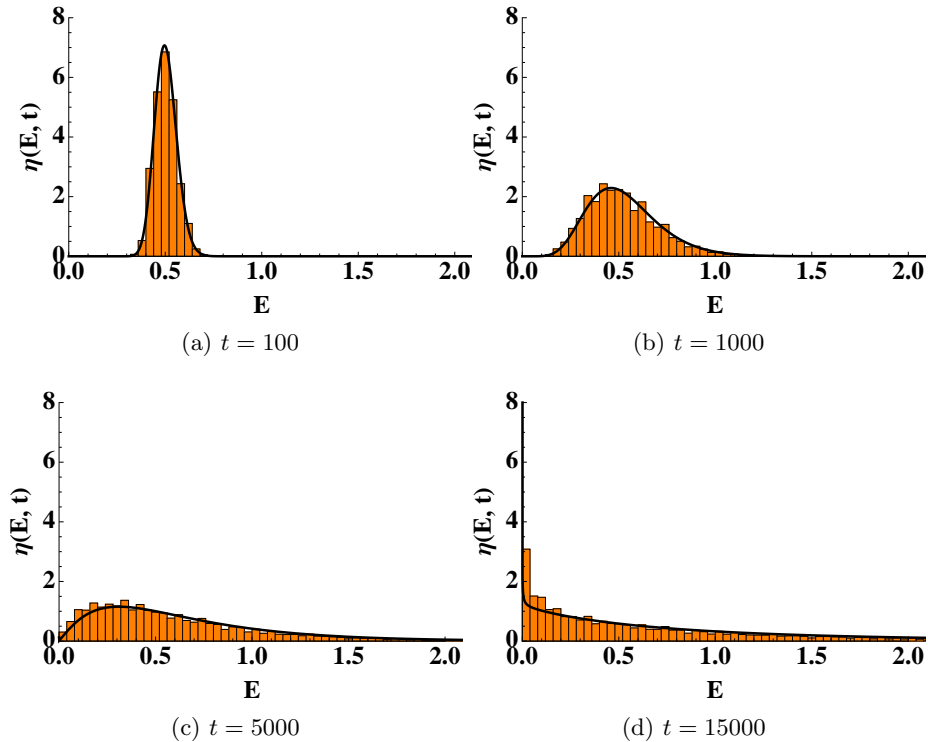
Figure 2.3: Energy distribution $\eta(E, t)$ of 10^5 particles following the exact Fermi-Ulam dynamics with small wall oscillation amplitude $a = 10^{-5}$ at times $t = 100$, $t = 1000$, $t = 5000$, and $t = 15000$. The histograms are generated from numerical simulations, and the smooth curve is the analytical solution Eq. (2.63) for the energy distribution of a particle ensemble in the corresponding quivering billiard.

the time $t = 15000$, a sufficient number of the particles have gained enough energy such that the system is no longer approximately quivering. Further energy gain is stunted by elliptic islands, so we see an excess of probability (an excess relative to the quivering billiard energy distribution) begin to build up at low energies. Figure 2.4 shows the same Fermi-Ulam model, with $u_c = 10^{-3}$, for successively smaller and smaller values of a and τ at time $t = 5000$. As a becomes smaller, we see the actual energy distribution converge to the distribution predicted by the quivering billiard.

The quivering limit of the Fermi-Ulam model given in Eqs. (2.70) and (2.71) is found by following the procedures described in Sec. 2.2. We first obtain the unbiased distribution,

$$P(u_b|0) = \frac{1}{2}\delta(u_b - 4u_c) + \frac{1}{2}\delta(u_b + 4u_c), \tag{2.72}$$
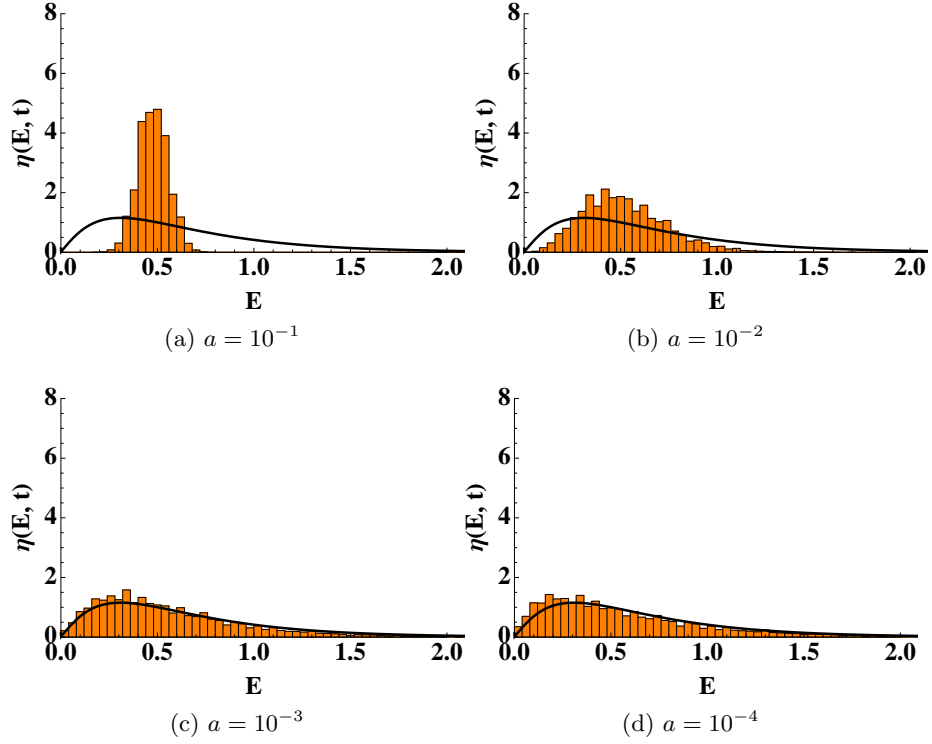
Figure 2.4: Energy distribution $\eta(E,t)$ at $t = 5000$ of 10^5 particles following the exact Fermi-Ulam dynamics for successively smaller wall oscillation amplitudes a. The histograms are generated from numerical simulations, and the smooth curve is the analytical solution Eq. (2.63) for the energy distribution of a particle ensemble in the corresponding quivering billiard. The case for $a = 10^{-5}$ is shown in the $t = 5000$ plot in Fig. 2.3

and then the biased distribution $P(u_b|v_{b-1})$,

$$
P(u_b|v_{b-1}) = \begin{cases} \frac{1}{2}\left(1 - \frac{u_b}{v_{b-1}}\right)\left[\delta(u_b - 4u_c) + \delta(u_b + 4u_c)\right], & v_{b-1} > 4u_c \\ \delta(u_b + 4u_c), & v_{b-1} \leq 4u_c. \end{cases} \tag{2.73}
$$

The drift and diffusion terms corresponding to this quivering billiard are found by following the procedures Sec. 2.3.4. We note that $M_2(\mathbf{q}_b) = 16u_c^2$ for the moving wall, and $M_2(\mathbf{q}_b) = 0$ for the stationary wall, so Eq. (2.45) yields $\overline{u^2} = (1/2)\,16u_c^2$. The coarse grained free flight distance is given simply by $\bar{l} = L$, so we find

$$
\begin{aligned}
g_1(E) &= \frac{32\sqrt{2m}\,u_c^2}{L}E^{\frac{1}{2}} = \alpha E^{\frac{1}{2}} \\
g_2(E) &= \frac{64\sqrt{2m}\,u_c^2}{L}E^{\frac{3}{2}} = 2\alpha E^{\frac{3}{2}}.
\end{aligned} \tag{2.74}
$$

40

(a) $t = 100$

(b) $t = 1000$

(c) $t = 5000$

(d) $t = 15000$

Figure 2.5: Energy distribution $\eta(E, t)$ of 10^5 particles at $t = 100$, $t = 1000$, $t = 5000$, and $t = 15000$ in a quivering billiard corresponding to the quivering limit of the Femi-Ulam model used in Fig. 2.3. The histograms are generated from numerical simulations, and the smooth curve is the analytical solution for the energy distribution given by Eq. (2.63).

The drift and diffusion terms are independent of time, so the rescaled time s is simply $s = \alpha t$. Using the same values for L, m, and u_c the we used in the Fermi-Ulam simulation, we find $\alpha \approx 4.53 \times 10^{-5}$. Figure 2.5 shows the evolving energy distribution in the simulated quivering billiard, with the analytical result predicted by Eq. (2.63) superimposed. Our analytical solution agrees very well with the numerical simulation.

For pedagogical purposes, we now construct and simulate a two-dimensional quivering billiard. For the billiard shape, we have chosen the six-circle clover introduced in Ref. [49], depicted here in Fig. 2.6. We set the normal wall velocities along the billiard boundary to be,

$$u(\mathbf{q}, t) = \begin{cases} u_c |\hat{\mathbf{n}}(\mathbf{q}) \cdot \hat{\mathbf{x}}|, & 0 \leq \Psi(t) < \frac{1}{2} \\ -u_c |\hat{\mathbf{n}}(\mathbf{q}) \cdot \hat{\mathbf{x}}|, & \frac{1}{2} \leq \Psi(t) < 1, \end{cases} \tag{2.75}$$

41
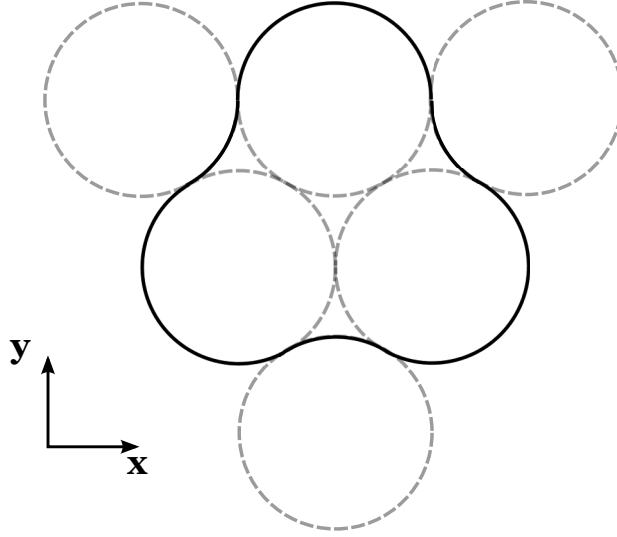
Figure 2.6: The six-circle clover billiard, constructed from sections of six adjacent equi-radii circles.
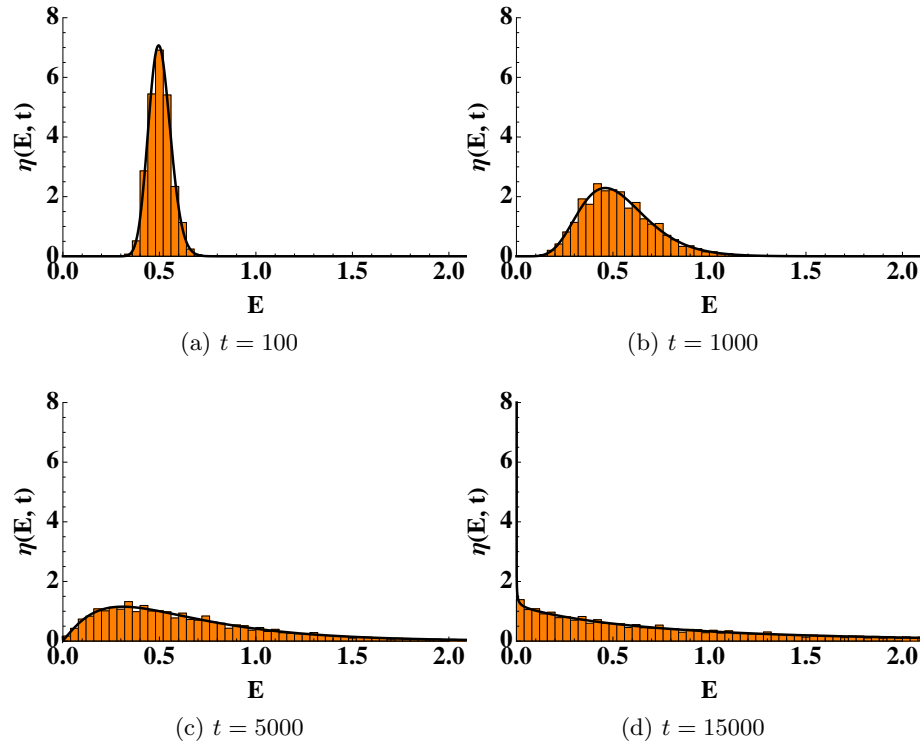

(a) $t = 100$

(b) $t = 1000$
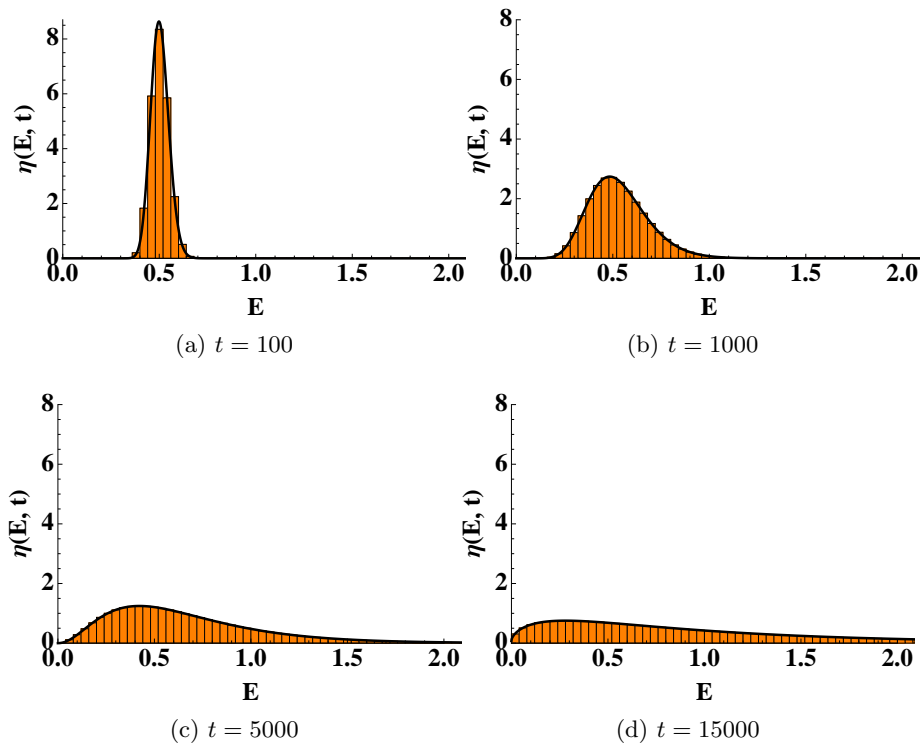
(c) $t = 5000$

(d) $t = 15000$

Figure 2.7: Energy distribution $\eta(E, t)$ of 10^5 particles at $t = 100$, $t = 1000$, $t = 5000$, and $t = 15000$ in a two-dimensional chaotic quivering billiard. The histograms are generated from numerical simulations, and the smooth curve is the analytical solution for the energy distribution given by Eq. (2.63).

where $\hat{\mathbf{n}}(\mathbf{q})$ is the outward unit normal to the wall at $\mathbf{q}$ and $\hat{\mathbf{x}}$ is the unit vector in the x-direction. This choice of wall velocities gives in the quivering limit,

$$P(u_b|0, \mathbf{q}_b) = \frac{1}{2}\delta(u_b - u_c|\hat{\mathbf{n}}(\mathbf{q}_b) \cdot \hat{\mathbf{x}}|) + \frac{1}{2}\delta(u_b + u_c|\hat{\mathbf{n}}(\mathbf{q}_b) \cdot \hat{\mathbf{x}}|) \tag{2.76}$$

$$P(u_b|v_{b-1}, \mathbf{q}_b, \theta_b) = \begin{cases} \left(1 - \frac{u_b}{v_{b-1}\sin(\theta_b)}\right) P(u_b|0, \mathbf{q}_b), & v_{b-1} > u_c|\hat{\mathbf{n}}(\mathbf{q}_b) \cdot \hat{\mathbf{x}}| \\ \delta(u_b + u_c|\hat{\mathbf{n}}(\mathbf{q}_b) \cdot \hat{\mathbf{x}}|), & v_{b-1} \leq u_c|\hat{\mathbf{n}}(\mathbf{q}_b) \cdot \hat{\mathbf{x}}|. \end{cases} \tag{2.77}$$

The six-circle clover constructed from equi-radii circles is fully chaotic [49], so over time scales greater than the clover's ergodic time scale, $\overline{u^2}$ is just $M_2(\mathbf{q})$ averaged uniformly over the billiard boundary. For any $\mathbf{q}$ on the boundary, we have $M_2(\mathbf{q}) = u_c^2 |\hat{\mathbf{n}}(\mathbf{q}) \cdot \hat{\mathbf{x}}|^2$, and from Fig. 2.6, we see the outward normals $\hat{\mathbf{n}}(\mathbf{q})$ are distributed uniformly around a unit circle, so we have $\overline{u^2} = (1/2)\, u_c^2$. The coarse grained free flight distance $\bar{l}$, over time scales greater than the ergodic time scale, is just the billiard's mean free path. For a two dimensional ergodic billiard, the mean free path is given by $\pi A/S$, where A is the billiard's area and S is the billiard's perimeter [49]. If the radius of the circles used to construct the six-circle clover is R, then the geometry of Fig. 2.6 gives $A = R^2(4\sqrt{3} + \pi)$ and $S = 4\pi R$. We thus have for the drift and diffusion coefficients,

$$\begin{aligned} g_1(E) &= \frac{2\sqrt{2m}\, u_c^2}{l}E^{\frac{1}{2}} = \alpha E^{\frac{1}{2}} \\ g_2(E) &= \frac{8\sqrt{2m}\, u_c^2}{3\, l}E^{\frac{3}{2}} = \frac{4}{3}\alpha E^{\frac{3}{2}}. \end{aligned} \tag{2.78}$$

where $l = R(\sqrt{3} + \pi/4)$ is the mean free path.

Figure 2.7 shows the energy evolution of a microcanonical ensemble of 10^5 particles in a quivering clover, with the distribution Eq. (2.63) superimposed. The particles have mass $m = 1$ and initial energy $E_0 = 1/2$. We constructed the clover with circles of radius $R = 1$ and set $u_c = 6.35 \times 10^{-3}$ to give $\alpha \approx 4.53 \times 10^{-5}$. Again, we see good agreement between the distribution predicted by Eq. (2.63) and the simulated energy distribution.

2.6 Summary and Conclusions

In this work, we have defined a particular fixed wall limit of time-dependent billiards, the quivering limit, and explored the evolution of particles and ensembles in the resulting quivering billiards. We have conjectured that any physically consistent, non-trivial, fixed wall limit of a time-dependent billiard must be physically equivalent to the quivering limit, and we have shown that the simplifications allowed by a physically consistent fixed wall limit come at a price: deterministic billiard dynamics become inherently stochastic. Although quivering is an idealized limit of billiard motion, we have shown that for smaller and smaller oscillation amplitudes and periods, time-dependent billiards become better and better approximated by quivering billiards. Billiards that quiver or approximately quiver behave universally; particle energy evolves diffusively, particle ensembles achieve a universal asymptotic energy distribution, and quadratic Fermi acceleration always occurs, regardless of a billiard's dimensionality or frozen dynamics. The mechanism for this quadratic Fermi acceleration is analogous to a resistive friction-like force, present due to the fluctuations induced by the erratic wall motion, as described by the fluctuation-dissipation relation in Eq. (2.54).

Through this work, we have gained some insight into issues that have been discussed in the previous literature. Namely, we concluded that in the quivering limit, the quasilinear approximation is exact, not an approximation. Also, we showed that the often used static wall approximation fails because it is unphysical and can not take into account the statistical bias towards inward moving wall collisions. Energy gain in the static wall approximation is a purely mixing effect; unbiased fluctuations in particle velocity produce an average increase in particle velocity squared, analogous to a Brownian random walk where unbiased fluctuations in position produce an average increase in squared distance from the initial position. From this observation, and the fact that the static wall approximation gives one half the asymptotic energy growth rate observed in exact systems, we conclude that in the quivering limit, half of the average energy gain observed in a time-dependent billiard is due to the mere presence of fluctuations, and half is due to the fact that energy gaining fluctuations are more likely than energy losing fluctuations.

We close by acknowledging that we have not given a rigorous mathematical proof showing that deterministic time-dependent billiards become stochastic quivering billiards in the quivering limit. One possible approach toward such a proof would be to define some sort of space of time-dependent billiards consisting of systems with different oscillation amplitudes and periods, define a metric to give some notion of distance in this space, and prove that particular sequences in this space with successively smaller amplitudes and periods are Cauchy sequences. One could then determine what properties the space of systems would need to posses in order to assure that these Cauchy sequences converge to limits, and then study the limits by studying the sequences that converge to them. Instead of a rigorous mathematical approach, we have taken a more intuitive approach and have attempted to justify our work by using physical reasoning and by showing consistency with previous results. We hope that the evidence is convincing enough to mitigate our mathematical deficiencies.

Chapter 3

Thermodynamics in a Chaos Bath: The Fokker-Planck Equation

3.1 Introduction

This chapter is the first of two which explore the microscopic Hamiltonian foundations of equilibrium and non-equilibrium thermodynamics. We will model a system of interest and its surrounding environment as a single closed classical Hamiltonian system, and we will determine the manner in which deterministic Hamiltonian chaos at the microscale produces effectively stochastic thermodynamic-like behavior at the mesoscale. Our work is preceded by a long history of research, dating back to Maxwell, Boltzmann, and Gibbs. See Refs. [65] and [66] for historical accounts. While the microscopic Hamiltonian foundations of equilibrium thermodynamics is firmly rooted in equilibrium statistical mechanics [67], non-equilibrium thermodynamics still lacks a firm microscopic foundation, due largely to intractable mathematics and the conceptual difficulty of reconciling the reversibility of Hamilton's equations with the irreversibility implied by the second law of thermodynamics and the inevitable increase of entropy [66]. These issues have been addressed with some success in more recent years using simplified microscopic frameworks such as thermostated dynamics, dynamical mappings, and stochastic dynamics - such methods produce consistent and physically meaningful results, but are they are described by non-Hamiltonian and/or
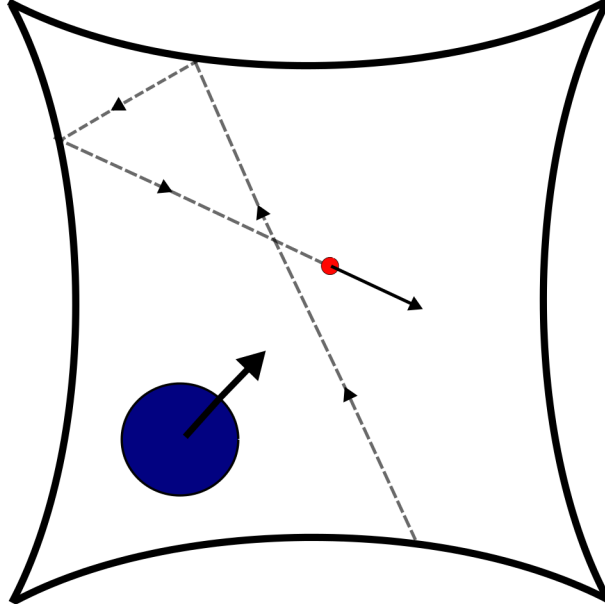
Figure 3.1: A two dimensional chaotically shaped billiard containing a point particle colored red and a particle of non-zero radius colored in blue. The point particle is considered fast and light, while the finite sized particle is considered slow and heavy.

discrete time dynamics and thus lack a physical, first-principles foundation [65, 66]. In this chapter and the next, we will "derive" thermodynamics from first principles in closed system Hamiltonian framework, and we will describe fluctuations, dissipation, irreversibility, and the notion of thermal equilibrium at the mesoscale in terms of the underlying microscopic Hamiltonian dynamics. In Chap. 3, we establish our framework and focus on equilibrium thermodynamics. In Chap. 4, we will use this framework to show that fluctuation theorems [13, 15, 17, 18, 19, 20] follow somewhat trivially, essentially by definition, from the very Hamiltonian reversibility symmetries which have caused over 100 years of difficulty and controversy in establishing the microscopic foundations of thermodynamics.

The work in this chapter is essentially a generalization and in depth exploration of Jarzynski's results in Ref. [68]. In Ref. [68], Jarzynski shows shows that when a slow heavy Hamiltonian system interacts with a fast, light, chaotic Hamiltonitan system, the slow degrees of freedom evolve diffusively and effectively "thermalize" to an equilibrium state with the fast degrees of freedom, and so Jarzynski concludes that the slow degrees of freedom behave analogously to a Brownian particle while the fast chaotic degrees of freedom behave analogously to an arbitrary (possibly low) dimensional thermal bath. We will therefore

refer to the fast chaotic Hamiltonian system as a "chaos bath." A simple example of such a system is depicted in Fig. 3.1. This system consists of a chaotic billiard which contains a light point particle and a heavy finite sized particle interacting through elastic collisions. Note that the heavy particle modifies the shape of the billiard by effectively punching a hole through its interior, so in order for Fig. 3.1 to be a valid example, the billiard shape must be chaotic for all heavy particle positions. As shown by Sinai in Ref. [69], this is indeed the case for the billiard pictured in Fig. 3.1.

The formalism established in Ref. [68] is valid only when the chaos bath is ergodic and mixing in such a manner that, when the slow degrees of freedom are "frozen" in place, the bath degrees of freedom chaotically explore the entirety of a fixed bounded energy shell in the bath's phase space. This requirement of ergodicity on an energy shell is somewhat restrictive in regards to thermodynamics: ergodicity on an energy shell requires all of the bath degrees of freedom to be mutually interacting, so the chaos bath formalism established in Ref. [68] is incapable of describing a multi-particle chaotic ideal gas. In a chaotic ideal gas, each gas particle is itself an independent chaos bath, so the joint bath state is unable to explore the entirety of an energy shell in the joint bath phase space. This is undesirable from a foundational standpoint - the multi-particle ideal gas ia a simple and physically relevant paradigmatic system which elucidates the relationship between statistical mechanics and thermodynamics, and from a multi-particle ideal gas, one can derive Fermi-Dirac, Bose-Einstein, and Maxwell-Boltzmann statistics [70]. Although a single chaos bath can not describe a multi-particle chaotic ideal gas, according to the Sinai-Boltzmann hypothesis [71], a single chaos bath can describe another paradigmatic thermodynamic system: the multi-particle chaotic dilute hard-sphere gas [66]. Because we are interested in the microscopic foundations of thermodynamics, we demand a formalism capable of describing both of these important thermodynamic systems, so in this chapter, we will extend the chaos bath formalism to allow for an arbitrary number of non-interacting baths. The extension to multiple baths is non-trivial; the non-interacting chaos baths effectively interact with each other through their effects on the evolution of the system of interest, so it is not a priori obvious whether or not an effective coupling between the baths will emerge in the mesoscale equations.

We acknowledge some previous older works which derive diffusive motion for a Brownian particle from a closed system deterministic Hamiltonian perspective [72, 73]. In these works, the bath Hamiltonian and interaction Hamiltonian are assumed to take a particular form, and diffusive Brownian motion only arises in the limit of an infinite number of bath particles, while in our work, we allow for arbitrary bath Hamiltonians, and diffusive motion arises for finite sized baths. Furthermore, in these older works, no connections to the first or second law of thermodynamics are presented, while the purpose of our work is to make these very connections. A formalism similar to ours is presented by Maes and Tasaki in Ref. [74], where a chaos bath serves as set of fast internal molecular degrees of freedom, and the corresponding center of mass serves as a set of slow heavy degrees of freedom. Maes and Tasaki discusses the applicability of the second law of thermodynamics to this system, but the presence of an external thermal environment is assumed, so the results do not derive from a first principles closed system deterministic Hamiltonian framework. In our work, the bath itself serves as the environment, and the slow and fast variables together comprise a closed system.

Our methods borrow heavily from a history of research related to ergodic adiabatic Hamiltonian systems, dating back to Ott's observation of the diffusive growth of errors to the conservation of the ergodic adiabatic invariant for slowly time-dependent ergodic Hamiltonian systems whose time-dependence is finitely slow [75]. Using Ott's result, Wilkinson [48] (and later independently Jarzynski [51]) showed that in addition to the reversible adiabatic force associated with infinitely slow time-dependence, a finitely slow time-dependent ergodic Hamiltonian system will experience an irreversible diffusive increase in energy, as if the agent controlling the time-dependence were subject to a dissipative friction-like force (dubbed "deterministic friction" by Wilkinson) from the Hamiltonian system. Expanding on these ideas, Berry and Robbins, in a series of papers, considered time-dependent ergodic fast Hamiltonian systems where the time-dependent parameters are actual physical degrees of freedom which evolve in time according their own slow Hamiltonian dynamics [76, 77, 78]. Using this framework, one can examine not only the effects of the time-dependence on the fast ergodic Hamiltonian system, but also the corresponding back reactions on the slow Hamiltonian system, which manifest themselves in the form of the so-called "adiabatic

reaction forces." As hinted at by Wilkinson, Berry and Robins showed that the slow Hamiltonian system is indeed subject to a genuine frictional force, as well as a magnetic like force (dubbed "geometric magnetism") which is a classical manifestation of the geometric phase associated with time-dependent adiabatic quantum systems [76, 79]. Later, in Ref. [68], Jarzynski added the concept of a rapidly fluctuating force to the framework of Berry and Robbins, and this facilitated connections between the ergodic Hamiltonian framework and thermodynamics (Brownian motion in particular).

The purpose of this chapter is to derive an effective mesoscale evolution equation for the probability distribution of a set of slow canonical position and momenta variables in the presence of multiple Hamiltonian chaos baths, and to determine to what extent this framework is able to reproduce classical equilibrium thermodynamics. Non-equilibrium thermodynamics will be addressed in chapter 4. Our work will proceed as follows. In Sec. 3.2, we establish our Hamiltonian framework and introduce some notation and identities. We then derive our central result, a Fokker-Planck diffusion equation, in Sec. 3.3. The derivation is lengthy and is broken down into six subsections. In Sec. 3.4, we find the stationary solution to the Fokker-Planck equation and clarify its relation to the notion of thermal equilibrium. We conclude in Sec. 3.5.

3.2 Hamiltonian Framework

Consider $N+1$ classical systems which evolve under Hamilton's equations: "the system of interest" and N "baths." The system of interest and baths together comprise "the universe." We denote the canonical variables of the system of interest and the αth bath by $(\mathbf{Q}, \mathbf{P})$ and $(\mathbf{q}^\alpha, \mathbf{p}^\alpha)$, respectively, where $\alpha \in \{1, 2, ..., N\}$. Throughout this work, Greek superscripts will index the baths, and Roman subscripts will index components of vectors and matrices, as well as orders of series expansions. For notational simplicity, we will often abbreviate the phase space coordinates by $\mathbf{X} = (\mathbf{Q}, \mathbf{P})$, and $\mathbf{x}^\alpha = (\mathbf{q}^\alpha, \mathbf{p}^\alpha)$. We also abbreviate the joint phase space coordinates of the baths by $\mathbb{X} = (\mathbf{x}^1, ..., \mathbf{x}^N)$, and abbreviate the joint phase space coordinates of the entire system plus baths universe by $(\mathbf{X}, \mathbb{X}) = (\mathbf{X}, \mathbf{x}^1, ..., \mathbf{x}^N)$. Assume a Hamiltonian of the form

$$\mathcal{H}(\mathbf{Q}, \mathbf{q}^1, ..., \mathbf{q}^N, \mathbf{P}, \mathbf{p}^1, ..., \mathbf{p}^N, \boldsymbol{\lambda}(t)) = H(\mathbf{Q}, \mathbf{P}, \boldsymbol{\lambda}(t)) + \sum_{\alpha=1}^{N} h^\alpha(\mathbf{q}^\alpha, \mathbf{p}^\alpha; \mathbf{Q}), \qquad (3.1)$$

where H is the system of interest's Hamiltonian, h^α is the αth bath's Hamiltonian (which includes the interaction between the system of interest and the αth bath), and $\boldsymbol{\lambda}(t)$ is a vector of time-dependent control parameters which can be manipulated by some external agent. For the moment, we are not interested in the questions and subtleties associated with controlling physical systems - we simply wish to acknowledge that the formalism presented here allows for a time-dependent Hamiltonian. We defer discussion of control to the next chapter of this thesis, where we study thermodynamics in a chaos bath at the level of individual trajectories.

The αth bath trajectory $\mathbf{x}^\alpha(t)$ evolves and interacts with the system of interest under $h^\alpha(\mathbf{x}^\alpha; \mathbf{Q})$, and the system of interest's Hamiltonian $H(\mathbf{X}, \boldsymbol{\lambda}(t))$ is assumed to be of the form

$$H(\mathbf{X}, \boldsymbol{\lambda}(t)) = \frac{1}{2}\mathbf{P}^T M^{-1} \mathbf{P} + V(\mathbf{Q}, \boldsymbol{\lambda}(t)), \qquad (3.2)$$

where M denotes a diagonal matrix of inertias (i.e $M_{ij} = M_i \delta_{ij}$), and $V(\mathbf{Q}, \boldsymbol{\lambda}(t))$ denotes the system of interest's potential energy. In our notation, the phase space points $\mathbf{X}$ and $\mathbf{x}^\alpha$ are viewed as column vectors, and the superscript T denotes the matrix transpose. We note that the energy associated with the interaction between the system of interest and the αth bath is considered to be part of the αth bath's energy. Applying Hamilton's equations of motion gives

$$
\begin{aligned}
\dot{\mathbf{Q}}(t) &= M^{-1}\mathbf{P}(t) & (3.3)\\
\dot{\mathbf{P}}(t) &= -\partial_{\mathbf{Q}} V(\mathbf{Q}, \boldsymbol{\lambda})|_{\mathbf{Q}(t), \boldsymbol{\lambda}(t)} - \sum_\alpha \partial_{\mathbf{Q}} h^\alpha(\mathbf{x}^\alpha; \mathbf{Q})|_{\mathbf{x}^\alpha(t), \mathbf{Q}(t)}\\
\dot{\mathbf{q}}^\alpha(t) &= \partial_{\mathbf{p}^\alpha} h^\alpha(\mathbf{x}^\alpha; \mathbf{Q})|_{\mathbf{x}^\alpha(t), \mathbf{Q}(t)}\\
\dot{\mathbf{p}}^\alpha(t) &= -\partial_{\mathbf{q}^\alpha} h^\alpha(\mathbf{x}^\alpha; \mathbf{Q})|_{\mathbf{x}^\alpha(t), \mathbf{Q}(t)} .
\end{aligned}
$$

The system of interest and the N baths together comprise a single closed system, and we now consider a statistical ensemble of such systems. Each member of this ensemble represents one realization of the dynamics in Eq. (3.3) generated from some distribution of initial conditions in the full system plus baths phase space. The quantity $\phi(\mathbf{X}, \mathbb{X}, t)$ will denote the distribution of these ensemble members over the full phase space at time t. In order to assure that ϕ vanishes at infinity, we assume that the surfaces of constant $\mathcal{H}$ in the full system plus baths phase space are bounded. Applying the conservation of ensemble members, the divergence theorem, and Eq. (3.3) leads to Liouville's equation for the time evolution of ϕ [65]:

$$\partial_t \phi(\mathbf{X}, \mathbb{X}, t) = -\{\phi(\mathbf{X}, \mathbb{X}, t), \mathcal{H}(\mathbf{X}, \mathbb{X}, \boldsymbol{\lambda}(t))\}_{\mathbf{X}, \mathbb{X}}, \tag{3.4}$$

where $\{f(\mathbf{X}, \mathbb{X}, t), g(\mathbf{X}, \mathbb{X}, t)\}_{\mathbf{X}, \mathbb{X}}$ denotes the Poisson bracket of two functions f and g evaluated in the full phase space:

$$\{f, g\}_{\mathbf{X}, \mathbb{X}} = \partial_{\mathbf{Q}} f \cdot \partial_{\mathbf{P}} g - \partial_{\mathbf{P}} f \cdot \partial_{\mathbf{Q}} g + \sum_{\alpha=1}^{N} \left(\partial_{\mathbf{q}^\alpha} f \cdot \partial_{\mathbf{p}^\alpha} g - \partial_{\mathbf{p}^\alpha} f \cdot \partial_{\mathbf{q}^\alpha} g\right). \tag{3.5}$$

Both sides of Eq. (3.5) are functions of the full phase space variables, but we have suppressed this explicit dependence for notational simplicity. Using Eq. (3.1), we can rewrite Liouville's equation as

$$\begin{aligned}
\partial_t \phi &= -\partial_{\mathbf{Q}} \cdot \left[M^{-1}\mathbf{P}\phi\right] - \partial_{\mathbf{P}} \cdot \left[\left(-\partial_{\mathbf{Q}} V - \sum_\alpha \partial_{\mathbf{Q}} h^\alpha\right)\phi\right] - \left\{\phi, \sum_\alpha h^\alpha\right\}_{\mathbb{X}} \tag{3.6} \\
&= -\partial_{\mathbf{Q}} \cdot \left[M^{-1}\mathbf{P}\phi\right] - \partial_{\mathbf{P}} \cdot \left[\left(-\partial_{\mathbf{Q}} V - \sum_\alpha \partial_{\mathbf{Q}} h^\alpha\right)\phi\right] - \sum_\alpha \{\phi, h^\alpha\}_{\mathbf{x}^\alpha},
\end{aligned}$$

where $\{f, g\}_{\mathbb{X}}$ and $\{f, g\}_{\mathbf{x}^\alpha}$ denote Poisson brackets in the joint bath phase space and α^{th} bath phase space, respectively, of the functions $f(\mathbf{X}, \mathbb{X}, t)$ and $g(\mathbf{X}, \mathbb{X}, t)$:

$$\begin{aligned}
\{f, g\}_{\mathbb{X}} &= \sum_{\alpha=1}^{N} \left(\partial_{\mathbf{q}^\alpha} f \cdot \partial_{\mathbf{p}^\alpha} g - \partial_{\mathbf{p}^\alpha} f \cdot \partial_{\mathbf{q}^\alpha} g \right), &(3.7) \\
\{f, g\}_{\mathbf{x}^\alpha} &= \partial_{\mathbf{q}^\alpha} f \cdot \partial_{\mathbf{p}^\alpha} g - \partial_{\mathbf{p}^\alpha} f \cdot \partial_{\mathbf{q}^\alpha} g.
\end{aligned}$$

The framework above is quite general, and we now impose two restrictions that will be critical for yielding thermodynamic-like behavior. First, we assume that for any fixed value of $\mathbf{Q}$, the evolution of $\mathbf{x}^\alpha(t)$ is ergodic and mixing on a bounded connected energy shell defined by a surface of constant h^α in $\mathbf{x}^\alpha$ phase space. These sets of energy shells, as well as the bath Hamiltonians themselves, are considered to be parametrized by the system of interest's position $\mathbf{Q}$. The ergodicity and mixing assumption implies that, for fixed $\mathbf{Q}$, the αth bath will relax to a microcanonical distribution over a single energy shell in the αth bath phase space. Second, we assume that the baths are comprised of microscopic, fast, light degrees of freedom, and that the system of interest is comprised of mesoscopic, heavy, slow degrees of freedom. Specifically, we assume that the characteristic time-scale over which the system of interest's Hamiltonian significantly changes is much larger than all characteristic ergodic time-scales associated with the baths. As the system of interest slowly evolves in time, the energy surfaces of constant h^α in $\mathbf{x}^\alpha$ space will slowly and continuously deform. For an infinitely large separation between the slow and fast time-scales, the marginal distribution of each bath will remain microcanonical and cling to a single deforming energy shell, and the phase space volumes enclosed by each energy shells (the so-called ergodic adiabatic invariant) will be conserved [75]. For a large but finite separation of time-scales, each individual marginal bath distribution will generally not be microcanonical and will occupy a continuum of energy shells, and violations of the conservation of the ergodic adiabatic invariants are expected to grow diffusively in time [75]. In this work, we will consider a large but finite separation of time-scales - the infinite time-scale separation distribution will emerge as a lowest order approximation.

When the above assumptions are imposed, the slowly evolving system of interest will experience rapidly fluctuating forces due to interactions with the rapidly evolving chaos baths. We are not interested in the dynamics taking place over the fast microscopic time-scales as-

sociated these fluctuations - dynamics occurring over these time-scales are considered to be, for all practical purposes, unobservable. Rather, we are interested in the dynamics taking place over the slow mesoscopic time-scale associated with the system of interest. Although the baths' phase space variables evolve on the fast time-scales, from Hamilton's equations (3.3) we see that changes in the energy E^α of the αth bath depend linearly $\dot{\mathbf{Q}}$ and thus occur over the mescscopic time-scale:

$$
\begin{aligned}
\dot{E}^\alpha(t) &= \frac{\mathrm{d}}{\mathrm{d}t} h^\alpha(\mathbf{x}^\alpha(t); \mathbf{Q}(t)) \\
&= \partial_{\mathbf{Q}} h^\alpha(\mathbf{x}^\alpha; \mathbf{Q})|_{\mathbf{x}^\alpha(t); \mathbf{Q}(t)} \cdot \dot{\mathbf{Q}}(t).
\end{aligned}
\tag{3.8}
$$

We therefore consider the bath energies to be mesoscopic variables in addition to the system of interest's phase space variables. Each bath will generally be distributed over a continuum of energy shells, and the distribution on each of the occupied shells will be approximately microcanonical. Roughly speaking, when $\mathbf{Q}$ changes significantly, each bath will travel in phase space to a new energy shell, but because the ergodic time-scales are much smaller than the system time-scales, the baths will have time to chaotically explore the entirety of the new energy shells before $\mathbf{Q}$ has time to further change significantly. We will therefore project the fast bath variables onto the energy shells determined by $E^\alpha = h^\alpha(\mathbf{x}; \mathbf{Q})$ and consider the slow time-scale evolution of $(\mathbf{X}, E^1, ..., E^N)$. For notational simplicity, we will often denote the variables $(E^1, ..., E^N)$ jointly by the boldface $\mathbf{E}$. We expect this reduced description to accurately depict the mesoscopic dynamics of the system of interest and baths on time-scales much greater than the baths' ergodic time-scales. The quantity of interest will be the statistical distribution of the mesoscopic variables, denoted by $W(\mathbf{X}, \mathbf{E}, t)$:

$$
W(\mathbf{X}, \mathbf{E}, t) = \int \mathrm{d}\mathbb{X}\, \phi(\mathbf{X}, \mathbb{X}, t) \prod_\alpha \delta(E^\alpha - h^\alpha(\mathbf{x}^\alpha; \mathbf{Q})).
\tag{3.9}
$$

Throughout this work, we will refer to ϕ and W as the microscopic and mesoscopic distributions, respectively. The goal of this section is to apply multiple time-scale techniques to Liouville's equation (3.6) to derive an approximate mesoscopic evolution equation for

$W(\mathbf{X}, \mathbf{E}, t)$.

Before beginning the derivation of the evolution equation for W, we introduce some notation and useful identities. We define $\Omega^\alpha(E^\alpha, \mathbf{Q})$ to be the volume of $\mathbf{x}^\alpha$ phase space enclosed by the E^α energy shell when the system of interest is located at $\mathbf{Q}$:

$$\Omega^\alpha(E^\alpha, \mathbf{Q}) = \int d\mathbf{x}^\alpha \, \Theta(E^\alpha - h^\alpha(\mathbf{x}^\alpha; \mathbf{Q})), \tag{3.10}$$

where $\Theta(y)$ is the unit step function, equal to 1 if $y \geq 0$, and 0 otherwise. We define $\Sigma^\alpha(E^\alpha, \mathbf{Q})$ to be the partial derivative of $\Omega^\alpha(E^\alpha, \mathbf{Q})$ with respect to E^α:

$$\begin{aligned} \Sigma^\alpha(E^\alpha, \mathbf{Q}) &= \partial_{E^\alpha} \Omega^\alpha(E^\alpha, \mathbf{Q}) \tag{3.11} \\ &= \int d\mathbf{x}^\alpha \, \delta(E^\alpha - h^\alpha(\mathbf{x}^\alpha; \mathbf{Q})) \end{aligned}$$

The quantity $\Sigma^\alpha(E^\alpha, \mathbf{Q})$ is the microcanonical partition function of the αth bath: it gives a measure of the number of states in lying on the E^α energy shell in $\mathbf{x}^\alpha$ phase space when the system of interest is located at $\mathbf{Q}$. Let $\langle f \rangle_{\mathbf{E}, \mathbf{Q}}$ denote a microcanonical average of an arbitrary observable $f(\mathbf{X}, \mathbb{X}, t)$ over the energy shells of energies $E^1, ..., E^N$ in the joint bath phase space when the system of interest is located at $\mathbf{Q}$:

$$\langle f \rangle_{\mathbf{E}, \mathbf{Q}} = \frac{1}{\prod_\alpha \Sigma^\alpha(E^\alpha, \mathbf{Q})} \int d\mathbb{X} \, f(\mathbf{X}, \mathbb{X}, t) \prod_\alpha \delta(E^\alpha - h^\alpha(\mathbf{x}^\alpha; \mathbf{Q})). \tag{3.12}$$

We thus have

$$W(\mathbf{X}, \mathbf{E}, t) = \left[\prod_\alpha \Sigma^\alpha(E^\alpha, \mathbf{Q}) \right] \langle \phi \rangle_{\mathbf{E}, \mathbf{Q}}. \tag{3.13}$$

We similarly let $\langle f \rangle_{E^\alpha, \mathbf{Q}}$ denote a microcanonical average over only the αth bath:

$$\langle f \rangle_{E^\alpha, \mathbf{Q}} = \frac{1}{\Sigma^\alpha(E^\alpha, \mathbf{Q})} \int d\mathbf{x}^\alpha \, f(\mathbf{X}, \mathbb{X}, t) \delta(E^\alpha - h^\alpha(\mathbf{x}^\alpha; \mathbf{Q})). \tag{3.14}$$

Next, we define $\mathbf{u}^\alpha(E^\alpha, \mathbf{Q})$ to be the microcanonical average force exerted by the αth bath on the system of interest when the system is located at $\mathbf{Q}$ and the αth bath has energy E^α:

$$
\begin{aligned}
\mathbf{u}^\alpha(E^\alpha, \mathbf{Q}) &= \langle -\partial_\mathbf{Q} h^\alpha \rangle_{\mathbf{E},\mathbf{Q}} \\
&= -\frac{1}{\Sigma^\alpha(E^\alpha, \mathbf{Q})} \int \mathrm{d}\mathbf{x}^\alpha \partial_\mathbf{Q} h^\alpha(\mathbf{x}^\alpha; \mathbf{Q})\, \delta(E^\alpha - h^\alpha(\mathbf{x}^\alpha; \mathbf{Q})).
\end{aligned}
\tag{3.15}
$$

We thus have from Eq. (3.10)

$$
\mathbf{u}^\alpha(E^\alpha, \mathbf{Q}) = \frac{1}{\Sigma^\alpha(E^\alpha, \mathbf{Q})} \partial_\mathbf{Q} \Omega^\alpha(E^\alpha, \mathbf{Q}),
\tag{3.16}
$$

which together with Eq. (3.11) yields the following useful relation:

$$
\begin{aligned}
\partial_\mathbf{Q} \Sigma^\alpha(E^\alpha, \mathbf{Q}) &= \partial^2_{\mathbf{Q} E^\alpha} \Omega^\alpha(E^\alpha, \mathbf{Q}) \\
&= \partial_{E^\alpha} \left[\Sigma^\alpha(E^\alpha, \mathbf{Q}) \mathbf{u}^\alpha(E^\alpha, \mathbf{Q}) \right].
\end{aligned}
\tag{3.17}
$$

Finally, we define the vector operator $\hat{\mathbf{D}}^\alpha$ by

$$
\hat{\mathbf{D}}^\alpha = \partial_\mathbf{P} - M^{-1}\mathbf{P}\partial_{E^\alpha}
\tag{3.18}
$$

This operator is a derivative in the direction of the system of interest's momentum, constrained to respect the conservation of energy. By this we mean the following: let $\mathcal{E}$ denote the total energy of the entire system of interest plus baths universe, let $f(\mathbf{X}, \mathbf{E}, t)$ be some function of the system of interest's phase space coordinates, each bath's energy, and time, and let $g^1(\mathbf{X}, \mathcal{E}, E^2, ..., E^N, t)$ be defined by

$$g^1(\mathbf{X}, \mathcal{E}, E^2, ..., E^N, t) = f\left(\mathbf{X}, \mathcal{E} - \sum_{\alpha \neq 1} E^\alpha - H(\mathbf{X}, \boldsymbol{\lambda}(t)), E^2, ..., E^N, t\right). \qquad (3.19)$$

The operator $\hat{\mathbf{D}}^1$ appears naturally when taking derivatives of g^1 with respect to $\mathbf{P}$:

$$
\begin{aligned}
\partial_{\mathbf{P}} g^1(\mathbf{X}, \mathcal{E}, E^2, ..., E^N, t) &= \partial_{\mathbf{P}} f(\mathbf{X}, \mathbf{E}, t)|_{E^1 = \mathcal{E} - \sum\limits_{\alpha \neq 1} E^\alpha - H(\mathbf{X}, \boldsymbol{\lambda}(t))} \qquad (3.20) \\
&\quad - M^{-1}\mathbf{P}\, \partial_{E_1} f(\mathbf{X}, \mathbf{E}, t)|_{E^1 = \mathcal{E} - \sum\limits_{\alpha \neq 1} E^\alpha - H(\mathbf{X}, \boldsymbol{\lambda}(t))} \\
&= \hat{\mathbf{D}}^1 f(\mathbf{X}, \mathbf{E}, t)\Big|_{E^1 = \mathcal{E} - \sum\limits_{\alpha \neq 1} E^\alpha - H(\mathbf{X}, \boldsymbol{\lambda}(t))}
\end{aligned}
$$

3.3   The Derivation

3.3.1   Order Parameter

In order to facilitate a series expansion of Liouville's equation, we begin by introducing the small parameter $\varepsilon \ll 1$ and defining a fast and a slow time. At the end of the derivation, we will set ε to 1. Let t denote the slow time, and let $\tau = \frac{t}{\varepsilon}$ denote the fast time. We assume that $\mathbf{X}$ evolves under Hamilton's equations in t, and that $\mathbf{x}^\alpha$ evolves under Hamilton's equations in τ:

$$
\begin{aligned}
\dot{\mathbf{Q}}(t) &= M^{-1}\mathbf{P}(t) \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (3.21) \\
\dot{\mathbf{P}}(t) &= -\partial_{\mathbf{Q}} V(\mathbf{Q}, \boldsymbol{\lambda})|_{\mathbf{Q}(t), \boldsymbol{\lambda}(t)} - \sum_\alpha \partial_{\mathbf{Q}} h^\alpha(\mathbf{x}^\alpha; \mathbf{Q})|_{\mathbf{x}^\alpha(\tau), \mathbf{Q}(t)} \\
\dot{\mathbf{q}}^\alpha(\tau) &= \partial_{\mathbf{p}^\alpha} h^\alpha(\mathbf{x}^\alpha; \mathbf{Q})|_{\mathbf{x}^\alpha(\tau), \mathbf{Q}(t)} \\
\dot{\mathbf{p}}^\alpha(\tau) &= -\partial_{\mathbf{q}^\alpha} h^\alpha(\mathbf{x}^\alpha; \mathbf{Q})|_{\mathbf{x}^\alpha(\tau), \mathbf{Q}(t)}.
\end{aligned}
$$

Rewriting the above equations in terms of the slow time t only, we have

$$\frac{d\mathbf{Q}(t)}{dt} = M^{-1}\mathbf{P}(t) \tag{3.22}$$

$$\frac{d\mathbf{P}(t)}{dt} = -\partial_{\mathbf{Q}}V(\mathbf{Q},\boldsymbol{\lambda})|_{\mathbf{Q}(t),\boldsymbol{\lambda}(t)} - \sum_{\alpha}\partial_{\mathbf{Q}}h^{\alpha}(\mathbf{x}^{\alpha};\mathbf{Q})|_{\mathbf{x}^{\alpha}(\frac{t}{\varepsilon}),\mathbf{Q}(t)}$$

$$\frac{d\mathbf{q}^{\alpha}(\frac{t}{\varepsilon})}{dt} = \frac{1}{\varepsilon}\partial_{\mathbf{p}^{\alpha}}h^{\alpha}(\mathbf{x}^{\alpha};\mathbf{Q})|_{\mathbf{x}^{\alpha}(\frac{t}{\varepsilon}),\mathbf{Q}(t)}$$

$$\frac{d\mathbf{p}^{\alpha}(\frac{t}{\varepsilon})}{dt} = -\frac{1}{\varepsilon}\partial_{\mathbf{p}^{\alpha}}h^{\alpha}(\mathbf{x}^{\alpha};\mathbf{Q})|_{\mathbf{x}^{\alpha}(\frac{t}{\varepsilon}),\mathbf{Q}(t)}.$$

The corresponding Liuoville equation for the full phase space distribution $\phi(\mathbf{X},\mathbb{X},t)$, evaluated on the slow time, is

$$\partial_t\phi = -\partial_{\mathbf{Q}}\cdot\left[M^{-1}\mathbf{P}\phi\right] - \partial_{\mathbf{P}}\cdot\left[\left(-\partial_{\mathbf{Q}}V - \sum_{\alpha}\partial_{\mathbf{Q}}h^{\alpha}\right)\phi\right] \tag{3.23}$$

$$-\frac{1}{\varepsilon}\left\{\phi,\sum_{\alpha}h^{\alpha}\right\}_{\mathbb{X}}.$$

We now make the ansatz that the full phase space distribution can be written as a power series in ε:

$$\phi = \phi_0 + \varepsilon\phi_1 + \varepsilon^2\phi_2 + ... \tag{3.24}$$

Substituting this ansatz into Eq. (3.23) yields, at order ε^{-1},

$$0 = \left\{\phi_0,\sum_{\alpha}h^{\alpha}\right\}_{\mathbb{X}}, \tag{3.25}$$

and, at order ε^k for $k \geq 0$,

$$\partial_t \phi_k = -\partial_{\mathbf{Q}} \cdot \left[M^{-1} \mathbf{P} \phi_k \right] - \partial_{\mathbf{P}} \cdot \left[\left(-\partial_{\mathbf{Q}} V - \sum_\alpha \partial_{\mathbf{Q}} h^\alpha \right) \phi_k \right] \tag{3.26}$$
$$- \left\{ \phi_{k+1}, \sum_\alpha h^\alpha \right\}_{\mathbb{X}} .$$

The power series expansion of ϕ also yields a power series expansion for the distribution $W(\mathbf{X}, \mathbf{E}, t)$, which can be seen from Eq. (3.9):

$$W = W_0 + \varepsilon W_1 + \varepsilon^2 W_2 + ... \tag{3.27}$$

We now proceed with our derivation by making repeated use of Eqs. (3.25) and (3.26) to find the functional forms of ϕ and W and the evolution equation for W, good to successively higher orders.

3.3.2 Functional forms of ϕ_0 and W_0

Reinserting the explicit functional dependencies, we rewrite Eq. (3.25) as

$$0 = \left\{ \phi_0(\mathbf{X}, \mathbb{X}, t), \sum_\alpha h^\alpha(\mathbf{x}^\alpha; \mathbf{Q}) \right\}_{\mathbb{X}} . \tag{3.28}$$

The above Poisson bracket is taken over the joint bath phase space, with $\mathbf{Q}, \mathbf{P}$, and t treated as fixed parameters. The Poisson bracket of ϕ_0 with the joint bath Hamiltonian vanishes, so ϕ_0 is a function the joint bath phase space variables only through the independent constants of the motion of the joint bath Hamiltonian [80] (by constant of the motion, we mean a constant of the motion of the joint bath phase space variables when $(\mathbf{Q}, \mathbf{P})$ is held fixed). When the system of interest is held fixed, the baths do not interact, and each bath is individually ergodic on an energy shell in its own phase space evolving under a time-independent Hamiltonian, so the independent constants of the motion are precisely the bath Hamiltonians h^α. Were this not the case, then the baths could not be ergodic on

their energy shells due to constraints imposed by the additional constants of the motion. We thus have

$$\phi_0(\mathbf{X}, \mathbb{X}, t) = A(\mathbf{X}, h^1(\mathbf{x}^1; \mathbf{Q}), ..., h^N(\mathbf{x}^N, \mathbf{Q}), t), \tag{3.29}$$

for some function A. The functional form of W_0 is then,

$$
\begin{aligned}
W_0(\mathbf{X}, \mathbf{E}, t) &= \int d\mathbb{X} \, \phi_0(\mathbf{X}, \mathbb{X}, t) \prod_\alpha \delta(E^\alpha - h^\alpha(\mathbf{x}^\alpha; \mathbf{Q})) \tag{3.30} \\
&= \left[\prod_\alpha \Sigma^\alpha(E^\alpha, \mathbf{Q}) \right] A(\mathbf{X}, \mathbf{E}, t).
\end{aligned}
$$

3.3.3 Evolution equation for W_0

To find the evolution equation for W_0, we will use the $O(\varepsilon^0)$ expression of Eq. (3.26):

$$
\begin{aligned}
\partial_t \phi_0 &= -\partial_{\mathbf{Q}} \cdot \left[M^{-1}\mathbf{P}\phi_0 \right] - \partial_{\mathbf{P}} \cdot \left[\left(-\partial_{\mathbf{Q}} V - \sum_\alpha \partial_{\mathbf{Q}} h^\alpha \right) \phi_0 \right] \tag{3.31} \\
&\quad - \left\{ \phi_1, \sum_\alpha h^\alpha \right\}_{\mathbb{X}}.
\end{aligned}
$$

From Eq. (3.29) we have

$$\partial_{\mathbf{Q}} \cdot \left[M^{-1}\mathbf{P}\phi_0(\mathbf{X}, \mathbb{X}, t) \right] = \partial_{\mathbf{Q}} \cdot \left[M^{-1}\mathbf{P} A(\mathbf{X}, h^1(\mathbf{x}^1; \mathbf{Q}), ..., h^N(\mathbf{x}^N; \mathbf{Q}), t) \right], \tag{3.32}$$

which gives

$$\partial_{\mathbf{Q}} \cdot \left[M^{-1}\mathbf{P}\phi_0(\mathbf{X}, \mathbb{X}, t) \right] = \partial_{\mathbf{Q}} \cdot \left[M^{-1}\mathbf{P}A(\mathbf{X}, \mathbf{E}, t) \right]_{E^\alpha = h^\alpha(\mathbf{x}^\alpha; \mathbf{Q})} \tag{3.33}$$

$$+ \sum_\beta \partial_{E^\beta} \left[M^{-1}\mathbf{P}A(\mathbf{X}, \mathbf{E}, t) \right]_{E^\alpha = h^\alpha(\mathbf{x}^\alpha; \mathbf{Q})} \cdot \partial_{\mathbf{Q}} h^\beta(\mathbf{x}^\beta; \mathbf{Q}).$$

Employing Eq. (3.29) again, as well as the definition in Eq. (3.18), we see that Eq. (3.31) can be written as

$$
\begin{aligned}
\partial_t A(\mathbf{X}, h^1(\mathbf{x}^1; \mathbf{Q}), ..., h^N(\mathbf{x}^N; \mathbf{Q}), t) \;=\; & -\partial_{\mathbf{Q}} \cdot \left[M^{-1}\mathbf{P}A(\mathbf{X}, \mathbf{E}, t) \right]_{E^\alpha = h^\alpha(\mathbf{x}^\alpha; \mathbf{Q})} \\
& -\; \partial_{\mathbf{P}} \cdot \left[-\partial_{\mathbf{Q}}V(\mathbf{Q}, \boldsymbol{\lambda}(t))A(\mathbf{X}, \mathbf{E}, t) \right]_{E^\alpha = h^\alpha(\mathbf{x}^\alpha; \mathbf{Q})} \\
& +\; \sum_\beta \hat{\mathbf{D}}^\beta \left[A(\mathbf{X}, \mathbf{E}, t) \right]_{E^\alpha = h^\alpha(\mathbf{x}^\alpha; \mathbf{Q})} \cdot \partial_{\mathbf{Q}} h^\beta(\mathbf{x}^\beta; \mathbf{Q}) \\
& -\; \left\{ \phi_1(\mathbf{X}, \mathbb{X}, t), \sum_\alpha h^\alpha(\mathbf{x}^\alpha; \mathbf{Q}) \right\}_{\mathbb{X}}.
\end{aligned}
\tag{3.34}
$$

Taking the microcanonical average of Eq. (3.34) over the bath energy shells $E^1, ..., E^N$ with the system of interest fixed at $\mathbf{Q}$, and using the definition of $\mathbf{u}^\alpha$ in Eq. (3.15) gives

$$
\begin{aligned}
\partial_t A(\mathbf{X}, \mathbf{E}, t) \;=\; & -\partial_{\mathbf{Q}} \cdot \left[M^{-1}\mathbf{P}A(\mathbf{X}, \mathbf{E}, t) \right] \\
& -\partial_{\mathbf{P}} \cdot \left[-\partial_{\mathbf{Q}}V(\mathbf{Q}, \boldsymbol{\lambda}(t))A(\mathbf{X}, \mathbf{E}, t) \right] \\
& -\sum_\alpha \mathbf{u}^\alpha(E^\alpha, \mathbf{Q}) \cdot \hat{\mathbf{D}}^\alpha A(\mathbf{X}, \mathbf{E}, t) \\
& -\left\langle \left\{ \phi_1, \sum_\alpha h^\alpha \right\}_{\mathbb{X}} \right\rangle_{\mathbf{E}, \mathbf{Q}}.
\end{aligned}
\tag{3.35}
$$

The last term on the right hand side of Eq. (3.35) can be written as

$$
\left\langle \left\{ \phi_1, \sum_\alpha h^\alpha \right\}_{\mathbb{X}} \right\rangle_{\mathbf{E}, \mathbf{Q}} = \sum_\alpha \int \left[\prod_{\beta \neq \alpha} \frac{\mathrm{d}\mathbf{x}^\beta \delta(E^\beta - h^\beta(\mathbf{x}^\beta, \mathbf{Q}))}{\Sigma^\beta(E^\beta, \mathbf{Q})} \right] \langle \{ \phi_1, h^\alpha \}_{\mathbf{x}^\alpha} \rangle_{E^\alpha, \mathbf{Q}}.
\tag{3.36}
$$

The term $\langle \{\phi_1, h^\alpha\}_{\mathbf{x}^\alpha}\rangle_{E^\alpha, \mathbf{Q}}$, and thus Eq. (3.36) itself, vanishes identically. This follows from the fact that the microcaonical average of the Poison bracket of any function with a corresponding Hamiltonian vanishes identically: see the supplementary material of Ref. [81] for a proof. We thus have a closed evolution equation for $A(\mathbf{X}, \mathbf{E}, t)$:

$$
\begin{aligned}
\partial_t A(\mathbf{X}, \mathbf{E}, t) \;=\; & -\partial_\mathbf{Q} \cdot \left[M^{-1}\mathbf{P} A(\mathbf{X}, \mathbf{E}, t) \right] \\
& -\partial_\mathbf{P} \cdot \left[-\partial_\mathbf{Q} V(\mathbf{Q}, \boldsymbol{\lambda}(t)) A(\mathbf{X}, \mathbf{E}, t) \right] \\
& -\sum_\alpha \mathbf{u}^\alpha(E^\alpha, \mathbf{Q}) \cdot \hat{\mathbf{D}}^\alpha A(\mathbf{X}, \mathbf{E}, t)
\end{aligned}
\tag{3.37}
$$

Multiplying both sides of Eq. (3.37) by $\prod_\alpha \Sigma^\alpha(E^\alpha, \mathbf{Q})$, using Eq. (3.30), and rearranging, we find

$$
\begin{aligned}
\partial_t W_0(\mathbf{X}, \mathbf{E}, t) \;=\; & -\partial_\mathbf{Q} \cdot \left[M^{-1}\mathbf{P} W_0(\mathbf{X}, \mathbf{E}, t) \right] \\
& -\partial_\mathbf{P} \cdot \left[-\partial_\mathbf{Q} V(\mathbf{Q}, \boldsymbol{\lambda}(t)) W_0(\mathbf{X}, \mathbf{E}, t) \right] \\
& -\sum_\alpha \hat{\mathbf{D}}^\alpha \cdot \left[\mathbf{u}^\alpha(E^\alpha, \mathbf{Q}) W_0(\mathbf{X}, \mathbf{E}, t) \right] \\
& +A(\mathbf{X}, \mathbf{E}, t)\mathbf{P}^T M^{-1}\partial_\mathbf{Q} \left[\prod_\alpha \Sigma^\alpha(E^\alpha, \mathbf{Q}) \right] \\
& +A(\mathbf{X}, \mathbf{E}, t)\sum_\alpha \hat{\mathbf{D}}^\alpha \cdot \left[\mathbf{u}^\alpha(E^\alpha, \mathbf{Q}) \prod_\beta \Sigma^\beta(E^\beta, \mathbf{Q}) \right] .
\end{aligned}
\tag{3.38}
$$

From the identity in Eq. (3.17), we have

$$
\sum_\alpha \hat{\mathbf{D}}^\alpha \cdot \left[\mathbf{u}^\alpha(E^\alpha, \mathbf{Q}) \prod_\beta \Sigma^\beta(E^\beta, \mathbf{Q}) \right] = -\mathbf{P}^T M^{-1}\partial_\mathbf{Q} \left[\prod_\alpha \Sigma^\alpha(E^\alpha, \mathbf{Q}) \right] ,
\tag{3.39}
$$

so the last two terms of Eq. (3.38) cancel, and we finally arrive at a closed evolution equation for W_0:

$$\partial_t W_0(\mathbf{X}, \mathbf{E}, t) \;=\; -\partial_{\mathbf{Q}} \cdot \left[M^{-1} \mathbf{P} W_0(\mathbf{X}, \mathbf{E}, t) \right] \tag{3.40}$$
$$-\partial_{\mathbf{P}} \cdot \left[-\partial_{\mathbf{Q}} V(\mathbf{Q}, \boldsymbol{\lambda}(t)) W_0(\mathbf{X}, \mathbf{E}, t) \right]$$
$$-\sum_\alpha \hat{\mathbf{D}}^\alpha \cdot \left[\mathbf{u}^\alpha(E^\alpha, \mathbf{Q}) W_0(\mathbf{X}, \mathbf{E}, t) \right].$$

This equation gives the time evolution of W in the adiabatic limit (when $\varepsilon \to 0$). The lack of second order or higher derivatives is indicative of the deterministic evolution of $(\mathbf{X}, \mathbf{E})$ in the adiabatic limit. The first two terms in Eq. (3.40) are also present in Liouville's equation for ϕ and represent the deterministic evolution of the system of interest under its own slow Hamiltonian dynamics. We can gain some intuition for the last term in Eq. (3.40) by expanding out the $\hat{\mathbf{D}}^\alpha$ operators and rearranging:

$$\partial_t W_0(\mathbf{X}, \mathbf{E}, t) \;=\; -\partial_{\mathbf{Q}} \cdot \left[M^{-1} \mathbf{P} W_0(\mathbf{X}, \mathbf{E}, t) \right] \tag{3.41}$$
$$-\partial_{\mathbf{P}} \cdot \left[\left(-\partial_{\mathbf{Q}} V(\mathbf{Q}, \boldsymbol{\lambda}(t)) + \sum_\alpha \mathbf{u}^\alpha(E^\alpha, \mathbf{Q}) \right) W_0(\mathbf{X}, \mathbf{E}, t) \right]$$
$$-\sum_\alpha \partial_{E^\alpha} \left[-\mathbf{P}^T M^{-1} \mathbf{u}^\alpha(E^\alpha, \mathbf{Q}) W_0(\mathbf{X}, \mathbf{E}, t) \right].$$

We thus see that in the adiabatic limit, the system of interest is subject to the additional forces $\mathbf{u}^\alpha$ from the baths. The rate of work done by the αth bath on the system of interest is thus $\dot{\mathbf{Q}}^T \mathbf{u}^\alpha = \mathbf{P}^T M^{-1} \mathbf{u}^\alpha$, and the corresponding rate of change of the αth bath's energy is thus $-\mathbf{P}^T M^{-1} \mathbf{u}^\alpha$, as demonstrated by the last term of Eq. (3.41). This change in energy is deterministic and coincides with the change in energy determined by the conservation of the ergodic adiabatic invariant $\Omega^\alpha(E^\alpha, \mathbf{Q})$.

3.3.4   Functional form of ϕ_1 and W_1

An equation for ϕ_1 can be found by equating Eq. (3.34) to Eq. (3.37), where Eq. (3.37) is evaluated for $E^\alpha = h^\alpha(\mathbf{x}^\alpha; \mathbf{Q})$:

$$\left\{ \phi_1(\mathbf{X}, \mathbb{X}, t), \sum_\alpha h^\alpha(\mathbf{x}^\alpha; \mathbf{Q}) \right\}_{\mathbb{X}} = -\sum_\alpha \hat{\mathbf{D}}^\alpha \left[A(\mathbf{X}, \mathbf{E}, t) \right]_{E^\alpha = h^\alpha(\mathbf{x}^\alpha; \mathbf{Q})} \tag{3.42}$$
$$\cdot \left[-\partial_\mathbf{Q} h^\alpha(\mathbf{x}^\alpha; \mathbf{Q}) - \mathbf{u}^\alpha(h^\alpha(\mathbf{x}^\alpha; \mathbf{Q}), \mathbf{Q}) \right]$$
$$= g(\mathbf{X}, \mathbb{X}, t),$$

where we define

$$g(\mathbf{X}, \mathbb{X}, t) = -\sum_\alpha \hat{\mathbf{D}}^\alpha \left[A(\mathbf{X}, \mathbf{E}, t) \right]_{E^\alpha = h^\alpha(\mathbf{x}^\alpha; \mathbf{Q})} \tag{3.43}$$
$$\cdot \left[-\partial_\mathbf{Q} h^\alpha(\mathbf{x}^\alpha; \mathbf{Q}) - \mathbf{u}^\alpha(h^\alpha(\mathbf{x}^\alpha; \mathbf{Q}), \mathbf{Q}) \right].$$

The function A is determined by Eq. (3.37), so Eq. (3.42) is a closed partial differential equation for ϕ_1 which will have a homogeneous and inhomogeneous solution, denoted ϕ_1^H and ϕ_1^I, respectively:

$$\phi_1(\mathbf{X}, \mathbb{X}, t) = \phi_1^H(\mathbf{X}, \mathbb{X}, t) + \phi_1^I(\mathbf{X}, \mathbb{X}, t). \tag{3.44}$$

$$\left\{ \phi_1^H(\mathbf{X}, \mathbb{X}, t), \sum_\alpha h^\alpha(\mathbf{x}^\alpha; \mathbf{Q}) \right\}_{\mathbb{X}} = 0. \tag{3.45}$$

$$\left\{ \phi_1^I(\mathbf{X}, \mathbb{X}, t), \sum_\alpha h^\alpha(\mathbf{x}^\alpha; \mathbf{Q}) \right\}_{\mathbb{X}} = g(\mathbf{X}, \mathbb{X}, t). \tag{3.46}$$

Applying the same logic used in Sec. 3.3.2, Eq. (3.45) implies the following functional form for ϕ_1^H:

$$\phi_1^H(\mathbb{X}, t) = B(\mathbf{X}, h^1(\mathbf{x}^1; \mathbf{Q}), ..., h^N(\mathbf{x}^N; \mathbf{Q}), t), \tag{3.47}$$

for some function B. Following Berry and Robbins in Ref. [77], the inhomogeneous solution is given by

$$\phi_1^I(\mathbf{X}, \mathbb{X}, t) \;\; = \;\; \int_{-\infty}^{0} \mathrm{d}t' g(\mathbf{X}, \mathbb{X}(t'), t), \tag{3.48}$$

where $\mathbf{x}^\alpha(0)$ is defined by the arguments of $\phi_1^I(\mathbf{X}, \mathbb{X}, t)$:

$$\mathbf{x}^\alpha(0) = \mathbf{x}^\alpha, \tag{3.49}$$

and $\mathbf{x}^\alpha(t')$ denotes the point to which $\mathbf{x}^\alpha$ evolves to under the Hamiltonian h^α after a time t', with $\mathbf{Q}$ held fixed.

The solution in Eq. (3.48) comes with some caveats which warrant some discussion. From Eq. (3.43), we see that $g(\mathbf{X}, \mathbb{X}(t'), t)$ is an oscillatory function of t': g is proportional the difference between the instantaneous forces of the baths of on the system of interest and the microcanonical average forces of the baths on the system of interest, and as the αth bath evolves under h^α (with $\mathbf{Q}$ held fixed), it chaotically and ergodically explores a single energy shell, so the instantaneous force oscillates indefinitely (and generally non-periodically) about the average force. We thus conclude that the integral in Eq. (3.48) does not converge, and ϕ_1^I as given in Eq. (3.48) is not a well-defined function. The indefinite oscillation also impedes the verification of ϕ_1^I as the solution to Eq. (3.46):

$$\left\{\phi_1^I, \sum_\alpha h^\alpha\right\}_{\mathbb{X}} = \int_{-\infty}^0 dt' \left\{g(\mathbf{X}, \mathbb{X}(t'), t), \sum_\alpha h^\alpha(\mathbf{x}^\alpha; \mathbf{Q})\right\}_{\mathbb{X}} \qquad (3.50)$$

$$= \int_{-\infty}^0 dt' \left\{g(\mathbf{X}, \mathbb{X}(t'), t), \sum_\alpha h^\alpha(\mathbf{x}^\alpha(t'); \mathbf{Q})\right\}_{\mathbb{X}(t')}$$

$$= \int_{-\infty}^0 dt' \frac{d}{dt'} g(\mathbf{X}, \mathbb{X}(t'), t)$$

$$= g(\mathbf{X}, \mathbb{X}, t) - \lim_{s \to -\infty} g(\mathbf{X}, \mathbb{X}(s), t).$$

The second line of Eq. (3.50) follows from the fact that $h^\alpha(\mathbf{x}^\alpha; \mathbf{Q}) = h^\alpha(\mathbf{x}^\alpha(t'); \mathbf{Q})$ for any t' when the αth bath evolves over a single energy shell, as well as the invariance of the Poisson bracket under a Hamiltonian flow, and the third line follows from a fundamental property of the Poisson bracket [80]. The term $\lim_{s \to -\infty} g(\mathbf{X}, \mathbb{X}(s), t)$ is ill-defined and makes ambiguous whether or not ϕ_1^I given in Eq. (3.48) is indeed a solution to Eq. (3.46). This offending term, as well as the non-convergence of the integral in Eq. (3.48), can be addressed introducing the convergence factor $e^{\gamma t'}$, where $\gamma > 0$. The function $\phi_{1,\gamma}^I$, given by

$$\phi_{1,\gamma}^I(\mathbf{X}, \mathbb{X}, t) = \int_{-\infty}^0 dt' g(\mathbf{X}, \mathbb{X}(t'), t) e^{\gamma t'}, \qquad (3.51)$$

is well-defined and related to the inhomogeneous solution ϕ_1^I by

$$\phi_1^I(\mathbf{X}, \mathbb{X}, t) = \lim_{\gamma \to 0} \phi_{1,\gamma}^I(\mathbf{X}, \mathbb{X}, t). \qquad (3.52)$$

Taking the Poisson bracket of Eq. (3.51) gives

$$\left\{\phi_{1,\gamma}^I, \sum_\alpha h^\alpha\right\}_{\mathbb{X}} = \int_{-\infty}^0 dt' \left\{g(\mathbf{X}, \mathbb{X}(t'), t) e^{\gamma t'}, \sum_\alpha h^\alpha(\mathbf{x}^\alpha; \mathbf{Q})\right\}_{\mathbb{X}} \qquad (3.53)$$

$$= \int_{-\infty}^0 dt' \left(\frac{d}{dt'} \left[g(\mathbf{X}, \mathbb{X}(t'), t) e^{\gamma t'}\right] - \gamma g(\mathbf{X}, \mathbb{X}(t'), t) e^{\gamma t'}\right)$$

$$= g(\mathbf{X}, \mathbb{X}, t) - \gamma \phi_{1,\gamma}^I(\mathbf{X}, \mathbb{X}, t),$$

and we find

$$
\lim_{\gamma \to 0} \left\{ \phi_{1,\gamma}^I, \sum_\alpha h^\alpha \right\}_{\mathbb{X}} = \left\{ \phi_1^I, \sum_\alpha h^\alpha \right\}_{\mathbb{X}} \tag{3.54}
$$
$$
= g(\mathbf{X}, \mathbb{X}, t).
$$

From the two preceding equations, we see that for non-zero γ, $\phi_{1,\gamma}^I$ is a well-defined function, but does not solve Eq. (3.46), and for $\gamma = 0$, $\phi_{1,\gamma}^I = \phi_1^I$ solves Eq. (3.46), but is not a well-defined function. Rather than a *function* of phase space variables, ϕ_1^I is well-defined as a *distribution* over phase space which is intended to be integrated over. From Eqs. (3.42) and (3.43), we see that the microcanonical average of g vanishes, and so the microcanonical average of the distribution ϕ_1^I vanishes as well. Using Eqs. (3.13) and (3.47), we thus find the following functional form for W_1:

$$
\begin{aligned}
W_1(\mathbf{X}, \mathbf{E}, t) &= \left[\prod_\alpha \Sigma^\alpha (E^\alpha, \mathbf{Q}) \right] \langle \phi_1^H + \phi_1^I \rangle_{\mathbf{E}, \mathbf{Q}} \tag{3.55} \\
&= \left[\prod_\alpha \Sigma^\alpha (E^\alpha, \mathbf{Q}) \right] \langle \phi_1^H \rangle_{\mathbf{E}, \mathbf{Q}} \\
&= \left[\prod_\alpha \Sigma^\alpha (E^\alpha, \mathbf{Q}) \right] B(\mathbf{X}, \mathbf{E}, t).
\end{aligned}
$$

The inhomogeneous term ϕ_1^I does not factor directly into the functional form of W_1, but as we will see, the time-evolution of ϕ_1^I influences the time-evolution of W_1, and ϕ_1^I will turn out to be the source of diffusion terms (and is thus related to fluctuations and dissipation) in the Fokker-Planck equation for W. This observation is reminiscent of the Gaussian white noise commonly used to model fluctuations in stochastic thermodynamics: the Gaussian white noise is a continuous non-differentiable fractal distribution over time with mean zero which is ill-defined as a function, and the fluctuations which it induces are the source of the diffusion terms in the corresponding Fokker-Planck equations [82].

67

3.3.5 Evolution equation for W_1

To find the evolution equation for W_1, we begin with the $O(\varepsilon)$ expression of Eq. (3.26):

$$
\begin{aligned}
\partial_t \phi_1 \;=\;& -\partial_\mathbf{Q} \cdot \left[M^{-1}\mathbf{P}\phi_1 \right] - \partial_\mathbf{P} \cdot \left[\left(-\partial_\mathbf{Q} V - \sum_\alpha \partial_\mathbf{Q} h^\alpha \right) \phi_1 \right] \\
& - \left\{ \phi_2, \sum_\alpha h^\alpha \right\}_{\mathbb{X}} .
\end{aligned}
\tag{3.56}
$$

Applying Eqs. (3.44) and (3.47), as well as the identity in Eq. (3.33) applied to B, and rearranging, we find

$$
\begin{aligned}
\partial_t B\big(\mathbf{X}, h^1(\mathbf{x}^1;\mathbf{Q}),...,h^N(\mathbf{x}^N;\mathbf{Q}),t\big) \;=\;& -\partial_\mathbf{Q} \cdot \left[M^{-1}\mathbf{P} B(\mathbf{X},\mathbf{E},t) \right]_{E^\alpha=h^\alpha(\mathbf{x}^\alpha;\mathbf{Q})} \\
& -\; \partial_\mathbf{P} \cdot \left[-\partial_\mathbf{Q} V(\mathbf{Q},\boldsymbol{\lambda}(t)) B(\mathbf{X},\mathbf{E},t) \right]_{E^\alpha=h^\alpha(\mathbf{x}^\alpha;\mathbf{Q})} \\
& +\; \sum_\beta \hat{\mathbf{D}}^\beta \left[B(\mathbf{X},\mathbf{E},t) \right]_{E^\alpha=h^\alpha(\mathbf{x}^\alpha;\mathbf{Q})} \cdot \partial_\mathbf{Q} h^\beta(\mathbf{x}^\beta;\mathbf{Q}) \\
& -\; \left\{ \phi_2(\mathbf{X},\mathbb{X},t), \sum_\alpha h^\alpha(\mathbf{x}^\alpha;\mathbf{Q}) \right\}_{\mathbb{X}} \\
& -\; \partial_t \phi_1^I(\mathbf{X},\mathbb{X},t) \\
& -\; \partial_\mathbf{Q} \cdot \left[M^{-1}\mathbf{P} \phi_1^I(\mathbf{X},\mathbb{X},t) \right] \\
& -\; \partial_\mathbf{P} \cdot \left[\left(-\partial_\mathbf{Q} V - \sum_\alpha \partial_\mathbf{Q} h^\alpha \right) \phi_1^I(\mathbf{X},\mathbb{X},t) \right] .
\end{aligned}
\tag{3.57}
$$

Taking the microcanonical average of the above equation, and noting that the microcanonical averages of ϕ_1^I, $\partial_t \phi_1^I$, and the Poisson bracket of ϕ_2 with the bath Hamiltonians all vanish, we find an evolution equation for B:

$$\partial_t B(\mathbf{X}, \mathbf{E}, t) \;=\; -\partial_{\mathbf{Q}} \cdot \left[M^{-1} \mathbf{P} B(\mathbf{X}, \mathbf{E}, t) \right] \tag{3.58}$$

$$-\partial_{\mathbf{P}} \cdot [-\partial_{\mathbf{Q}} V(\mathbf{Q}, \boldsymbol{\lambda}(t)) B(\mathbf{X}, \mathbf{E}, t)]$$

$$-\sum_\alpha \mathbf{u}^\alpha(E^\alpha, \mathbf{Q}) \cdot \hat{\mathbf{D}}^\alpha B(\mathbf{X}, \mathbf{E}, t)$$

$$+\left\langle -\partial_{\mathbf{Q}} \cdot \left[M^{-1} \mathbf{P} \phi_1^I \right] - \partial_{\mathbf{P}} \cdot \left[\left(-\partial_{\mathbf{Q}} V - \sum_\alpha \partial_{\mathbf{Q}} h^\alpha \right) \phi_1^I \right] \right\rangle_{\mathbf{E}, \mathbf{Q}}.$$

Multiplying both sides of Eq. (3.58) by $\prod_\alpha \Sigma^\alpha(E^\alpha, \mathbf{Q})$, rearranging, and applying the identity in Eq. (3.39), we find an evolution equation for W_1:

$$\partial_t W_1(\mathbf{X}, \mathbf{E}, t) \;=\; -\partial_{\mathbf{Q}} \cdot \left[M^{-1} \mathbf{P} W_1(\mathbf{X}, \mathbf{E}, t) \right] \tag{3.59}$$

$$-\partial_{\mathbf{P}} \cdot [-\partial_{\mathbf{Q}} V(\mathbf{Q}, \boldsymbol{\lambda}(t)) W_1(\mathbf{X}, \mathbf{E}, t)]$$

$$-\sum_\alpha \hat{\mathbf{D}}^\alpha \cdot [\mathbf{u}^\alpha(E^\alpha, \mathbf{Q}) W_1(\mathbf{X}, \mathbf{E}, t)]$$

$$+\left\langle -\partial_{\mathbf{Q}} \cdot \left[M^{-1} \mathbf{P} \phi_1^I \right] \right\rangle_{\mathbf{E}, \mathbf{Q}} \prod_\alpha \Sigma^\alpha(E^\alpha, \mathbf{Q})$$

$$+\left\langle -\partial_{\mathbf{P}} \cdot \left[\left(-\partial_{\mathbf{Q}} V - \sum_\alpha \partial_{\mathbf{Q}} h^\alpha \right) \phi_1^I \right] \right\rangle_{\mathbf{E}, \mathbf{Q}} \prod_\alpha \Sigma^\alpha(E^\alpha, \mathbf{Q}).$$

The first three terms on the right hand side of Eq. (3.59) are the same terms present in the evolution equation for W_0: they represent the adiabatic effects on the evolution of W_1. The last two terms are microcanonical averages which, after being evaluated in Sec. 3.3.6 below, will yield second order derivatives and produce a Fokker-Planck equation for the evolution of W. These terms are thus intimately related to fluctuations and dissipation and are indicative of the stochastic evolution of $(\mathbf{X}, \mathbf{E})$ for large but finite time-scale separations between the system of interest and the baths.

3.3.6 The Fokker-Planck equation

We must now evaluate the function

$$J(\mathbf{X}, \mathbf{E}, t) \quad = \quad \langle -\partial_{\mathbf{Q}} \cdot [M^{-1}\mathbf{P}\phi_1^I]\rangle_{\mathbf{E},\mathbf{Q}} \prod_\alpha \Sigma^\alpha(E^\alpha, \mathbf{Q}) \tag{3.60}$$

$$+ \left\langle -\partial_{\mathbf{P}} \cdot \left[\left(-\partial_{\mathbf{Q}}V - \sum_\alpha \partial_{\mathbf{Q}}h^\alpha\right)\phi_1^I\right]\right\rangle_{\mathbf{E},\mathbf{Q}} \prod_\alpha \Sigma^\alpha(E^\alpha, \mathbf{Q}).$$

Using the product rule on the first term on the right hand side of Eq. (3.60), and using the fact that the microcanonical average of ϕ_1^I vanishes, we find

$$\langle -\partial_{\mathbf{Q}} \cdot [M^{-1}\mathbf{P}\phi_1^I]\rangle_{\mathbf{E},\mathbf{Q}} \prod_\alpha \Sigma^\alpha(E^\alpha, \mathbf{Q}) \quad = \quad \int d\mathbf{x}^1...d\mathbf{x}^N \Big[\phi_1^I(\mathbf{X}, \mathbb{X}, t)$$

$$\times \mathbf{P}^T M^{-1}\partial_{\mathbf{Q}} \prod_\alpha \delta(E^\alpha - h^\alpha(\mathbf{x}^\alpha; \mathbf{Q})) \Big].$$

$$\tag{3.61}$$

From the identity

$$\partial_{\mathbf{Q}}\left[\prod_\alpha \delta(E^\alpha - h^\alpha(\mathbf{x}^\alpha; \mathbf{Q}))\right] \quad = \quad -\sum_\alpha \partial_{\mathbf{Q}}h^\alpha(\mathbf{x}^\alpha; \mathbf{Q})\partial_{E^\alpha}\left[\prod_\beta \delta(E^\beta - h^\beta(\mathbf{x}^\beta; \mathbf{Q}))\right]$$

$$= \quad -\sum_\alpha \partial_{E^\alpha}\left[\partial_{\mathbf{Q}}[h^\alpha(\mathbf{x}^\alpha; \mathbf{Q})]\prod_\beta \delta(E^\beta - h^\beta(\mathbf{x}^\beta; \mathbf{Q}))\right],$$

$$\tag{3.62}$$

we then have

$$\langle -\partial_{\mathbf{Q}} \cdot [M^{-1}\mathbf{P}\phi_1^I]\rangle_{\mathbf{E},\mathbf{Q}} \prod_\alpha \Sigma^\alpha(E^\alpha, \mathbf{Q}) \quad = \quad \sum_\alpha \mathbf{P}^T M^{-1}\partial_{E^\alpha} \int d\mathbb{X}\Big[\left(-\partial_{\mathbf{Q}}h^\alpha(\mathbf{x}^\alpha; \mathbf{Q})\right)$$

$$\times \phi_1^I(\mathbf{X}, \mathbb{X}, t)\prod_\beta \delta(E^\beta - h^\beta(\mathbf{x}^\beta; \mathbf{Q}))\Big].$$

$$\tag{3.63}$$

Note that

$$\langle \partial_{\mathbf{P}} \cdot \partial_{\mathbf{Q}} V \phi_1^I \rangle_{\mathbf{E},\mathbf{Q}} = \partial_{\mathbf{P}} \cdot \left[\partial_{\mathbf{Q}} V \langle \phi_1^I \rangle_{\mathbf{E},\mathbf{Q}} \right] \tag{3.64}$$
$$= 0,$$

so the second term on the right hand side of Eq. (3.60) reduces to

$$\left\langle \partial_{\mathbf{P}} \cdot \sum_\alpha \partial_{\mathbf{Q}} h^\alpha \phi_1^I \right\rangle_{\mathbf{E},\mathbf{Q}} \prod_\alpha \Sigma^\alpha(E^\alpha, \mathbf{Q}) = -\sum_\alpha \partial_{\mathbf{P}} \cdot \int d\mathbb{X} \Big[\big(-\partial_{\mathbf{Q}} h^\alpha(\mathbf{x}^\alpha; \mathbf{Q}) \big)$$
$$\times \phi_1^I(\mathbf{X}, \mathbb{X}, t) \prod_\beta \delta(E^\beta - h^\beta(\mathbf{x}^\beta; \mathbf{Q})) \Big], \tag{3.65}$$

and we find

$$J(\mathbf{X}, \mathbf{E}, t) = -\sum_\alpha \hat{\mathbf{D}}^\alpha \cdot \int d\mathbb{X} \Big[\big(-\partial_{\mathbf{Q}} h^\alpha(\mathbf{x}^\alpha; \mathbf{Q}) \big) \phi_1^I(\mathbf{X}, \mathbb{X}, t) \tag{3.66}$$
$$\times \prod_\beta \delta(E^\beta - h^\beta(\mathbf{x}^\beta; \mathbf{Q})) \Big]$$
$$= -\sum_\alpha \hat{\mathbf{D}}^\alpha \cdot \left[\left(\prod_\beta \Sigma^\beta(E^\beta, \mathbf{Q}) \right) \langle (-\partial_{\mathbf{Q}} h^\alpha) \phi_1^I \rangle_{E^1,\dots,E^N,\mathbf{Q}} \right].$$

Substituting in the expression for ϕ_1^I from Eqs. (3.43) and (3.48) gives

$$
\begin{aligned}
J(\mathbf{X}, \mathbf{E}, t) \;=\;& \sum_{\alpha,\beta,i,j} \hat{D}_i^\alpha \int d\mathbb{X} \left[\left(\prod_\sigma \delta(E^\sigma - h^\sigma(\mathbf{x}^\sigma; \mathbf{Q})) \right) \right. \hspace{2cm} (3.67) \\
& \times \int_{-\infty}^0 dt' \left[\left(- \partial_{Q_i} h^\alpha(\mathbf{x}^\alpha; \mathbf{Q}) \right) \right. \\
& \times \left(- \partial_{Q_j} h^\beta(\mathbf{x}^\beta(t'); \mathbf{Q}) - u_j^\beta(h^\beta(\mathbf{x}^\beta(t'); \mathbf{Q}), \mathbf{Q}) \right) \\
& \left. \left. \times \hat{D}_j^\beta \left[A(\mathbf{X}, \mathbf{E}, t) \right]_{E^\xi = h^\xi(\mathbf{x}^\xi(t'); \mathbf{Q})} \right] \right] \\
=\;& \sum_{\alpha,\beta,i,j} \hat{D}_i^\alpha \left[\left(\prod_\sigma \Sigma^\sigma(E^\sigma, \mathbf{Q}) \right) \right. \\
& \left. \times \int_{-\infty}^0 dt' \left\langle \left(- \partial_{Q_i} h^\alpha \right)\left(- \partial_{Q_j} h_{t'}^\beta - u_j^\beta \right) \right\rangle_{\mathbf{E},\mathbf{Q}} \hat{D}_j^\beta A(\mathbf{X}, \mathbf{E}, t) \right],
\end{aligned}
$$

where we use $h_{t'}^\beta$ to denote $h^\beta(\mathbf{x}^\beta(t'); \mathbf{Q})$ inside of a microcanonical average. Note that the microcanonical average inside the time integral in the above equation can be written as

$$
\begin{aligned}
\left\langle \left(- \partial_{Q_i} h^\alpha \right)\left(- \partial_{Q_j} h_{t'}^\beta - u_j^\beta \right) \right\rangle_{\mathbf{E},\mathbf{Q}} \;=\;& \left\langle \left(- \partial_{Q_i} h^\alpha \right)\left(- \partial_{Q_j} h_{t'}^\beta - u_j^\beta \right) \right\rangle_{E^\alpha, E^\beta, \mathbf{Q}} \hspace{1cm} (3.68) \\
=\;& \left\langle \left(- \partial_{Q_i} h^\alpha \right)\left(- \partial_{Q_j} h_{t'}^\beta \right) \right\rangle_{E^\alpha, E^\beta, \mathbf{Q}} \\
& - u_i^\alpha(E^\alpha, \mathbf{Q}) u_j^\beta(E^\beta, \mathbf{Q}) \\
=\;& \left\langle \left(- \partial_{Q_i} h^\alpha - u_i^\alpha \right)\left(- \partial_{Q_j} h_{t'}^\beta - u_j^\beta \right) \right\rangle_{E^\alpha, E^\beta, \mathbf{Q}}.
\end{aligned}
$$

The αth, β^{th} component of the time integral Eq. (3.67) is thus the integrated covariance of the force of bath α of the system of interest at time $t = 0$ with the force of bath β on the system of interest at previous times, when the system of interest is held fixed. If the system of interest is fixed, bath α and bath β have no effect on each other for $\alpha \neq \beta$, so this correlation will vanish unless $\alpha = \beta$. We thus see that no effective coupling between the baths emerges at order ε. Employing our chaotic assumption, we assume that each bath is individually mixing to a sufficient degree such that the time integral in Eq. (3.67) converges. We now define the integrated covariance matrix $L_{ij}^\alpha(E^\alpha, \mathbf{Q})$ by

$$L_{ij}^{\alpha}(E^{\alpha}, \mathbf{Q}) = 2 \int_{-\infty}^{0} dt' \left\langle \left(-\partial_{Q_i} h^{\alpha} - u_i^{\alpha} \right) \left(-\partial_{Q_j} h_{t'}^{\alpha} - u_j^{\alpha} \right) \right\rangle_{E^{\alpha}, \mathbf{Q}}, \qquad (3.69)$$

which is given in matrix notation by

$$L^{\alpha}(E^{\alpha}, \mathbf{Q}) = 2 \int_{-\infty}^{0} dt' \left\langle \left(-\partial_{\mathbf{Q}} h^{\alpha} - \mathbf{u}^{\alpha} \right) \left(-\partial_{\mathbf{Q}} h_{t'}^{\alpha} - \mathbf{u}^{\alpha} \right)^{T} \right\rangle_{E^{\alpha}, \mathbf{Q}}. \qquad (3.70)$$

The function J can now be written as

$$J(\mathbf{X}, \mathbf{E}, t) = \frac{1}{2} \sum_{\alpha, i, j} \hat{D}_i^{\alpha} \left[\left(\prod_{\sigma} \Sigma^{\sigma}(E^{\sigma}, \mathbf{Q}) \right) L_{ij}^{\alpha}(E^{\alpha}, \mathbf{Q}) \hat{D}_j^{\alpha} A(\mathbf{X}, \mathbf{E}, t) \right]. \qquad (3.71)$$

The evolution equation for W_1 is thus

$$\begin{aligned}
\partial_t W_1(\mathbf{X}, \mathbf{E}, t) = \ & -\partial_{\mathbf{Q}} \cdot \left[M^{-1} \mathbf{P} W_1(\mathbf{X}, \mathbf{E}, t) \right] \qquad\qquad (3.72)\\
& -\partial_{\mathbf{P}} \cdot \left[-\partial_{\mathbf{Q}} V(\mathbf{Q}, \boldsymbol{\lambda}(t)) W_1(\mathbf{X}, \mathbf{E}, t) \right] \\
& -\sum_{\alpha} \hat{\mathbf{D}}^{\alpha} \cdot \left[\mathbf{u}^{\alpha}(E^{\alpha}, \mathbf{Q}) W_1(\mathbf{X}, \mathbf{E}, t) \right] \\
& \frac{1}{2} \sum_{\alpha, i, j} \hat{D}_i^{\alpha} \left[\Sigma^{\alpha}(E^{\alpha}, \mathbf{Q}) L_{ij}^{\alpha}(E^{\alpha}, \mathbf{Q}) \hat{D}_j^{\alpha} \left(\frac{W_0(\mathbf{X}, \mathbf{E}, t)}{\Sigma^{\alpha}(E^{\alpha}, \mathbf{Q})} \right) \right],
\end{aligned}$$

where we have made us of Eq. (3.30) to write A in terms of W_0, and the fact that $\hat{D}_i^{\alpha} \Sigma^{\beta} = 0$ unless $\alpha = \beta$. We thus see the emergence of second order derivatives in the evolution equation for W_1.

We now have evolution equations for W_0 and W_1, good to order ϵ. Noting that $W = W_0 + \varepsilon W_1$ up to $O(\varepsilon)$, we have

$$\partial_t W(\mathbf{X}, \mathbf{E}, t) = \partial_t W_0(\mathbf{X}, \mathbf{E}, t) + \varepsilon \partial_t W_1(\mathbf{X}, \mathbf{E}, t). \qquad (3.73)$$

73

Using the above equation along with Eqs. (3.41) and (3.72), neglecting terms of order ε^2 and higher, and setting $\varepsilon = 1$, we finally arrive at the Fokker-Planck equation for W:

$$
\begin{aligned}
\partial_t W \quad = \quad & -\partial_{\mathbf{Q}} \cdot \left[M^{-1} \mathbf{P} W \right] - \partial_{\mathbf{P}} \cdot \left[-\partial_{\mathbf{Q}} V \, W \right] - \sum_\alpha \hat{\mathbf{D}}^\alpha \cdot \left[\mathbf{u}^\alpha W \right] \qquad (3.74) \\
& + \frac{1}{2} \sum_{\alpha,i,j} \hat{D}_i^\alpha \left[\Sigma^\alpha L_{ij}^\alpha \hat{D}_j^\alpha \left(\frac{W}{\Sigma^\alpha} \right) \right],
\end{aligned}
$$

where we now suppress the explicit functional dependences for simplicity. Alternatively, in a more standard form, we have

$$
\begin{aligned}
\partial_t W \quad = \quad & -\partial_{\mathbf{Q}} \cdot \left[M^{-1} \mathbf{P} W \right] - \partial_{\mathbf{P}} \cdot \left[-\partial_{\mathbf{Q}} V \, W \right] \qquad\qquad\qquad\qquad (3.75) \\
& - \sum_{\alpha,i} \hat{D}_i^\alpha \left[\left(u_i^\alpha + \frac{1}{2\Sigma^\alpha} \sum_j \hat{D}_j^\alpha \left[\Sigma^\alpha L_{ij}^\alpha \right] \right) W \right] + \frac{1}{2} \sum_{\alpha,i,j} \hat{D}_i^\alpha \hat{D}_j^\alpha \left[L_{ij}^{\text{sy}\,\alpha} W \right],
\end{aligned}
$$

where $L^{\text{sy}\,\alpha}$ denotes the symmetric part of L^α (the anti-symmetric part of L^α makes no contribution to the quadratic form in the last term on the right hand side of the above equation).

In the next chapter of this thesis, where we use stochastic differential equations to discuss thermodynamics in a chaos bath at the level of individual trajectories, we will discuss the physical meaning of the terms appearing in Eq. (3.75) relation to the adiabatic reaction forces and fluctuation forces of Berry, Robbins, Wilkinson, and Jarzynski [48, 68, 77]. As we will show, to interpret these terms in a physically meaningful manner, one must consider the Stratonovich interpretation of the Fokker-Planck equation and the corresponding stochastic differential equations. An additional force, the noise induced drift appears, only in the Stratonovich interpretation, and explicit expression for the rapidly fluctuating force can only be found examining the stochastic differential equations corresponding to Eq. (3.75). At this point, we merely mention that the adiabatic (or Born-Oppenheimer) force, deterministic frictional force, and geomagnetic magnetic force associated with the αth bath are given by $\mathbf{u}^\alpha$, $-\frac{1}{2\Sigma^\alpha} \partial_{E^\alpha} \left[\Sigma^\alpha L^{\text{sy}\,\alpha} \right] M^{-1} \mathbf{P}$, and $-\frac{1}{2\Sigma^\alpha} \partial_{E^\alpha} \left[\Sigma^\alpha L^{\text{ay}\,\alpha} \right] M^{-1} \mathbf{P}$, respectively, where $L^{\text{ay}\,\alpha}$

denotes the anti-symmetric part of L^α. The fluctuations are implied by the diffusion terms (the second order derivatives) in Eq. (3.75), and we thus see an emergent fluctuation-dissipation relation involving the symmetric part of diffusion matrix and the strength of the deterministic frictional force [83]. This relation was previously noted by Wilkinson, Berry, Robbins, and Jarzynski for the case of a single bath [48, 68, 77].

3.4 Thermal equilibrium

In this section, we find the form of the stationary solution to Eq. (3.74) and discuss its relation to the notion of thermodynamic equilibrium. First, we verify when the control parameters $\boldsymbol{\lambda}$ are held fixed, a stationary distribution for Eq. (3.74) indeed exists, and we show that the form of this stationary distribution implies a particular statistical sharing of the total energy of the universe between the baths and system of interest. We then show that any time-dependent solution to Eq. (3.74) monotonically approaches (in the sense of relative entropy) the stationary distribution in time. This fact, along with the statistical sharing of energy, implies that the relaxation process is analogous to the thermalization of a small system with a large thermal environment, and so the stationary distribution is analogous to a thermal equilibrium distribution. We next discuss the Shannon entropy of the stationary distribution in terms of thermodynamic entropy, and we close this section by showing that, upon coarse graining the bath states, the stationary distribution becomes a Boltzmann distribution in a particular thermodynamic limit.

3.4.1 Stationary distribution

Consider the time-independent distribution W_s, defined by

$$W_s(\mathbf{X}, \mathbf{E}, \boldsymbol{\lambda}) \;=\; \mathcal{G}(E^U(\mathbf{X}, \mathbf{E}, \boldsymbol{\lambda})) \prod_\alpha \Sigma^\alpha(E^\alpha, \mathbf{Q}), \tag{3.76}$$

where $E^U(\mathbf{X}, \mathbf{E}, \boldsymbol{\lambda})$ denotes the total energy of the entire system plus baths universe:

$$
\begin{aligned}
E^U(\mathbf{X}, \mathbf{E}, \boldsymbol{\lambda}) &= H(\mathbf{X}, \boldsymbol{\lambda}) + \sum_\alpha E^\alpha \qquad (3.77) \\
&= \frac{1}{2}\mathbf{P}^T M^{-1}\mathbf{P} + V(\mathbf{Q}, \boldsymbol{\lambda}) + \sum_\alpha E^\alpha,
\end{aligned}
$$

and $\mathcal{G}(E^U)$ is an arbitrary fixed distribution over the total energy of the universe. Note the following identities:

$$
\begin{aligned}
\partial_\mathbf{Q} \mathcal{G}(E^U(\mathbf{X}, \mathbf{E}, \boldsymbol{\lambda})) &= \partial_\mathbf{Q} V(\mathbf{Q}, \boldsymbol{\lambda})\, \partial_{E^U} \mathcal{G}(E^U)\Big|_{E^U = H(\mathbf{X}, \boldsymbol{\lambda}) + \sum_\alpha E^\alpha}, \qquad (3.78) \\
\partial_\mathbf{P} \mathcal{G}(E^U(\mathbf{X}, \mathbf{E}, \boldsymbol{\lambda})) &= M^{-1}\mathbf{P}\, \partial_{E^U} \mathcal{G}(E^U)\Big|_{E^U = H(\mathbf{X}, \boldsymbol{\lambda}) + \sum_\alpha E^\alpha}, \\
\partial_{E^\beta} \mathcal{G}(E^U(\mathbf{X}, \mathbf{E}, \boldsymbol{\lambda})) &= \partial_{E^U} \mathcal{G}(E^U)\Big|_{E^U = H(\mathbf{X}, \boldsymbol{\lambda}) + \sum_\alpha E^\alpha}.
\end{aligned}
$$

To show that W_s as defined in Eq. (3.76) is indeed a stationary solution to the Fokker-Planck equation, we must verify

$$
\begin{aligned}
0 &= -\partial_\mathbf{Q} \cdot \left[M^{-1}\mathbf{P} W_s \right] - \partial_\mathbf{P} \cdot \left[-\partial_\mathbf{Q} V\, W_s \right] - \sum_\alpha \hat{\mathbf{D}}^\alpha \cdot \left[\mathbf{u}^\alpha W_s \right] \qquad (3.79) \\
&\quad + \frac{1}{2}\sum_{\alpha,i,j} \hat{D}_i^\alpha \left[\Sigma^\alpha L_{ij}^\alpha \hat{D}_j^\alpha \left(\frac{W_s}{\Sigma^\alpha} \right) \right].
\end{aligned}
$$

Making use of the identities in Eqs. (3.17), (3.18) and (3.78), the first three terms on the right hand side of Eq. (3.79) can be written as

$$-\partial_{\mathbf{Q}} \cdot [M^{-1}\mathbf{P}W_s] = -\left[\prod_{\beta}\Sigma^{\beta}\right]\mathbf{P}^T M^{-1}\partial_{\mathbf{Q}}V \, \partial_{E_u}\mathcal{G} \tag{3.80}$$

$$-\mathcal{G}\mathbf{P}^T M^{-1}\sum_{\alpha}\partial_{E^{\alpha}}\left[\mathbf{u}^{\alpha}\left[\prod_{\beta}\Sigma^{\beta}\right]\right],$$

$$-\partial_{\mathbf{P}} \cdot [-\partial_{\mathbf{Q}}V \, W_s] = \left[\prod_{\beta}\Sigma^{\beta}\right]\mathbf{P}^T M^{-1}\partial_{\mathbf{Q}}V \, \partial_{E_u}\mathcal{G},$$

$$\sum_{\alpha}\hat{\mathbf{D}}^{\alpha} \cdot [\mathbf{u}^{\alpha}W_s] = \mathcal{G}\mathbf{P}^T M^{-1}\sum_{\alpha}\partial_{E^{\alpha}}\left[\mathbf{u}^{\alpha}\left[\prod_{\beta}\Sigma^{\beta}\right]\right],$$

respectively. The sum of the above three thus cancel, and we are left to verify that the last term in Eq. (3.79) vanishes. Note that this term can be written as

$$\frac{1}{2}\sum_{\alpha,i,j}\hat{D}_i^{\alpha}\left[\Sigma^{\alpha}L_{ij}^{\alpha}\hat{D}_j^{\alpha}\left(\frac{W_s}{\Sigma^{\alpha}}\right)\right] = \frac{1}{2}\sum_{\alpha,i,j}\hat{D}_i^{\alpha}\left[\left(\prod_{\beta}\Sigma^{\beta}\right)L_{ij}^{\alpha}\hat{D}_j^{\alpha}\mathcal{G}\right]. \tag{3.81}$$

The term $\hat{D}_j^{\alpha}\mathcal{G}$ can be shown to vanish by employing the identities in Eq. (3.78):

$$\hat{D}_j^{\alpha}\mathcal{G} = \partial_{P_j}\mathcal{G} - M_{jj}^{-1}P_j\partial_{E^{\alpha}}\mathcal{G} \tag{3.82}$$

$$= M_{jj}^{-1}P_j\partial_{E_u}\mathcal{G} - M_{jj}^{-1}P_j\partial_{E_u}\mathcal{G}$$

$$= 0.$$

All terms in Eq. (3.79) vanish, so we have verified that W_s is indeed a stationary solution to the Fokker-Planck equation. The microcanonical partition function $\Sigma^{\alpha}(E^{\alpha}, \mathbf{Q})$ is a measure of the number of ways in which bath α can have energy E^{α} when the system of interest is located at $\mathbf{Q}$, so the form of Eq. (3.76) tells us that, when the system of interest and baths are distributed according to W_s, if the total energy of the universe is found to be $\mathcal{E}$, then the probability for the system of interest to be found with canonical coordinates

$\mathbf{X}$ is proportional to the total number of ways in which the baths can jointly share the remaining energy: $\mathcal{E} - H(\mathbf{X}, \boldsymbol{\lambda})$. Thus, W_s represents a state in which the total energy of the universe is shared statistically, with the statistics of a particular partitioning of energy being determined by the number of phase space configurations which allow that partition. Such a statistical sharing leads to the canonical thermal equilibrium distribution in statistical mechanics, so in this sense, the distribution W_s is analogous to a thermalized equilibrium state.

3.4.2 Thermalization

We have shown that the state W_s is a stationary solution to the Fokker-Planck equation and is reminiscent of a thermal equilibrium state, but in order to legitimately use the term "thermalization," we must show that any time-dependent solution $W(\mathbf{X}, \mathbf{E}, t)$ will continuously and monotonically approach W_s when $\boldsymbol{\lambda}$ is held fixed. To show that this is indeed the case, we will consider the time derivative of the relative entropy $\mathbb{D}(W_t || W_s)$, which is defined as the Kullback-Leibler divergence:

$$\mathbb{D}(W_t || W_s) \;=\; \int \mathrm{d}\mathbf{X} \mathrm{d}\mathbf{E} \, W(\mathbf{X}, \mathbf{E}, t) \ln \left[\frac{W(\mathbf{X}, \mathbf{E}, t)}{W_s(\mathbf{X}, \mathbf{E}, \boldsymbol{\lambda})} \right], \tag{3.83}$$

where we use W_t to abbreviate $W(\mathbf{X}, \mathbf{E}, t)$ for notational simplicity. This relative entropy is non-negative and only vanishes when $W_t = W_s$ almost everywhere with respect to W_t [84]. Taking a time derivative gives

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} \mathbb{D}(W_t || W_s) \;&=\; \int \mathrm{d}\mathbf{X} \mathrm{d}\mathbf{E} \left(\partial_t W_t \ln \left[\frac{W_t}{W_s} \right] + \partial_t W_t \right) \\ &=\; \int \mathrm{d}\mathbf{X} \mathrm{d}\mathbf{E} \, \partial_t W_t \ln \left[\frac{W_t}{W_s} \right]. \end{aligned} \tag{3.84}$$

The second line of the above equation follows from the fact that W_t is a normalized to the constant unity. From Eq. (3.74), we see that Eq. (3.84) is comprised of four terms:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{D}(W_t\|W_s) = \int \mathrm{d}\mathbf{X}\mathrm{d}\mathbf{E}\left[-\partial_\mathbf{Q}\cdot\left[M^{-1}\mathbf{P}W_t\right]\ln\left[\frac{W_t}{W_s}\right]\right.$$
$$-\partial_\mathbf{P}\cdot\left[-\partial_\mathbf{Q}V\,W_t\right]\ln\left[\frac{W_t}{W_s}\right]-\sum_\alpha\hat{\mathbf{D}}^\alpha\cdot\left[\mathbf{u}^\alpha W_t\right]\ln\left[\frac{W_t}{W_s}\right]$$
$$\left.+\frac{1}{2}\sum_{\alpha,i,j}\hat{D}_i^\alpha\left[\Sigma^\alpha L_{ij}^\alpha\hat{D}_j^\alpha\left(\frac{W_t}{\Sigma^\alpha}\right)\right]\ln\left[\frac{W_t}{W_s}\right]\right].$$

(3.85)

Integrating the first term on the right hand side of Eq. (3.85) by parts, noting that that W_t and W_s vanish at infinity and are normalized to constants, yields,

$$\int \mathrm{d}\mathbf{X}\mathrm{d}\mathbf{E}-\partial_\mathbf{Q}\cdot\left[M^{-1}\mathbf{P}W_t\right]\ln\left[\frac{W_t}{W_s}\right] = \int \mathrm{d}\mathbf{X}\mathrm{d}\mathbf{E}W_s\partial_\mathbf{Q}\left[\frac{W_t}{W_s}\right]\cdot M^{-1}\mathbf{P}$$
$$= -\int \mathrm{d}\mathbf{X}\mathrm{d}\mathbf{E}\frac{W_t}{W_s}\mathbf{P}^TM^{-1}\partial_\mathbf{Q}W_s.$$

(3.86)

Using the definition $W_s = \mathcal{G}\prod_\alpha\Sigma^\alpha$ from Eq. (3.76), as well as the identity in Eq. (3.17), the above expression can be written as

$$-\int \mathrm{d}\mathbf{X}\mathrm{d}\mathbf{E}\frac{W_t}{W_s}\mathbf{P}^TM^{-1}\partial_\mathbf{Q}W_s = \int \mathrm{d}\mathbf{X}\mathrm{d}\mathbf{E}\left[-\frac{W_t}{W_s}\mathbf{P}^TM^{-1}\partial_\mathbf{Q}V\,\partial_{E^u}\mathcal{G}\prod_\beta\Sigma^\beta\right.$$
$$-\sum_\alpha\mathbf{P}^TM^{-1}W_t\partial_{E^\alpha}\mathbf{u}^\alpha$$
$$\left.-\frac{W_t}{W_s}\sum_\alpha\mathbf{P}^TM^{-1}\mathbf{u}^\alpha\mathcal{G}\partial_{E^\alpha}\prod_\beta\Sigma^\beta\right],$$

(3.87)

where $\mathcal{G}$ is evaluated at $E^U = H(\mathbf{X},\boldsymbol{\lambda})+\sum_\alpha E^\alpha$. By similar considerations, the second term on the right hand side of Eq. (3.85) can be written as

$$- \int d\mathbf{X} d\mathbf{E} \, \partial_{\mathbf{P}} \cdot [-\partial_{\mathbf{Q}} V \, W_t] \ln \left[\frac{W_t}{W_s} \right] = \int d\mathbf{X} d\mathbf{E} \frac{W_t}{W_s} \mathbf{P}^T M^{-1} \partial_{\mathbf{Q}} V \, \partial_{E^u} \mathcal{G} \prod_{\beta} \Sigma^{\beta},$$

(3.88)

and the third term can be written as

$$- \int d\mathbf{X} d\mathbf{E} \sum_{\alpha} \hat{\mathbf{D}}^{\alpha} \cdot [\mathbf{u}^{\alpha} W_t] \ln \left[\frac{W_t}{W_s} \right] = \int d\mathbf{X} d\mathbf{E} \sum_{\alpha} \left[\mathbf{P}^T M^{-1} W_t \partial_{E^{\alpha}} \mathbf{u}^{\alpha} \right.$$
$$\left. + \frac{W_t}{W_s} \mathbf{P}^T M^{-1} \mathbf{u}^{\alpha} \mathcal{G} \partial_{E^{\alpha}} \prod_{\beta} \Sigma^{\beta} \right].$$

(3.89)

We thus see that the first three terms on the right hand side of Eq. (3.85) sum to zero, so only the fourth term remains. Integrating by parts again and rearranging yields

$$\frac{d}{dt} \mathbb{D}(W_t \| W_s) = -\frac{1}{2} \int d\mathbf{X} d\mathbf{E} \sum_{\alpha} \frac{(\Sigma^{\alpha})^2}{W_t} \left[\hat{\mathbf{D}}^{\alpha} \frac{W_t}{\Sigma^{\alpha}} \right]^T L^{\text{sy} \, \alpha} \left[\hat{\mathbf{D}}^{\alpha} \frac{W_t}{\Sigma^{\alpha}} \right]$$
$$+ \frac{1}{2} \int d\mathbf{X} d\mathbf{E} \sum_{\alpha} \Sigma^{\alpha} \left[\hat{\mathbf{D}}^{\alpha} \mathcal{G} \right]^T L^{\alpha} \left[\hat{\mathbf{D}}^{\alpha} \frac{W_t}{\Sigma^{\alpha}} \right].$$

(3.90)

Note that $L^{\text{sy} \, \alpha}$ appears in the first integral in the above equation because the quadratic form eliminates any contribution from $L^{\text{ay} \, \alpha}$. The term $\hat{\mathbf{D}}^{\alpha} \mathcal{G}$ vanishes as in Eq. (3.82), so we have

$$\frac{d}{dt} \mathbb{D}(W_t \| W_s) = -\frac{1}{2} \int d\mathbf{X} d\mathbf{E} \sum_{\alpha} \frac{(\Sigma^{\alpha})^2}{W_t} \left[\hat{\mathbf{D}}^{\alpha} \frac{W_t}{\Sigma^{\alpha}} \right]^T L^{\text{sy} \, \alpha} \left[\hat{\mathbf{D}}^{\alpha} \frac{W_t}{\Sigma^{\alpha}} \right].$$

(3.91)

We thus see that $\frac{d}{dt} \mathbb{D}(W_t \| W_s) \le 0$, and that any W_t approaches W_s monotonically in time, whenever $L^{\text{sy} \, \alpha}$ is non-negative for each α.

We now make a heuristic argument for the non-negativity of $L^{\text{sy} \, \alpha}$ for strongly chaotic

systems. Employing the definition $L^{\text{sy}\,\alpha} = \frac{1}{2}\big[L^\alpha + (L^\alpha)^T\big]$, and making use of the time translation invariance of correlation functions for stationary ergodic processes [11], we find from Eq. (3.70)

$$
\begin{aligned}
L^{\text{sy},\alpha} &= \int_{-\infty}^0 dt' \left\langle \big(-\partial_{\mathbf{Q}} h^\alpha - \mathbf{u}^\alpha\big)\big(-\partial_{\mathbf{Q}} h^\alpha_{t'} - \mathbf{u}^\alpha\big)^T \right\rangle_{E^\alpha,\mathbf{Q}} \\
&\quad + \int_{-\infty}^0 dt' \left\langle \big(-\partial_{\mathbf{Q}} h^\alpha_{t'} - \mathbf{u}^\alpha\big)\big(-\partial_{\mathbf{Q}} h^\alpha - \mathbf{u}^\alpha\big)^T \right\rangle_{E^\alpha,\mathbf{Q}} \\
&= \int_{-\infty}^0 dt' \left\langle \big(-\partial_{\mathbf{Q}} h^\alpha - \mathbf{u}^\alpha\big)\big(-\partial_{\mathbf{Q}} h^\alpha_{t'} - \mathbf{u}^\alpha\big)^T \right\rangle_{E^\alpha,\mathbf{Q}} \\
&\quad + \int_{-\infty}^0 dt' \left\langle \big(-\partial_{\mathbf{Q}} h^\alpha - \mathbf{u}^\alpha\big)\big(-\partial_{\mathbf{Q}} h^\alpha_{-t'} - \mathbf{u}^\alpha\big)^T \right\rangle_{E^\alpha,\mathbf{Q}} \\
&= \int_{-\infty}^\infty dt' \left\langle \big(-\partial_{\mathbf{Q}} h^\alpha - \mathbf{u}^\alpha\big)\big(-\partial_{\mathbf{Q}} h^\alpha_{t'} - \mathbf{u}^\alpha\big)^T \right\rangle_{E^\alpha,\mathbf{Q}}.
\end{aligned}
\tag{3.92}
$$

For future reference, we note that similar manipulations applied to $L^{\text{ay}\,\alpha} = \frac{1}{2}\big[L^\alpha - (L^\alpha)^T\big]$ yield

$$
\begin{aligned}
L^{\text{ay}\,\alpha} &= \int_{-\infty}^0 dt' \left\langle \big(-\partial_{\mathbf{Q}} h^\alpha - \mathbf{u}^\alpha\big)\big(-\partial_{\mathbf{Q}} h^\alpha_{t'} - \mathbf{u}^\alpha\big)^T \right\rangle_{E^\alpha,\mathbf{Q}} \\
&\quad - \int_0^\infty dt' \left\langle \big(-\partial_{\mathbf{Q}} h^\alpha - \mathbf{u}^\alpha\big)\big(-\partial_{\mathbf{Q}} h^\alpha_{t'} - \mathbf{u}^\alpha\big)^T \right\rangle_{E^\alpha,\mathbf{Q}}.
\end{aligned}
\tag{3.93}
$$

Now define the matrix $C^\alpha(t)$ by

$$
C^\alpha(t) = \left\langle \big(-\partial_{\mathbf{Q}} h^\alpha - \mathbf{u}^\alpha\big)\big(-\partial_{\mathbf{Q}} h^\alpha_t - \mathbf{u}^\alpha\big)^T \right\rangle_{E^\alpha,\mathbf{Q}}.
\tag{3.94}
$$

Note that C^α is also a function of $\mathbf{Q}$ and E^α, but we suppress this functional dependence for simplicity. Also note that $C(0)$ is a genuine non-negative covariance matrix. The matrix $L^{\text{sy}\,\alpha}$ can thus be written as

$$L^{\text{sy }\alpha} = \int_{-\infty}^{\infty} \mathrm{d}t'\, C^{\alpha}(t'). \tag{3.95}$$

We now assume that for each bath, when the system of interest is held fixed, correlations in observables of the bath's state decay exponentially fast with time, and we assume that the correlation time-scale is much smaller than all other time-scales associated with the bath. We thus suppose that $C^{\alpha}(t)$ can be written as

$$C^{\alpha}(t) = f^{\alpha}(t)e^{-\frac{|t|}{\tau^{\alpha}}}, \tag{3.96}$$

where τ^{α} denotes the correlation time of the αth bath, $\ln\left(|f^{\alpha}(t)|\right)$ grows slower that $|t|$ as $|t|$ approaches infinity, and the natural time-scale associated with f^{α} is much larger than τ^{α}. Note that $C^{\alpha}(0) = f^{\alpha}(0)$. The matrix f^{α} will thus change very little over times that scale like τ^{α}, and integrand in Eq. (3.95) will make negligible contributions to the integral for times much greater than τ^{α}, so we may expand f in a power series and write

$$L^{\text{sy }\alpha} = \int_{-\infty}^{\infty} \mathrm{d}t' \left[C^{\alpha}(0) + \sum_{n=0}^{\infty} \frac{1}{n!}(f^{\alpha})^{(n)}(0)\, t'^{n} \right] e^{-\frac{|t|}{\tau^{\alpha}}}. \tag{3.97}$$

The odd powers of t' in the above equation will vanish, and integrating the even powers gives

$$L^{\text{sy }\alpha} = 2\tau^{\alpha} \left[C^{\alpha}(0) + \sum_{n=1}^{\infty} (f^{\alpha})^{(2n)}(0)\, (\tau^{\alpha})^{2n} \right] \tag{3.98}$$

The natural time-scale associated with f^{α} is much greater than τ^{α}, so the infinite sum in the above equation will be negligible compared to $C^{\alpha}(0)$, and we find

$$
\begin{aligned}
L^{\mathrm{sy}\,\alpha} &\approx 2\tau^{\alpha}C^{\alpha}(0) && (3.99) \\
&= 2\tau^{\alpha}\left\langle \left(-\partial_{\mathbf{Q}}h^{\alpha}-\mathbf{u}^{\alpha}\right)\left(-\partial_{\mathbf{Q}}h^{\alpha}-\mathbf{u}^{\alpha}\right)^{T}\right\rangle_{E^{\alpha},\mathbf{Q}}.
\end{aligned}
$$

The above expression is a non-negative definite covariance matrix, so we see that for a strongly chaotic bath, $L^{\mathrm{sy}\alpha}$ will be dominated by a non-negative matrix, with possible non-non-negative corrections being negligible.

In this subsection, we have shown, given the presence strong chaos at the micro-scale, any time-dependent solution to the Fokker Planck equation (3.74) will continually relax towards the form of the stationary distribution given by Eq. (3.76). Because W_s represents a statistical sharing of energy, with the statistics determined by the number of allowed phase space configurations, we conclude that the relaxation towards W_s provides closed system chaos bath analog to the relaxation of an open system to a thermal equilibrium distribution. The system of interest thus "thermalizes" with the chaos baths in the same manner in which a small open system thermalizes with the surrounding environment. This observation was previously made by Jarzynski for the case of a single bath [68].

We close this subsection by noting that we have not shown that the form of W_s in Eq. (3.76) gives a unique stationary solution to the Fokker-Planck equation, so we can not say for certain whether or not all initial distributions will actually reach W_s in the long-time limit. We have shown that $\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{D}(W_t||W_s) \leq 0$, regardless of the uniqueness of the stationary distribution, but we have not shown that $\mathbb{D}(W_t||W_s) \to 0$ as $t \to \infty$: the relative entropy between W_t and W_s may monotonically decrease to some non-zero positive number. If $\mathbb{D}(W_t||W_s)$ is observed to converge to a positive number, then W_t must be converging towards some other stationary distribution W_s', distinct from W_s on some set of non-zero measure with respect to W_s'. Regardless of the uniqueness of the form of stationary solution to the Fokker-Planck equation (3.74), however, in the next subsection, we will show that the form of the stationary given in Eq. (3.76) is unique in the sense that it follows from a microcanonical stationary solution at the microscale.

3.4.3 Entropy

For the remainder of this thesis chapter, we will assume that the system of interest and baths together have a well-defined constant total energy $\mathcal{E}$ such that the distribution $\mathcal{G}$ in Eq. (3.76) is a Dirac delta distribution centred at $\mathcal{E}$. The total energy of the universe will remain delta distributed if the control parameters $\boldsymbol{\lambda}$ are held fixed, and we denote this particular stationary distribution by

$$W_s(\mathbf{X}, \mathbf{E}; \mathcal{E}, \boldsymbol{\lambda}) = \frac{\delta(\mathcal{E} - H(\mathbf{X}, \boldsymbol{\lambda}) - \sum_\alpha E^\alpha) \prod_\alpha \Sigma^\alpha(E^\alpha, \mathbf{Q})}{\tilde{\Sigma}^U(\mathcal{E}, \boldsymbol{\lambda})}, \tag{3.100}$$

where the partition function $\tilde{\Sigma}^U(\mathcal{E}, \boldsymbol{\lambda})$ is a normalization factor. We now show that $\tilde{\Sigma}^U(\mathcal{E}, \boldsymbol{\lambda})$ is equal to the microcanonical partition of the entire universe.

Consider the microcanonical microscopic stationary distribution ϕ_s, given by

$$\phi_s(\mathbf{X}, \mathbb{X}; \mathcal{E}, \boldsymbol{\lambda}) = \frac{\delta(\mathcal{E} - H(\mathbf{X}, \boldsymbol{\lambda}) - \sum_\alpha h^\alpha(\mathbf{x}^\alpha; \mathbf{Q}))}{\Sigma^U(\mathcal{E}, \boldsymbol{\lambda})}, \tag{3.101}$$

where the normalization factor $\Sigma^U(\mathcal{E}, \boldsymbol{\lambda})$, is the microcanonical partition function of the universe:

$$\Sigma^U(\mathcal{E}, \boldsymbol{\lambda}) = \int d\mathbf{X} d\mathbb{X} \delta(\mathcal{E} - H(\mathbf{X}, \boldsymbol{\lambda}) + \sum_\alpha h^\alpha(\mathbf{x}^\alpha; \mathbf{Q})). \tag{3.102}$$

Projecting the bath variables onto their energy shells, we find

$$\langle \phi_s \rangle_{\mathbf{E}, \mathbf{Q}} = \int d\mathbb{X} \, \phi_s(\mathbf{X}, \mathbb{X}; \mathcal{E}, \boldsymbol{\lambda}) \prod_\alpha \delta(E^\alpha - h^\alpha(\mathbf{x}^\alpha; \mathbf{Q})) \tag{3.103}$$

$$= \frac{\delta(\mathcal{E} - H(\mathbf{X}, \boldsymbol{\lambda}) - \sum_\alpha E^\alpha) \prod_\alpha \Sigma^\alpha(E^\alpha, \mathbf{Q})}{\Sigma^U(\mathcal{E}, \boldsymbol{\lambda})}.$$

Both $\langle \phi_s \rangle_{\mathbf{E}, \mathbf{Q}}$ and W_s as given in Eq. (3.100) are normalized to unity, so we conclude that $\tilde{\Sigma}^U = \Sigma^U$, and that W_s follows from an underlying microcanonical microscopic stationary

distribution.

Following the standard correspondence between statistical mechanics and thermodynamics, [70] the thermodynamic entropy of the universe, denoted by $S^U(\mathcal{E}, \boldsymbol{\lambda})$, is given by the microcanonical entropy of the universe:

$$S^U(\mathcal{E}, \boldsymbol{\lambda}) = k_B \ln \Sigma^U(\mathcal{E}, \boldsymbol{\lambda}), \tag{3.104}$$

where k_B is Boltzmann's constant. We are working in an approximation which assumes that each bath remains microcanonically distributed over the energy shells which they occupy in phase space, so thermodynamic entropy of each individual bath, s^α, is the corresponding microcanonical entropy:

$$s^\alpha(E^\alpha, \mathbf{Q}) = k_B \ln \Sigma^\alpha(E^\alpha, \mathbf{Q}), \tag{3.105}$$

Note that we are being sloppy with units in Eqs. (3.104) and (3.105): we should be including factors which divide the microcaonical partition functions in order to make the arguments of the natural logs unitless. Such factors will only shift the entropies by over all constants, so we disregard them and similar factors throughout. In terms of entropies, the stationary distribution can be written as

$$W_s(\mathbf{X}, \mathbf{E}; \mathcal{E}, \boldsymbol{\lambda}) = \delta(\mathcal{E} - H(\mathbf{X}, \boldsymbol{\lambda}) - \sum_\alpha E^\alpha) e^{\frac{\sum_\alpha s^\alpha(E^\alpha, \mathbf{Q}) - S^U(\mathcal{E}, \boldsymbol{\lambda})}{k_B}}. \tag{3.106}$$

We define the entropy of a stationary solution to the Fokker-Planck equation to be the distribution's differential Shannon Entropy multiplied by Boltzmann's constant. The entropy of W_s, denoted by S^W, is thus

$$S^W(\mathcal{E}, \boldsymbol{\lambda}) = -k_B \int d\mathbf{X} d\mathbf{E} \, W_s \ln[W_s] \tag{3.107}$$

$$= S^U(\mathcal{E}, \boldsymbol{\lambda}) - \sum_\alpha \langle s^\alpha \rangle_{\mathcal{E}, \boldsymbol{\lambda}}^W - k_B \langle \ln \delta(0) \rangle_{\mathcal{E}, \boldsymbol{\lambda}}^W,$$

where $\langle ... \rangle_{\mathcal{E}, \boldsymbol{\lambda}}^W$ denotes an average over W_s. The delta function is a divergence which reflects the fact that W_s is a singular distribution constrained to lie on a single surface of constant $H + \sum_\alpha E^\alpha$ in $(\mathbf{X}, \mathbf{E})$ space: the entirety of the probability measure lies on a surface of Lebesgue measure zero. This divergence simply indicates the presence of a constraint, and the associated infinite information is physically irrelevant information which can be ignored. More formally, noting that $(\mathbf{X}, \mathbf{E})$ over-specifies the state of the universe by one degree of freedom, we can define a physically equivalent, non-singular, reduced distribution $\tilde{W}_s$ by tracing out one of the bath variables from W_s, say bath number one for example:

$$\tilde{W}_s(\mathbf{X}, E^2, ..., E^N; \mathcal{E}, \boldsymbol{\lambda}) = \int dE^1 W_s(\mathbf{X}, \mathbf{E}; \mathcal{E}, \boldsymbol{\lambda}) \tag{3.108}$$

$$= e^{\frac{s^1(\mathcal{E} - H(\mathbf{X}, \boldsymbol{\lambda}) - \sum_{\alpha \neq 1} E^\alpha, \mathbf{Q}) + \sum_{\alpha \neq 1} s^\alpha(E^\alpha, \mathbf{Q}) - S^U(\mathcal{E}, \boldsymbol{\lambda})}{k_B}}.$$

The entropy of this physically equivalent reduced distribution, given by,

$$-k_B \int d\mathbf{X} dE^2 ... dE^N \, \tilde{W}_s \ln \tilde{W}_s = S^U(\mathcal{E}, \boldsymbol{\lambda}) - \sum_\alpha \langle s^\alpha \rangle_{\mathcal{E}, \boldsymbol{\lambda}}^W, \tag{3.109}$$

is equal to the physically meaningful part of the entropy of the full distribution. The thermodynamic entropy of the universe can thus be written the sum of the entropy of W_s and the average thermodynamic entropies of the individual baths:

$$S^U(\mathcal{E}, \boldsymbol{\lambda}) = S^W(\mathcal{E}, \boldsymbol{\lambda}) + \sum_\alpha \langle s^\alpha \rangle_{\mathcal{E}, \boldsymbol{\lambda}}^W. \tag{3.110}$$

3.4.4 Coarse grained bath states

Consider now the distribution W_B which results from projecting the individual bath energies out of W_s onto surfaces of constant total bath energy E^B:

$$
\begin{aligned}
W_B(\mathbf{X}, E^B; \mathcal{E}, \boldsymbol{\lambda}) &= \int d\mathbf{E}\, \delta(E^B - \sum_\alpha E^\alpha) W_s(\mathbf{X}, \mathbf{E}; \mathcal{E}, \boldsymbol{\lambda}) \qquad (3.111) \\
&= \int d\mathbf{E}\, \delta(E^B - \sum_\alpha E^\alpha) \frac{\delta(\mathcal{E} - H(\mathbf{X}, \boldsymbol{\lambda}) - \sum_\alpha E^\alpha) \prod_\alpha \Sigma^\alpha(E^\alpha, \mathbf{Q})}{\Sigma^U(\mathcal{E}, \boldsymbol{\lambda})}.
\end{aligned}
$$

This distribution is a coarse grained distribution relative to W_s - the bath energy states lumped together into states of total bath energy, with the statistical weight of a particular energy configuration within a coarse grained state given by $\prod_\alpha \Sigma^\alpha(E^\alpha, \mathbf{Q})$. In the next subsection, we will show that W_B reduces to a Boltzmann distribution in an appropriately defined thermodynamic limit. In this subsection, we examine the physical properties of W_B and derive some useful identities.

Denote the microcanonical partition function associated with the microcanonical distribution in the joint bath phase space by Σ^B:

$$
\Sigma^B(E^B, \mathbf{Q}) = \int d\mathbb{X}\, \delta\big(E^B - \sum_\alpha h^\alpha(\mathbf{x}^\alpha; \mathbf{Q})\big). \qquad (3.112)
$$

This quantity gives a measure of the total number of ways in which the baths can jointly share an energy E_B when the system of interest is located at $\mathbf{Q}$. This partition function can also be written as

$$
\begin{aligned}
\Sigma^B(E^B, \mathbf{Q}) &= \int d\mathbb{X} d\mathbf{E}\, \delta(E^B - \sum_\alpha E^\alpha) \prod_\alpha \delta(E^\alpha - h^\alpha(\mathbf{x}^\alpha; \mathbf{Q})) \qquad (3.113) \\
&= \int d\mathbf{E}\, \delta(E^B - \sum_\alpha E^\alpha) \prod_\alpha \Sigma^\alpha(E^\alpha, \mathbf{Q}).
\end{aligned}
$$

Using this expression for Σ^B, W_B can be written as

$$W_B(\mathbf{X}, E^B; \mathcal{E}, \boldsymbol{\lambda}) \quad = \quad \frac{\delta(\mathcal{E} - H(\mathbf{X}, \boldsymbol{\lambda}) - E^B)\,\Sigma^B(E^B, \mathbf{Q})}{\Sigma^U(\mathcal{E}, \boldsymbol{\lambda})}. \qquad (3.114)$$

This distribution is constrained to lie on a surface of constant $H(\mathbf{X}, \boldsymbol{\lambda}) + E^B$ and is therefore singular. A physically equivalent non-singular distribution, P_s, can be obtained by tracing out E_B:

$$
\begin{aligned}
P_s(\mathbf{X}; \mathcal{E}, \boldsymbol{\lambda}) \quad &= \quad \int \mathrm{d}E^B \, W_B(\mathbf{X}, E^B; \mathcal{E}, \boldsymbol{\lambda}) & (3.115) \\
&= \quad \frac{\Sigma^B(\mathcal{E} - H(\mathbf{X}, \boldsymbol{\lambda}), \mathbf{Q})}{\Sigma^U(\mathcal{E}, \boldsymbol{\lambda})}.
\end{aligned}
$$

This distribution must be normalized, so we find the identity

$$\Sigma^U(\mathcal{E}, \boldsymbol{\lambda}) \quad = \quad \int\limits_{0 \leq H(\mathbf{X}, \boldsymbol{\lambda}) \leq \mathcal{E}} \mathrm{d}\mathbf{X}\, \Sigma^B(\mathcal{E} - H(\mathbf{X}, \boldsymbol{\lambda}), \mathbf{Q}) \qquad (3.116)$$

Denoting by S^B the microcanonical entropy corresponding to Σ^B,

$$S^B(E^B, \mathbf{Q}) \quad = \quad k_B \ln \Sigma^B(E^B, \mathbf{Q}), \qquad (3.117)$$

the coarse grained stationary distribution can be written as

$$P_s(\mathbf{X}; \mathcal{E}, \boldsymbol{\lambda}) \quad = \quad e^{\frac{S^B(\mathcal{E} - H(\mathbf{X}, \boldsymbol{\lambda}), \mathbf{Q}) - S^U(\mathcal{E}, \boldsymbol{\lambda})}{k_B}}. \qquad (3.118)$$

The entropy of P_s, denoted by S^P, is given by

$$S^P(\mathcal{E}, \boldsymbol{\lambda}) = -k_B \int d\mathbf{X} \, P_s(\mathbf{X}; \mathcal{E}, \boldsymbol{\lambda}) \ln P_s(\mathbf{X}; \mathcal{E}, \boldsymbol{\lambda}) \tag{3.119}$$

$$= S^U(\mathcal{E}, \boldsymbol{\lambda}) - \langle S^B \rangle^P_{\mathcal{E}, \boldsymbol{\lambda}},$$

where $\langle .. \rangle^P_{\mathcal{E}, \boldsymbol{\lambda}}$ denotes an average over P_s. Inserting delta functions into Eq. (3.119) and making use of Eq. (3.113), we also have

$$S^P(\mathcal{E}, \boldsymbol{\lambda}) = -k_B \int d\mathbf{X} dE^B \, \delta(\mathcal{E} - H(\mathbf{X}, \boldsymbol{\lambda}) - E^B) \frac{\Sigma^B(E^B, \mathbf{Q})}{\Sigma^U(\mathcal{E}, \boldsymbol{\lambda})} \ln \left[\frac{\Sigma^B(E^B, \mathbf{Q})}{\Sigma^U(\mathcal{E}, \boldsymbol{\lambda})} \right]$$

$$= -k_B \int d\mathbf{X} dE^B d\mathbf{E} \, \delta(\mathcal{E} - H(\mathbf{X}, \boldsymbol{\lambda}) - \sum_\alpha E^\alpha) \delta(E_B - \sum_\alpha E^\alpha)$$

$$\times \frac{\prod_\alpha \Sigma^\alpha(E^\alpha, \mathbf{Q})}{\Sigma^U(\mathcal{E}, \boldsymbol{\lambda})} \ln \left[\frac{\Sigma^B(E^B, \mathbf{Q})}{\Sigma^U(\mathcal{E}, \boldsymbol{\lambda})} \right]$$

$$= S^U(\mathcal{E}, \boldsymbol{\lambda}) - \sum_\alpha \langle s^\alpha \rangle^W_{\mathcal{E}, \boldsymbol{\lambda}} + k_B \left\langle \ln \frac{\Sigma^B}{\prod_\alpha \Sigma^\alpha} \right\rangle^W_{\mathcal{E}, \boldsymbol{\lambda}}. \tag{3.120}$$

The last term on the final line of the right hand side of Eq. (3.120) is a measure of the entropy of P_s relative to W_s. This relative entropy characterizes the information which is lost when the individual bath energy configuration state are coarse grained into states of total bath energy. For the case of a single bath, we have $\Sigma^1 = \Sigma^B$, and the relative entropy will therefore vanish. This must be the case: no actual coarse graining occurs for the case of a single bath, so no information can be lost. Using Eq. (3.110), we find

$$S^P(\mathcal{E}, \boldsymbol{\lambda}) = S^W(\mathcal{E}, \boldsymbol{\lambda}) + k_B \left\langle \ln \frac{\Sigma^B}{\prod_\alpha \Sigma^\alpha} \right\rangle^W_{\mathcal{E}, \boldsymbol{\lambda}} \tag{3.121}$$

The relative entropy in the above equation is non-negative, so we see that $S^P \geq S^W$: the disorder associated with S^P is generally greater than the disorder associated with S^W due to the information lost as a result of the coarse graining.

Consider now the partial derivative of Σ^B with respect to E_B:

$$
\begin{aligned}
\partial_{E^B}\Sigma^B(E^B,\mathbf{Q}) &= \int d\mathbf{E}\, \partial_{E^B}\delta(E^B - \sum_\alpha E^\alpha)\prod_\alpha \Sigma^\alpha(E^\alpha,\mathbf{Q}) &\text{(3.122)}\\
&= \int d\mathbf{E}\,\delta(E^B - \sum_\alpha E^\alpha)\left[\prod_\alpha \Sigma^\alpha(E^\alpha,\mathbf{Q})\right]\frac{\partial_{E^\beta}\Sigma^\beta(E^\beta,\mathbf{Q})}{\Sigma^\beta(E^\beta,\mathbf{Q})}\\
&= \frac{1}{N}\int d\mathbf{E}\,\delta(E^B - \sum_\alpha E^\alpha)\prod_\alpha \Sigma^\alpha(E^\alpha,\mathbf{Q})\left[\sum_\beta \frac{\partial_{E^\beta}\Sigma^\beta(E^\beta,\mathbf{Q})}{\Sigma^\beta(E^\beta,\mathbf{Q})}\right].
\end{aligned}
$$

The second line of the above equation follows from integrating by parts and making use of the fact that, when operating on the delta function, $\partial_{E^B} = -\partial_{E^\beta}$ for any one of the N bath energies E^β, and the third line follows from applying this fact to all N bath energies. We thus have, for any β,

$$
\begin{aligned}
\frac{1}{\Sigma^B(E^B,\mathbf{Q})}\partial_{E^B}\Sigma^B(E^B,\mathbf{Q}) &= \left\langle \frac{1}{\Sigma^\beta}\partial_{E_\beta}\Sigma^\beta \right\rangle_{E^B,\mathbf{Q}} &\text{(3.123)}\\
&= \frac{1}{N}\sum_\alpha \left\langle \frac{1}{\Sigma^\alpha}\partial_{E_\alpha}\Sigma^\alpha \right\rangle_{E^B,\mathbf{Q}},
\end{aligned}
$$

where $\langle ... \rangle_{E^B,\mathbf{Q}}$ denotes an average over the microcanonical joint bath distribution when the baths share a total energy E^B and the system of interest is located at $\mathbf{Q}$. Invoking standard thermodynamic definitions [70], the microcanonical temperature T^B of the joint bath system is given by

$$
\begin{aligned}
\frac{1}{T^B(E^B,\mathbf{Q})} &= \partial_{E^B}S^B(E^B,\mathbf{Q}) &\text{(3.124)}\\
&= k_B\partial_{E^B}\ln\Sigma^B(E^B,\mathbf{Q}).
\end{aligned}
$$

The microcanonical temperatures of the individual baths are likewise given by

$$\frac{1}{T^\alpha(E^\alpha, \mathbf{Q})} = \partial_{E^\alpha} s^\alpha(E^\alpha, \mathbf{Q}) \tag{3.125}$$

$$= k_B \partial_{E^\alpha} \ln \Sigma^\alpha(E^\alpha, \mathbf{Q}),$$

so from Eq. (3.123), we see that the inverse joint bath temperature is related to the individual bath temperatures by

$$\frac{1}{T^B(E^B, \mathbf{Q})} = \left\langle \frac{1}{T^\beta} \right\rangle_{E^B, \mathbf{Q}}, \tag{3.126}$$

for each β. We note that the above temperatures are well-defined, regardless of the size of the baths - they are simply notational conveniences with no a priori physical interpretations. However, for baths with many, many degrees of freedom, we expect these temperatures to behave in accordance with our ordinary notions of temperature [70].

Finally, consider the partial derivative of Σ^B with respect to $\mathbf{Q}$:

$$\partial_\mathbf{Q} \Sigma^B(E^B, \mathbf{Q}) = \int d\mathbf{E}\, \delta(E^B - \sum_\alpha E^\alpha) \partial_\mathbf{Q} \prod_\alpha \Sigma^\alpha(E^\alpha, \mathbf{Q}) \tag{3.127}$$

$$= \frac{1}{N} \Sigma^B(E^B, \mathbf{Q}) \sum_\alpha \left\langle \frac{1}{\Sigma^\alpha} \partial_\mathbf{Q} \Sigma^\alpha \right\rangle_{E^B, \mathbf{Q}}$$

$$= \frac{1}{N} \Sigma^B(E^B, \mathbf{Q}) \sum_\alpha \left\langle \frac{1}{\Sigma^\alpha} \partial_{E^B} [\Sigma^\alpha \mathbf{u}^\alpha] \right\rangle_{E^B, \mathbf{Q}},$$

where we have made use of Eq. (3.17). The average total adiabatic force exerted on the system of interest, denoted by $\mathbf{u}^B$, is given by

$$\mathbf{u}^B(E^B, \mathbf{Q}) = \sum_\alpha \langle \mathbf{u}^\alpha \rangle_{E^B, \mathbf{Q}} \tag{3.128}$$

$$= \frac{1}{\Sigma^B(E^B, \mathbf{Q})} \int d\mathbf{E}\, \delta(E^B - \sum_\alpha E^\alpha) \prod_\alpha \Sigma^\alpha(E^\alpha, \mathbf{Q}) \left[\sum_\beta \mathbf{u}^\beta(E^\beta, \mathbf{Q}) \right].$$

Making use of Eqs. (3.122) and (3.127), we find the anolog of the useful identity in Eq. (3.17):

$$\partial_{E^B} \left[\Sigma^B(E^B, \mathbf{Q}) \mathbf{u}^B(E^B, \mathbf{Q}) \right] \;\; = \;\; \partial_{\mathbf{Q}} \Sigma^B(E^B, \mathbf{Q}). \qquad (3.129)$$

3.4.5   The thermodynamic limit

We now consider the stationary distribution in a limit where the number of joint bath degrees of freedom, N_B, becomes much larger than system of interest's number of degrees of freedom, N_S. The exact nature of this limit, which we refer to as the thermodynamic limit, we be specified shortly. This limit could correspond a very large number of baths, or to a few (or a single) baths with many, many degrees of freedom. We assume the special case where Σ^B, as a function of total bath energy, grows exponentially at a rate linearly dependent on N_B, such that S^B is extensive in N_B in the large N_B limit. We acknowledge that in order for S^B to be extensive, it may be required to modify definition in Eq. (3.117) by subtracting the log of the factorial of the number of *identical* bath systems, as per the Gibbs mixing paradox [70], but this modification will be inconsequential for what follows. Here, we are not concerned with determining exactly what classes of baths satisfy the extensivity assumption, but we do note that for some simple cases of practical interest, such as a bath of ideal gas particles, the assumption is indeed satisfied [70]. Under the extensivity assumption, as N_B grows larger and larger, states of very large total bath energy become more and more dominate in their contribution to Σ^U, and thus become the dominate states of non-negligible probability in Eq. (3.115).

The thermodynamic limit is defined as taking N_B to infinity such E^B/N_B approaches a constant while holding N_S constant and scaling the system of interest and baths such that $||\mathbf{u}^B||/N_B$ approaches zero. For the example pictured in Fig. 3.1, the thermodynamic limit corresponds to adding more and more point particles to the billiard interior while simultaneously increasing the size of the billiard such that the density of point particles remains constant. The total bath energy is thus extensive by definition in the thermodynamic limit. For argument's sake, we suppose that $H(\mathbf{X}, \boldsymbol{\lambda})$ is extensive in N_S, but all we really require

is that the system of interest's energy be held constant as more and more bath degrees of freedom are added in the thermodynamic limit.

As $N_B \to \infty$, states of large total bath energy (large compared to the energy of the system of interest) are the only statistically non-negligible states, so may assume that the total bath energy is always very close to the total energy of the universe. The energy of the universe is therefore extensive in N_B, and we may safely approximate the joint bath entropy $S^B(\mathcal{E} - H(\mathbf{X}, \boldsymbol{\lambda}), \mathbf{Q})$ in the thermodynamic limit by expanding it in a power series in its first argument, evaluated at $\mathcal{E}$:

$$
\begin{aligned}
S^B(\mathcal{E} - H(\mathbf{X}, \boldsymbol{\lambda}), \mathbf{Q}) &= S^B(\mathcal{E}, \mathbf{Q}) - \partial_\mathcal{E}[S^B(\mathcal{E}, \mathbf{Q})]H(\mathbf{X}, \boldsymbol{\lambda}) \\
&\quad + \frac{1}{2}\partial^2_{\mathcal{E}\mathcal{E}}[S^B(\mathcal{E}, \mathbf{Q})]H(\mathbf{X}, \boldsymbol{\lambda})^2 + \ldots \\
&= S^B(\mathcal{E}, \mathbf{Q}) - H(\mathbf{X}, \boldsymbol{\lambda})\left(\frac{1}{T^B(\mathcal{E}, \mathbf{Q})} + O\left(\frac{N_s}{N_b}\right)\right).
\end{aligned}
\tag{3.130}
$$

As $N_S/N_B \to 0$, we have

$$
S^B(\mathcal{E} - H(\mathbf{X}, \boldsymbol{\lambda}), \mathbf{Q}) = S^B(\mathcal{E}, \mathbf{Q}) - \frac{H(\mathbf{X}, \boldsymbol{\lambda})}{T^B(\mathcal{E}, \mathbf{Q})},
\tag{3.131}
$$

so the coarse grained stationary distribution in Eq. (3.115) can be written as

$$
\begin{aligned}
P_s(\mathbf{X}; \mathcal{E}, \boldsymbol{\lambda}) &\underset{N_B \to \infty}{=} e^{-\frac{H(\mathbf{X}, \boldsymbol{\lambda}) - T^B(\mathcal{E}, \mathbf{Q})S^B(\mathcal{E}, \mathbf{Q})}{k_B T^B(\mathcal{E}, \mathbf{Q})} - \frac{S^U(\mathcal{E}, \boldsymbol{\lambda})}{k_B}} \\
&= e^{-\frac{H(\mathbf{X}, \boldsymbol{\lambda}) - T^B(\mathcal{E}, \mathbf{Q})S^B(\mathcal{E}, \mathbf{Q}) - \mathcal{F}(\mathcal{E}, \boldsymbol{\lambda}, \mathbf{Q})}{k_B T^B(\mathcal{E}, \mathbf{Q})}},
\end{aligned}
\tag{3.132}
$$

where we define $\mathcal{F}(\mathcal{E}, \boldsymbol{\lambda}, \mathbf{Q})$ by

$$
\mathcal{F}(\mathcal{E}, \boldsymbol{\lambda}, \mathbf{Q}) = -T^B(\mathcal{E}, \mathbf{Q})S^U(\mathcal{E}, \boldsymbol{\lambda}).
\tag{3.133}
$$

Equation (3.132) resembles a canonical distribution with a position dependent temperature and free energy. This "free energy" does not appear to be a true free energy because, due

93

to the position dependent temperature, it is not a constant multiple of the log of a partition function. However, using the identity in Eq. (3.128), one can show that the gradient of T^B with respect to the system's position is given by

$$\partial_{\mathbf{Q}} T^B(\mathcal{E}, \mathbf{Q}) = \partial_{\mathcal{E}}[T^B(\mathcal{E}, \mathbf{Q})]\mathbf{u}^B(\mathcal{E}, \mathbf{Q}) - T^B(\mathcal{E}, \mathbf{Q})\partial_{\mathcal{E}}\mathbf{u}^B(\mathcal{E}, \mathbf{Q})$$
$$-k_B[T^B(\mathcal{E}, \mathbf{Q})]^2\partial_{\mathcal{E}\mathcal{E}}^2\mathbf{u}^B(\mathcal{E}, \mathbf{Q}). \tag{3.134}$$

The temperature T^B is an intensive quantity, while the energy of the universe, is an extensive quantity, so in the limit $N_B \to \infty$ with$\|\mathbf{u}^B\|/N_B \to 0$, all the terms on the right hand side of Eq. (3.134) vanish for every $\mathbf{Q}$. Therefore, T^B will be independent of $\mathbf{Q}$ in the thermodynamic limit:

$$T^B(\mathcal{E}, \mathbf{Q}) \underset{\text{TDlimit}}{=} T^B(\mathcal{E}). \tag{3.135}$$

Also note that,

$$\partial_{\mathbf{Q}} S^B(\mathcal{E}, \mathbf{Q}) = \frac{\mathbf{u}^B(\mathcal{E}, \mathbf{Q})}{T^B(\mathcal{E}, \mathbf{Q})} + k_B\partial_{\mathcal{E}}\mathbf{u}^B(\mathcal{E}, \mathbf{Q}), \tag{3.136}$$

the thermodynamic limit of which yields

$$\partial_{\mathbf{Q}} S^B(\mathcal{E}, \mathbf{Q}) \underset{\text{TDlimit}}{=} \frac{\mathbf{u}^B(\mathcal{E}, \mathbf{Q})}{T^B(\mathcal{E})}. \tag{3.137}$$

The average total adiabatic bath force is therefore equal to the gradient of $T^B(\mathcal{E})S^B(\mathcal{E}, \mathbf{Q})$ in the thermodynamic limit:

$$\mathbf{u}^B(\mathcal{E}, \mathbf{Q}) \underset{\text{TDlimit}}{=} -\partial_{\mathbf{Q}}[-T^B(\mathcal{E})S^B(\mathcal{E}, \mathbf{Q})], \tag{3.138}$$

implying that the quantity $-T^B(\mathcal{E})S^B(\mathcal{E}, \mathbf{Q})$ limits to the potential of the mean force $\mathbf{u}^B$. Note that Eq. (3.17) can be used to derive similar relations for each individual bath:

$$\partial_{\mathbf{Q}} T^{\alpha}(\mathcal{E}^{\alpha}, \mathbf{Q}) = \partial_{E^{\alpha}}[T^{\alpha}(E^{\alpha}, \mathbf{Q})]\mathbf{u}^{\alpha}(E^{\alpha}, \mathbf{Q}) - T^{\alpha}(E^{\alpha}, \mathbf{Q})\partial_{E^{\alpha}}\mathbf{u}^{\alpha}(E^{\alpha}, \mathbf{Q})$$

$$-k_B[T^{\alpha}(E^{\alpha}, \mathbf{Q})]^2\partial^2_{E^{\alpha}E^{\alpha}}\mathbf{u}^{\alpha}(E^{\alpha}, \mathbf{Q}), \tag{3.139}$$

$$\partial_{\mathbf{Q}} s^{\alpha}(\mathcal{E}, \mathbf{Q}) = \frac{\mathbf{u}^{\alpha}(E^{\alpha}, \mathbf{Q})}{T^{\alpha}(E^{\alpha}, \mathbf{Q})} + k_B\partial_{E^{\alpha}}\mathbf{u}^{\alpha}(E^{\alpha}, \mathbf{Q}). \tag{3.140}$$

If the entropy of bath α is extensive in the number of its degrees of freedom, then in a thermodynamic limit applied to bath α only (as opposed to all of the bath degrees of freedom), T^{α} becomes independent of $\mathbf{Q}$ and $-T^{\alpha}(E^{\alpha})s^{\alpha}(E^{\alpha}, \mathbf{Q})$ becomes the potential of the mean force $\mathbf{u}^{\alpha}$.

With the free energy and temperature position independent, P_s reduces to a genuine Boltzmann distribution with an effective system Hamiltonian H^{eff}

$$P_s(\mathbf{X}; \mathcal{E}, \boldsymbol{\lambda}) \underset{\text{TDlimit}}{=} e^{-\frac{H^{\text{eff}}(\mathbf{X}, \boldsymbol{\lambda}, \mathcal{E}) - \mathcal{F}(\mathcal{E}, \boldsymbol{\lambda})}{k_B T^B(\mathcal{E})}}, \tag{3.141}$$

where the effective Hamiltonian is given by the system Hamiltonian plus the potential of mean force:

$$H^{\text{eff}}(\mathbf{X}, \mathcal{E}, \boldsymbol{\lambda}) = H(\mathbf{X}, \boldsymbol{\lambda}) - T^B(\mathcal{E})S^B(\mathcal{E}, \mathbf{Q}), \tag{3.142}$$

and the free energy is given by

$$\mathcal{F}(\mathcal{E}, \boldsymbol{\lambda}) \underset{\text{TDlimit}}{=} -T^B(\mathcal{E}, \mathbf{Q})S^U(\mathcal{E}, \boldsymbol{\lambda}) \tag{3.143}$$

$$= -k_B T^B(\mathcal{E}, \mathbf{Q})\ln\int\limits_{0 \leq H(\mathbf{X}, \boldsymbol{\lambda}) \leq \mathcal{E}} d\mathbf{X}\,\Sigma^B(\mathcal{E} - H(\mathbf{X}, \boldsymbol{\lambda}), \mathbf{Q})$$

$$\underset{\text{TDlimit}}{=} -k_B T^B(\mathcal{E})\ln\int\limits_{0 \leq H(\mathbf{X}, \boldsymbol{\lambda}) \leq \mathcal{E}} d\mathbf{X}\,e^{-\frac{H^{\text{eff}}(\mathbf{X}, \boldsymbol{\lambda}, \mathcal{E})}{k_B T^B(\mathcal{E})}}.$$

Equation (3.143) is the standard statistical mechanics definition of the Helmholtz free energy [70]. Employing Eq. (3.119), we also have

$$
\begin{aligned}
\mathcal{F}(\mathcal{E}, \boldsymbol{\lambda}) \quad \underset{\text{TDlimit}}{=} \quad & -T^B(\mathcal{E}, \mathbf{Q}) S^P(\mathcal{E}, \boldsymbol{\lambda}) - T^B(\mathcal{E}, \mathbf{Q}) \left\langle S^B \right\rangle^P_{\mathcal{E}, \boldsymbol{\lambda}} \\
= \quad & -T^B(\mathcal{E}, \mathbf{Q}) S^P(\mathcal{E}, \boldsymbol{\lambda}) \\
& -T^B(\mathcal{E}, \mathbf{Q}) \int\limits_{0 \leq H(\mathbf{X}, \boldsymbol{\lambda}) \leq \mathcal{E}} d\mathbf{X}\, P_s(\mathbf{X}; \mathcal{E}, \boldsymbol{\lambda}) k_B \ln[\Sigma^B(\mathcal{E} - H(\mathbf{X}, \boldsymbol{\lambda}), \mathbf{Q})] \\
\underset{\text{TDlimit}}{=} \quad & -T^B(\mathcal{E}) S^P(\mathcal{E}, \boldsymbol{\lambda}) \\
& -T^B(\mathcal{E}) \int\limits_{0 \leq H(\mathbf{X}, \boldsymbol{\lambda}) \leq \mathcal{E}} d\mathbf{X}\, P_s(\mathbf{X}; \mathcal{E}, \boldsymbol{\lambda}) \left[-\frac{H^{\text{eff}}(\mathbf{X}, \mathcal{E}, \boldsymbol{\lambda})}{T^B(\mathcal{E})} \right] \\
= \quad & \left\langle H^{\text{eff}} \right\rangle^P_{\mathcal{E}, \boldsymbol{\lambda}} - T^B(\mathcal{E}) S^P(\mathcal{E}, \boldsymbol{\lambda}).
\end{aligned}
\tag{3.144}
$$

The last line of Eq. (3.144) is the standard thermodynamic definition of the Helmholtz free energy [70]. Finally, we note that in the thermodynamic limit, the microcanonical temperature of the universe, defined by

$$
\frac{1}{T^U(\mathcal{E}, \boldsymbol{\lambda})} = \partial_{\mathcal{E}} S^U(\mathcal{E}, \boldsymbol{\lambda}),
\tag{3.145}
$$

becomes independent of $\boldsymbol{\lambda}$ and is equal to the joint bath temperature T^B:

$$
\begin{aligned}
\frac{1}{T^U(\mathcal{E}, \boldsymbol{\lambda})} \quad = \quad & k_B \partial_{\mathcal{E}} \ln \Sigma^U(\mathcal{E}, \boldsymbol{\lambda}) \\
= \quad & \frac{k_B}{\Sigma^U(\mathcal{E}, \boldsymbol{\lambda})} \int\limits_{0 \leq H(\mathbf{X}, \boldsymbol{\lambda}) \leq \mathcal{E}} d\mathbf{X}\, \partial_{E^B} \Sigma^B(E^B, \mathbf{Q})|_{E^B = \mathcal{E} - H(\mathbf{X}, \boldsymbol{\lambda})} \\
= \quad & \frac{1}{\Sigma^U(\mathcal{E}, \boldsymbol{\lambda})} \int\limits_{0 \leq H(\mathbf{X}, \boldsymbol{\lambda}) \leq \mathcal{E}} d\mathbf{X}\, \Sigma^B(E^B, \mathbf{Q}) \partial_{E^B} S^B(E^B, \mathbf{Q})|_{E^B = \mathcal{E} - H(\mathbf{X}, \boldsymbol{\lambda})} \\
\underset{\text{TDlimit}}{=} \quad & \frac{1}{\Sigma^U(\mathcal{E}, \boldsymbol{\lambda})} \int\limits_{0 \leq H(\mathbf{X}, \boldsymbol{\lambda}) \leq \mathcal{E}} d\mathbf{X}\, \frac{\Sigma^B(\mathcal{E} - H(\mathbf{X}, \boldsymbol{\lambda}), \mathbf{Q}))}{T^B(\mathcal{E}, \mathbf{Q})} \left[1 + O\left(\frac{N_S}{N_B} \right) \right] \\
\underset{\text{TDlimit}}{=} \quad & \frac{1}{T^B(\mathcal{E})}.
\end{aligned}
\tag{3.146}
$$

3.5 Summary and Conclusions

In this chapter, we have shown that the chaos bath framework establishes a microscopic deterministic, Hamiltonian setting which yields thermodynamic evolution at the mesoscale. When a slow heavy mesoscopic Hamiltonian system interacts with any number of light fast micorscopic chaotic Hamiltonian systems, the chaotic systems' energies evolve on mesoscopic time-scales, and the joint distribution of the mesoscopic system and bath energies evolves diffusively under a Fokker-Planck equation. This Fokker-Planck equation can be derived from first principles from the underlying Hamiltonian dynamics. If the control parameters are held fixed, so that the system of interest and baths jointly evolve as an isolated system, the mesoscale variables thermalize with each other and relax to a stationary equilibrium distribution. We have studied this stationary distribution in detail, and we have shown that it implies a Boltzmann distribution in thermodynamic limit. Thus, the framework established here contains the classical framework of equilibrium statistical mechanics as a special case. The hallmarks of thermodynamic processes, however (namely fluctuations, dissipation, and thermalization), always emerge in the presence of a chaos bath regardless of the size of the bath or the existence of a thermodynamic limit.

Our derivation of the Boltzmann distribution demonstrates that ordinary equilibrium thermodynamics follows from the thermodynamic limit of the chaos bath framework, but the derivation itself is nothing special - deriving an equilibrium Boltzmann distribution from the assumption of molecular chaos and many degrees of freedom is a standard exercise in equilibrium statistical mechanics and is the essence of the results first established by Maxwell, Boltzmann, and Gibbs [65, 66, 67, 70]. The central result of this chapter is the Fokker-Planck equation (3.74) to which the stationary distribution is a solution. In addition to equilibrium thermodynamics, the Fokker-Planck equation connects non-equilibrium thermodynamics at the mesoscale to the underlying Hamiltonian chaos at the microscale. In the next chapter, we will use this micro-meso connection to interpret important results in non-equilibrium thermodynamics at both the microscale and mesoscale.

Chapter 4

Thermodynamics in a Chaos Bath: The Langevin Equation

4.1 Introduction

In this chapter, we continue our study of thermodynamics in a chaos bath and consider the mesoscopic dynamics of individual trajectories. The distribution of $(\mathbf{X}, \mathbf{E})$ evolves under a diffusion equation, so we expect the corresponding trajectories to evolve under a stochastic differential equation (denoted SDE for short) [9]. From the Fokker-Planck equation (3.75), we will derive an SDE for the mesoscopic evolution of $(\mathbf{X}, \mathbf{E})$ in the form of a generalized underdamped Langevin equation. The purpose of this chapter is to utilize the Langevin equation to derive the chaos bath anaolog of important results from stochastic thermodynamics, namely, fluctuation theorems [13, 15, 17, 18, 19, 20]. After the stochastic derivation, we will show that the fluctuation theorems follow trivially, essentially by definition, from generic properties of the underlying microscopic Hamiltonian dynamics, and this will be the central result of Chap. 4. Our approach will yield first principles Hamiltonian definitions heat and entropy production, as well as a statement of the first law of thermodynamics at the level of individual trajectories in a chaos bath. The second law of thermodynamics will follow from the fluctuation theorems in the thermodynamic limit.

The outline of this chapter is as follows. In Sec. 4.2 we derive the Langevin equation

and discuss the physical meaning of the various forces which emerge. Then, in Sec. 4.3, we review the basics of stochastic thermodynamics and find corresponding expressions for heat flow, entropy production, and the first law of thermodynamics in a chaos bath by utilizing the path integral formalism. In Sec. 4.4, we derive the chaos bath equivalent of fluctuation theorems from both mesoscopic and microscopic considerations, and we conclude in Sec. 4.5.

4.2 The Langevin equation

Fokker-Planck equations and stochastic differential equations are intimately related - for every white noise driven stochastic differential equation, there is a corresponding Fokker-Planck equation which describes the time evolution of distributions of the stochastic variables [9]. The prescription to find a Fokker-Plack equation corresponding to a given stochastic differential equation (denoted SDE for short) can be found in standard references like [9] and is outlined here. Suppose we are given an Ito stochastic differential equation for some vector function $\mathbf{Y}(t)$:

$$dY_i(t) = B_i(\mathbf{Y}(t), t)dt + \sqrt{\sigma}_{ij}(\mathbf{Y}(t), t)d\mathcal{W}_j(t'). \tag{4.1}$$

Summation over repeated subscripts is implied in Eq. (4.1), and $d\boldsymbol{\mathcal{W}}(t)$ denotes the increment of a vector of independent standard Wiener processes with the following statistics:

$$\langle d\mathcal{W}_i(t) \rangle = 0 \tag{4.2}$$

$$\langle d\mathcal{W}_i(t)d\mathcal{W}_j(t) \rangle = dt\, \delta_{ij}\, \delta(t - t'), \tag{4.3}$$

where δ_{ij} is the Kronecker delta, and the expectation $\langle ... \rangle$ is taken over all realizations of $\boldsymbol{\mathcal{W}}$. The time-derivative of $\boldsymbol{\mathcal{W}}(t)$, denoted by $\boldsymbol{\xi}(t) = \dot{\boldsymbol{\mathcal{W}}}(t)$, is defined formally to be a Gaussian white-noise vector. The function $\mathbf{B}(\mathbf{Y}, t)$ is called the "drift vector," and the matrix $\sqrt{\sigma}(\mathbf{Y}, t)$ is defined indirectly as a square root of the matrix σ:

$$\sigma(\mathbf{Y}, t) = \sqrt{\sigma}(\mathbf{Y}, t)\sqrt{\sigma}^T(\mathbf{Y}, t), \tag{4.4}$$

where σ is called the "diffusion matrix." Note that $\sigma(\mathbf{Y}, t)$ is symmetric and non-negative by definition. If $\mathbf{Y}$ and $\mathcal{W}$ are n and p-dimensional respectively, then $\mathbf{B}$ is n-dimensional, $\sqrt{\sigma}$ is an $n \times p$ matrix, and σ is an $n \times n$ matrix. Note that $\sqrt{\sigma}$ is not uniquely defined: if R is a $p \times m$ semi-orthogonal matrix for any m (meaning that $RR^T = I_p$, where I_p is the $p \times p$ identity matrix), we can define a new $n \times m$ matrix $\sqrt{\tilde{\sigma}} = \sqrt{\sigma}R$, and we find

$$\sigma = \sqrt{\sigma}\sqrt{\sigma}^T = \sqrt{\tilde{\sigma}}\sqrt{\tilde{\sigma}}^T. \tag{4.5}$$

This freedom will be a source of complications in the work to follow.

Given the SDE in Eq. (4.1) the probability distribution for $\mathbf{Y}$, denoted by $P(\mathbf{Y}, t)$, can be shown to evolve under the following Fokker-Planck equation [9]:

$$\partial_t P(\mathbf{Y}, t) = -\partial_{\mathbf{Y}} \cdot [\mathbf{B}(\mathbf{Y}, t)P(\mathbf{Y}, t)] + \frac{1}{2}\partial^2_{Y_i Y_j}[\sigma_{ij}(\mathbf{Y}, t)P(\mathbf{Y}, t)]. \tag{4.6}$$

We will apply this correspondence in reverse to derive an SDE for $(\mathbf{X}, \mathbf{E})$ from Fokker-Planck equation in Eq. (3.75). Due to the non-uniqueness of $\sqrt{\sigma}$, Eq. (3.75) corresponds to many SDE's with noise vectors of varying dimensionality, so the reverse correspondence will introduce some mathematical ambiguities which we will need to resolve by appealing to physical considerations.

4.2.1 Derivation

For simplicity, we will derive our Langevin equation assuming the presence of only one bath. The generalization to multiple baths is straightforward. We denote the number of degrees of freedom of the system of interest by N_S, where by "degree of freedom," we mean the number of position coordinates, implying the phase space of the system of interest is $2N_S$-dimensional. We begin our derivation with the Fokker-Planck equation in (3.75) for the case of $N = 1$ bath (so no summation over α is required). Expanding out the $\hat{\mathbf{D}}$ operators and rearranging, we find

$$\partial_t W_t = -\partial_{\mathbf{Q}} \cdot \left[M^{-1}\mathbf{P} W_t \right] - \partial_{\mathbf{P}} \cdot \left[\left(-\frac{\partial V}{\partial \mathbf{Q}} + \mathbf{u} - \frac{1}{2\Sigma}\partial_E \left[\Sigma\, L \right] M^{-1}\mathbf{P} \right) W_t \right] \qquad (4.7)$$

$$- \partial_E \left[\left(-\mathbf{P}^T M^{-1}\mathbf{u} - \frac{1}{2}\mathrm{tr}[M^{-1}L^{\mathrm{sy}}] + \mathbf{P}^T M^{-1}\frac{1}{2\Sigma}\partial_E \left[\Sigma\, L \right] M^{-1}\mathbf{P} \right) W_t \right]$$

$$+ \frac{1}{2}\partial^2_{P_i P_j}\left[L^{\mathrm{sy}}_{ij} W \right] + \frac{1}{2}\partial^2_{P_i E}\left[-L^{\mathrm{sy}}_{ij} M_j^{-1} P_j W_t \right] + \frac{1}{2}\partial^2_{E P_i}\left[-P_j M_j^{-1} L^{\mathrm{sy}}_{ji} W_t \right]$$

$$+ \frac{1}{2}\partial^2_{EE}\left[\mathbf{P}^T M^{-1} L^{\mathrm{sy}} M^{-1}\mathbf{P} W_t \right].$$

We again assume summation over repeated subscripts, and the function tr denotes the matrix trace. The term M_i^{-1} is the inverse inertia of the ith degree of freedom of the system of interest: $M_i^{-1} = M_{ii}^{-1}$, where M_{ij}^{-1} is an element of diagonal matrix M^{-1}. By denoting the joint vector $(\mathbf{X}, \mathbf{E})$ by $\mathbf{Y}$, Eq. (4.7) can be cast in the form of Eq. (4.6). The $2N_S + 1$ dimensional drift vector $\mathbf{B}(\mathbf{Y}, t)$ is given by the function

$$\mathbf{B}(\mathbf{Y}, t) = \begin{pmatrix} M^{-1}\mathbf{P} \\ -\partial_{\mathbf{Q}}V + \mathbf{u} - \frac{1}{2\Sigma}\partial_E\left[\Sigma\, L \right] M^{-1}\mathbf{P} \\ -\mathbf{P}^T M^{-1}\mathbf{u} - \frac{1}{2}\mathrm{tr}[M^{-1}L^{\mathrm{sy}}] + \mathbf{P}^T M^{-1}\frac{1}{2\Sigma}\partial_E\left[\Sigma\, L \right] M^{-1}\mathbf{P} \end{pmatrix}. \qquad (4.8)$$

Note that the explicit time-dependence of $\mathbf{B}$ arises from the time-dependence of the control parameters $\boldsymbol{\lambda}(t)$ appearing in the system of interest's potential energy function. The $(2N_S + 1) \times (2N_S + 1)$ time-independent diffusion matrix is given by

$$\sigma(\mathbf{Y}) = \begin{pmatrix} 0_{N_S} & 0_{N_S} & \mathbf{0}_{N_S} \\ 0_{N_S} & L^{\mathrm{sy}} & -L^{\mathrm{sy}}M^{-1}\mathbf{P} \\ \mathbf{0}_{N_S}^T & -\mathbf{P}^T M^{-1}(L^{\mathrm{sy}})^T & \mathbf{P}^T M^{-1}L^{\mathrm{sy}}M^{-1}\mathbf{P} \end{pmatrix}, \qquad (4.9)$$

where $\mathbf{0}_{N_S}$ denotes the N_S dimensional zero vector and 0_{N_S} denotes the $N_S \times N_S$ zero matrix. The elements of $\sigma(\mathbf{Y})$ are themselves matrices with dimensions given by the following schematic:

$$
\begin{pmatrix}
N_S \times N_S & N_S \times N_S & N_S \times 1 \\
N_S \times N_S & N_S \times N_S & N_S \times 1 \\
1 \times N_S & 1 \times N_S & 1 \times 1
\end{pmatrix}.
\tag{4.10}
$$

With these definitions, Eq. (4.7) can be written in the form of Eq. (4.6):

$$
\partial_t W_t = -\partial_{\mathbf{Y}} \cdot [\mathbf{B}(\mathbf{Y}, t) W_t] + \frac{1}{2} \partial^2_{Y_i Y_j} [\sigma_{ij}(\mathbf{Y}) W_t].
\tag{4.11}
$$

In order to write an SDE corresponding to Eq. (4.11), the matrix σ defined in Eq. (4.9) must be non-negative, and we must find an appropriate square root $\sqrt{\sigma}$. If we can find a square root, we will have $\sigma = \sqrt{\sigma}\sqrt{\sigma}^T$, so σ will be non-negative by definition. The square root, if it exists, is not unique, but from the correspondence between Eqs. (4.1) and Eq. (4.6), we know that if $\sqrt{\sigma}$ does exist, it will be a $(2N_S + 1) \times m$ dimensional matrix for some integer m, where m is the number of independent noise terms appearing in the SDE for $\mathbf{Y}$. The integer m is not known a priori, but we know that the physical origin of the noise terms is the chaotic fluctuating force of the bath on the system of interest, and this force has N_S components. We therefore expect N_S independent noises to enter the SDE for $\mathbf{Y}$, and consequently expect $\sqrt{\sigma}$, if it exists, be a $(2N_S + 1) \times N_S$ dimensional matrix. We now show, by construction, that $\sqrt{\sigma}$ does indeed exist.

The matrix L^{sy} is by-definition real and symmetric, and in Sec. 3.4, we showed that L^{sy} is non-negative when the bath is strongly chaotic. Therefore, for a strongly chaotic bath, L^{sy} can be diagonalized by a real orthogonal transformation:

$$
L^{\mathrm{sy}} = O \operatorname{diag}(\Lambda_i) O^T,
\tag{4.12}
$$

The matrix $\operatorname{diag}(\Lambda_i)$ is an $N_S \times N_S$ diagonal matrix of the real non-negative eigenvalues Λ_i of L^{sy}, and O is a real orthogonal $N_S \times N_S$ matrix. We define the matrix $\sqrt{L^{\mathrm{sy}}}$ by

$$
\sqrt{L^{\mathrm{sy}}} = O \operatorname{diag}(\sqrt{\Lambda_i}) O^T.
\tag{4.13}
$$

Because $OO^T = I_{N_S}$, where I_{N_S} is the $N_S \times N_S$ identity matrix, we see immediately that

$$L^{\text{sy}} = \sqrt{L^{\text{sy}}}\sqrt{L^{\text{sy}}}^T. \tag{4.14}$$

We also see that $\sqrt{L^{\text{sy}}}$ is itself symmetric. Utilizing $\sqrt{L^{\text{sy}}}$, we find the following expression for $\sqrt{\sigma}$:

$$\sqrt{\sigma}(\mathbf{Y}) = \begin{pmatrix} 0 \\ \sqrt{L^{\text{sy}}}(E, \mathbf{Q}) \\ -\mathbf{P}^T M^{-1}\sqrt{L^{\text{sy}}}(E, \mathbf{Q}) \end{pmatrix}, \tag{4.15}$$

where, the dimensions of the sub-matrices of $\sqrt{\sigma}$ are given by,

$$\begin{pmatrix} N_S \times N_S \\ N_S \times N_S \\ 1 \times N_S \end{pmatrix}. \tag{4.16}$$

One can immediately verify that $\sigma = \sqrt{\sigma}\sqrt{\sigma}^T$. Note that we still have the freedom to multiply $\sqrt{\sigma}$ by some $N_S \times N_S$ orthogonal matrix R. If we write $(\sqrt{\sigma})' = \sqrt{\sigma}R$, we have

$$(\sqrt{\sigma})'(\mathbf{Y}) = \begin{pmatrix} 0 \\ \sqrt{L^{\text{sy}}}(E, \mathbf{Q})R \\ -\mathbf{P}^T M^{-1}\sqrt{L^{\text{sy}}}(E, \mathbf{Q})R \end{pmatrix}, \tag{4.17}$$

and we see that $\sigma = (\sqrt{\sigma})'(\sqrt{\sigma})'^T$. We will show shortly that this freedom is of no physical consequence.

We now have all the necessary pieces to write a stochastic differential equation corresponding to the Fokker-Planck equation in (3.75). Following the correspondence between Eq. (4.1) and Eq. (4.6), we arrive at the desired SDE:

$$dY_i(t) = B_i(\mathbf{Y}(t), t)dt + \sqrt{\sigma}_{ij}(\mathbf{Y}(t))d\mathcal{W}_j, \tag{4.18}$$

In more transparent notation, we have

$$d\mathbf{Q} = M^{-1}\mathbf{P}\,dt \tag{4.19}$$

$$d\mathbf{P} = \left(-\partial_{\mathbf{Q}}V + \mathbf{u} - \frac{1}{2\Sigma}\partial_E\left[\Sigma\,L\right]M^{-1}\mathbf{P}\right)dt + \sqrt{L^{\mathrm{sy}}}\,d\boldsymbol{\mathcal{W}}$$

$$dE = \left(-\mathbf{P}^T M^{-1}\mathbf{u} - \frac{1}{2}\mathrm{tr}[M^{-1}L^{\mathrm{sy}}] + \mathbf{P}^T M^{-1}\frac{1}{2\Sigma}\partial_E\left[\Sigma\,L^{\mathrm{sy}}\right]M^{-1}\mathbf{P}\right)dt$$
$$\quad -\mathbf{P}^T M^{-1}\sqrt{L^{\mathrm{sy}}}\,d\boldsymbol{\mathcal{W}}.$$

For simplicity, we have suppressed the functional dependencies of all of the quantities in Eq. (4.19). The $\mathbf{Q}$ and $\mathbf{P}$ equations together have the form of an underdamped Langevin equation, so we will refer to Eq. (4.19) as the Langevin equation for $\big(\mathbf{X}(t), E(t)\big)$. The multi-bath generalization is given by

$$d\mathbf{Q} = M^{-1}\mathbf{P}\,dt \tag{4.20}$$

$$d\mathbf{P} = -\partial_{\mathbf{Q}}V\,dt + \sum_{\alpha}\left(\mathbf{u}^{\alpha} - \frac{1}{2\Sigma^{\alpha}}\partial_{E^{\alpha}}\left[\Sigma^{\alpha}\,L^{\alpha}\right]M^{-1}\mathbf{P}\right)dt + \sqrt{L^{\mathrm{sy}\,\alpha}}\,d\boldsymbol{\mathcal{W}}^{\alpha}$$

$$dE^{\alpha} = \left(-\mathbf{P}^T M^{-1}\mathbf{u}^{\alpha} - \frac{1}{2}\mathrm{tr}[M^{-1}L^{\mathrm{sy}\,\alpha}] + \mathbf{P}^T M^{-1}\frac{1}{2\Sigma^{\alpha}}\partial_{E^{\alpha}}\left[\Sigma\,L^{\mathrm{sy}\,\alpha}\right]M^{-1}\mathbf{P}\right)dt$$
$$\quad -\mathbf{P}^T M^{-1}\sqrt{L^{\mathrm{sy}\,\alpha}}\,d\boldsymbol{\mathcal{W}}^{\alpha},$$

where $\boldsymbol{\mathcal{W}}^{\alpha}$ and $\boldsymbol{\mathcal{W}}^{\beta}$ are independent for $\alpha \neq \beta$.

From Eq. (4.19), we can see why choosing the $\sqrt{\sigma}$ as in Eq. (4.15) is equivalent to choosing $(\sqrt{\sigma})' = \sqrt{\sigma}R$ as in Eq. (4.17). If we had chosen to use $(\sqrt{\sigma})'$, then we would have obtained the same stochastic differential equation, but with the noise vector $\boldsymbol{\mathcal{W}}' = R\boldsymbol{\mathcal{W}}$ instead of $\boldsymbol{\mathcal{W}}$. Utilizing the orthogonality of R, it is straightforward to show that $\boldsymbol{\mathcal{W}}$ and $\boldsymbol{\mathcal{W}}'$ have identical statistics; $\boldsymbol{\mathcal{W}}' = R\boldsymbol{\mathcal{W}}$ is also a vector of independent standard Weiner processes. Thus, the SDE's obtained by using $\sqrt{\sigma}$ and $(\sqrt{\sigma})'$ are physically identical.

4.2.2 Energy conservation and the Stratonovich calculus

By the conservation of energy, the quantity $-dE^{\alpha}$ is the incremental change in the energy of the system of due to the force exerted by bath α. On the other hand, the expression

for the force on the system of interest due to bath α, denoted here by $\widetilde{\mathbf{F}}^\alpha$, is, according to Eq. (4.20),

$$\widetilde{\mathbf{F}}^\alpha = \mathbf{u}^\alpha - \frac{1}{2\Sigma^\alpha}\partial_{E^\alpha}\left[\Sigma^\alpha L^\alpha\right]M^{-1}\mathbf{P} + \sqrt{L^{\text{sy}\,\alpha}}\boldsymbol{\xi}^\alpha. \tag{4.21}$$

so applying the definition incremental energy change = force $\times$ displacement, where the displacement is given by $\mathrm{d}\mathbf{Q} = M^{-1}\mathbf{P}\mathrm{d}t$, we have

$$
\begin{aligned}
\mathrm{d}\mathbf{Q}^T\widetilde{\mathbf{F}}^\alpha &= \left(\mathbf{P}^T M^{-1}\mathbf{u}^\alpha - \mathbf{P}^T M^{-1}\frac{1}{2\Sigma^\alpha}\partial_{E^\alpha}\left[\Sigma^\alpha L^\alpha\right]M^{-1}\mathbf{P}\right)\mathrm{d}t \\
&\quad + \mathbf{P}^T M^{-1}\sqrt{L^{\text{sy}\alpha}}\mathrm{d}\boldsymbol{\mathcal{W}}^\alpha.
\end{aligned}
\tag{4.22}
$$

We thus see that $\mathrm{d}\mathbf{Q}^T\widetilde{\mathbf{F}}^\alpha \neq -\mathrm{d}E^\alpha$, so the conservation of energy appears to be in contradiction with the definition of incremental energy change. This apparent contradiction arises because we have written our SDE in the Ito form, where the regular rules of calculus do not always apply. In particular, the differential of the system of interest's Hamiltonian along trajectories must be calculated with a modified chain rule known as Ito's formula [9]. Ito's formula introduces terms in addition to those given by the normal chain rule, so the differential change in the system of interest's energy along a trajectory is not given by force $\times$ displacement in the Ito calculus.

In order to use the standard definitions of work, changes in energy, and other physical quantities, we must write our SDE in the Stratonovich form, where the regular rules of calculus apply [85]. The essential difference between the Ito and Stratonovich calculus lies in the discretization conventions for defining stochastic integrals. In short, the Ito integral of some function $G(\mathbf{Y},t)$ along a stochastic trajectory $\mathbf{Y}(t)$ is defined as

$$\int_0^T G(\mathbf{Y}(t),t)\mathrm{d}\mathbf{Y}(t) \underset{\text{Lim}}{=} \sum G(\mathbf{Y}(t_i),t_i)\left[\mathbf{Y}(t_{i+1}) - \mathbf{Y}(t_i)\right], \tag{4.23}$$

while the corresponding Strantonivich integral is defined as

$$\int_0^T G\big(\mathbf{Y}(t), t\big) \circ \mathrm{d}\mathbf{Y}(t) \quad \underset{\text{Lim}}{=} \sum G\big(\tfrac{\mathbf{Y}(t_i) + \mathbf{Y}(t_{i+1})}{2}, \tfrac{t_i + t_{i+1}}{2}\big) \cdot \big[\mathbf{Y}(t_{i+1}) - \mathbf{Y}(t_i)\big], \quad (4.24)$$

where the operation "Lim" corresponds to the limit of an infintely fine discretization of the interval $[0, T]$. The Ito integral uses an initial point discretization scheme, while the Stratonovich integral uses a mid-point discretization scheme. If $\mathbf{Y}(t)$ is a differentiable function, both discretization schemes will be equivalent.

The Stratonovich calculus is preferred from a physical standpoint because stochastic noises, which often induce continuous but non-differential fractal sample paths, always represent idealized limits or approximations of more regular processes when used to model physical phenomena. As shown by Sussmann [86], if a sequence of differentiable functions limits to a continuous, possibly non-differentiable function (such as the sample path of a Gaussian white noise process), then, under suitable Lipschitz continuity conditions, corresponding sequences of ordinary differential equations driven by those differentiable functions will limit to corresponding differential equations driven by the limiting continuous function in the Stratonovich form, not the Ito form. Key to proving this result is the fact that the set of differentiable functions defined over some compact subset of the real numbers is dense in the set of continuous functions defined over that compact subset [86]. Therefore, equations of motion for physical quantities evaluated along trajectories (such as work and energy), when defined in the context of ordinary deterministic Hamiltonian dynamics, will retain their form when the deterministic dynamics are approximated by Stratonovich stochastic dynamics. It is for these reasons that the Stratonovich calculus is used in stochastic thermodynamics, where physically meaningful results are dependent on physically meaningful definitions of work and heat along individual trajectories of mesoscale thermodynamic systems [16, 13]. In addition to the above mentioned benefits, the symmetric nature of the Stratonovich discretization is convenient for deriving fluctuation theorems, as we will show in Sec. 4.4.

The general Ito SDE given in Eq. (4.1) can be written as the following Stratonovich SDE:

$$dY_i(t) = B_i^s(\mathbf{Y}(t), t)dt + \sigma_{ij}(\mathbf{Y}(t), t) \circ d\mathcal{W}_j(t), \tag{4.25}$$

where the Stratonovich drift vector $\mathbf{B}^s$ is related to the Ito drift vector $\mathbf{B}$ from Eq. (4.1) by the formula

$$B_i^s(\mathbf{Y}, t) = B_i(\mathbf{Y}, t) - \frac{1}{2}\frac{\partial\sqrt{\sigma}_{ik}(\mathbf{Y}, t)}{\partial Y_j}\sqrt{\sigma}_{jk}(\mathbf{Y}, t). \tag{4.26}$$

Applying this rule to (4.9), for the case of a single bath, we find

$$\mathbf{B}^s(\mathbf{Y}, t) = \mathbf{B}(\mathbf{Y}, t) + \begin{pmatrix} \mathbf{0} \\ \frac{1}{2}\partial_E\left[\sqrt{L^{\mathrm{sy}}}\right]\sqrt{L^{\mathrm{sy}}}M^{-1}\mathbf{P} \\ \frac{1}{2}\mathrm{tr}[M^{-1}L] - \frac{1}{2}\mathbf{P}^T M^{-1}\partial_E\left[\sqrt{L^{\mathrm{sy}}}\right]\sqrt{L^{\mathrm{sy}}}M^{-1}\mathbf{P} \end{pmatrix}, \tag{4.27}$$

which, with Eqs. (4.8) and (4.20), gives the Langevin equation in the Stratonovich form:

$$
\begin{aligned}
d\mathbf{Q} &= M^{-1}\mathbf{P}dt \\
d\mathbf{P} &= -\partial_\mathbf{Q} V dt + \left(\mathbf{u} - \frac{1}{2\Sigma}\partial_E\left[\Sigma\, L\right]M^{-1}\mathbf{P} + \frac{1}{2}\partial_E\left[\sqrt{L^{\mathrm{sy}}}\right]\sqrt{L^{\mathrm{sy}}}M^{-1}\mathbf{P}\right)dt \\
&\quad + \sqrt{L^{\mathrm{sy}}} \circ d\mathcal{W} \\
dE &= \left(-\mathbf{P}^T M^{-1}\mathbf{u} + \mathbf{P}^T M^{-1}\frac{1}{2\Sigma}\partial_E\left[\Sigma\, L^{\mathrm{sy}}\right]M^{-1}\mathbf{P}\right. \\
&\quad \left. -\mathbf{P}^T M^{-1}\frac{1}{2}\partial_E\left[\sqrt{L^{\mathrm{sy}}}\right]\sqrt{L^{\mathrm{sy}}}M^{-1}\mathbf{P}\right)dt - \mathbf{P}^T M^{-1}\sqrt{L^{\mathrm{sy}}} \circ d\mathcal{W}.
\end{aligned}
\tag{4.28}
$$

Generalizing to multiple baths, we have

$$\mathrm{d}\mathbf{Q} = M^{-1}\mathbf{P}\mathrm{d}t \tag{4.29}$$

$$\mathrm{d}\mathbf{P} = -\partial_{\mathbf{Q}}V\mathrm{d}t + \sum_{\alpha}\left(\mathbf{u}^{\alpha} - \frac{1}{2\Sigma}\partial_{E^{\alpha}}\left[\Sigma\,L^{\alpha}\right]M^{-1}\mathbf{P}\right.$$

$$\left. +\frac{1}{2}\partial_{E^{\alpha}}\left[\sqrt{L^{\mathrm{sy}\,\alpha}}\right]\sqrt{L^{\mathrm{sy}\,\alpha}}M^{-1}\mathbf{P}\right)\mathrm{d}t + \sqrt{L^{\mathrm{sy}\,\alpha}}\circ\mathrm{d}\boldsymbol{\mathcal{W}}^{\alpha}$$

$$\mathrm{d}E^{\alpha} = \left(-\mathbf{P}^{T}M^{-1}\mathbf{u}^{\alpha} + \mathbf{P}^{T}M^{-1}\frac{1}{2\Sigma}\partial_{E^{\alpha}}\left[\Sigma\,L^{\mathrm{sy}\,\alpha}\right]M^{-1}\mathbf{P}\right.$$

$$\left. -\mathbf{P}^{T}M^{-1}\frac{1}{2}\partial_{E^{\alpha}}\left[\sqrt{L^{\mathrm{sy}\,\alpha}}\right]\sqrt{L^{\mathrm{sy}\,\alpha}}M^{-1}\mathbf{P}\right)\mathrm{d}t - \mathbf{P}^{T}M^{-1}\sqrt{L^{\mathrm{sy}\,\alpha}}\circ\mathrm{d}\boldsymbol{\mathcal{W}}^{\alpha}.$$

The force of bath α on the system of interest, denoted by $\mathbf{F}^{\alpha}$, is thus given by

$$\mathbf{F}^{\alpha} = \mathbf{u}^{\alpha} - \frac{1}{2\Sigma}\partial_{E^{\alpha}}\left[\Sigma\,L^{\alpha}\right]M^{-1}\mathbf{P} - \frac{1}{2}\partial_{E^{\alpha}}\left[\sqrt{L^{\mathrm{sy}\,\alpha}}\right]\sqrt{L^{\alpha\,\mathrm{sy}}}M^{-1}\mathbf{P} + \sqrt{L^{\mathrm{sy}\,\alpha}}\circ\boldsymbol{\xi}^{\alpha}, \tag{4.30}$$

and we find

$$\mathrm{d}\mathbf{Q}^{T}\mathbf{F}^{\alpha} = \left(\mathbf{P}^{T}M^{-1}\mathbf{u}^{\alpha} - \mathbf{P}^{T}M^{-1}\frac{1}{2\Sigma}\partial_{E^{\alpha}}\left[\Sigma\,L^{\mathrm{sy}\,\alpha}\right]M^{-1}\mathbf{P}\right.$$

$$\left. +\mathbf{P}^{T}M^{-1}\frac{1}{2}\partial_{E^{\alpha}}\left[\sqrt{L^{\mathrm{sy}\,\alpha}}\right]\sqrt{L^{\mathrm{sy}\,\alpha}}M^{-1}\mathbf{P}\right)\mathrm{d}t + -\mathbf{P}^{T}M^{-1}\sqrt{L^{\mathrm{sy}\,\alpha}}\circ\mathrm{d}\boldsymbol{\mathcal{W}}^{\alpha}$$

$$= -\mathrm{d}E^{\alpha}, \tag{4.31}$$

so we see that the normal definition of incremental energy change is consistent with the conservation of energy.

4.2.3 Adiabatic reaction forces

Equation (4.30) gives a mesoscopic expression for the forces exerted by a bath on the system of interest, defined using the physically meaningful form of stochastic calculus, and we now discuss the terms which appear in the context of the adiabatic reaction forces

associated with the ergodic adiabatic Hamiltonian framework [48, 68, 75, 77, 78]. These forces are back reactions (in the sense of Newton's third law) to the forces which the system of interest exerts on the baths due to its slow time-evolution. The terms in Eq. (4.30) can be broken into five physically distinct parts which we will discuss separately:

$$\mathbf{F}^\alpha \;=\; \mathbf{F}_A^\alpha + \mathbf{F}_{GM}^\alpha + \mathbf{F}_{DF}^\alpha + \mathbf{F}_F^\alpha + \mathbf{F}_{ND}^\alpha. \tag{4.32}$$

The first force in Eq. (4.32) is the adiabatic force:

$$\mathbf{F}_A^\alpha = \mathbf{u}^\alpha. \tag{4.33}$$

As mentioned in Sec. 3.3.3, $\mathbf{F}_A^\alpha$ is the force which governs the conservation of the baths' ergodic adiabatic invariants when the separation of the system and bath time-scales is infinitely large. We can see the explicit relationship between $\mathbf{u}^\alpha$ and the conservation of the ergodic adiabatic invariant Ω^α, defined in Eq. (3.10), by noting that

$$\begin{aligned} \mathrm{d}\Omega^\alpha &= \partial_{\mathbf{Q}}\Omega^\alpha \cdot \mathrm{d}\mathbf{Q} + \partial_{E^\alpha}\Omega^\alpha \mathrm{d}E^\alpha \tag{4.34}\\ &= \Sigma^\alpha \mathbf{u}^\alpha \cdot \mathrm{d}\mathbf{Q} + \Sigma^\alpha \mathrm{d}E^\alpha, \tag{4.35} \end{aligned}$$

where we have made use of Eqs. (3.11) and (3.15). In the adiabatic limit, $\mathrm{d}\Omega^\alpha = 0$, so we have

$$\mathrm{d}E^\alpha = -\mathbf{u}^\alpha \cdot \mathrm{d}\mathbf{Q}. \tag{4.36}$$

The above expression gives the increment of work done on bath α by the system of interest in the adiabatic limit, so the corresponding adiabatic force on bath α is $-\mathbf{u}^\alpha$, and the reaction force exerted on the system of interest is thus $\mathbf{u}^\alpha$. Because system of interest changes infinitely slowly in the adiabatic limit, the evolution governed by $\mathbf{u}^\alpha$ is both quasi-static

109

and reversible (in the thermodynamic sense), and the adiabatic force results in a reversible exchange of energy between the system and bath.

The adiabatic force is often called the Born-Oppenheimer force in reference to the Born-Oppenheimer approximation used in quantum chemistry. In the Born-Oppenheimer approximation, one utilizes the large separation of time scales between the motion of the nuclei and electrons of a molecule to write approximate molecular wave function [87]. One first calculates the electronic wave function with the nuclei held fixed, then calculates the nuclear wave function subject to average potential created by the electronic wave function. The effect of the average potential on the nuclei is analogous to the effect of the average force $\mathbf{u}^\alpha$ on the system of interest. This analogy becomes stronger in the thermodynamic limit applied to bath α, where as shown in Sec. 3.4.5, $\mathbf{u}^\alpha$ is the gradient of a potential.

The next two forces in Eq. (4.30), $\mathbf{F}^\alpha_{GM}$ and $\mathbf{F}^\alpha_{DF}$, are called geometric magnetism and deterministic friction, respectively. They arise from the symmetric and anti-symmetric parts of L^α in the term $-\frac{1}{2\Sigma^\alpha}\partial_{E^\alpha}\left[\Sigma^\alpha L^\alpha\right]M^{-1}\mathbf{P}$. Geometric magnetism is given by the expression

$$
\begin{aligned}
\mathbf{F}^\alpha_{GM} &= -\frac{1}{2\Sigma^\alpha}\partial_{E^\alpha}\left[\Sigma^\alpha L^{\mathrm{ay}\,\alpha}\right]M^{-1}\mathbf{P} \\
&= -K^{\mathrm{ay}\,\alpha}M^{-1}\mathbf{P} \\
&= -K^{\mathrm{ay}\,\alpha}\dot{\mathbf{Q}},
\end{aligned}
\tag{4.37}
$$

where K^{ay} is an anti-symmetric 2-form, given by

$$
K^{\mathrm{ay}\,\alpha} = \frac{1}{2\Sigma^\alpha}\partial_{E^\alpha}\left[\Sigma^\alpha L^{\mathrm{ay}\,\alpha}\right]
\tag{4.38}
$$

This force was first described by Berry and Robbins in [77, 78]. Note that if the system of interest has N_S degrees of freedom, then there exists a natural isomorphism between 2-forms and $(N_S - 2)$-forms, so if $N_S = 3$, there is a natural vector field $\mathbf{b}^\alpha$ associated with the 2-form $K^{\mathrm{ay}\,\alpha}$, and the tensor contraction in Eq. (4.37) can be written as a vector cross product, just like the ordinary magnetic force [77, 80]:

$$\mathbf{F}^\alpha_{GM} = (M^{-1}\mathbf{P}) \times \mathbf{b}^\alpha \tag{4.39}$$

$$= \dot{\mathbf{Q}} \times \mathbf{b}^\alpha.$$

Regardless of dimensionality, the form of $L^{\mathrm{ay}\,\alpha}$ given in Eq. (3.93) can be used to write $K^{\mathrm{ay}\,\alpha}$ in the language of exterior calculus:

$$K^{\mathrm{ay}\,\alpha} = \frac{1}{2\Sigma^\alpha} \partial_{E^\alpha} \left[\Sigma^\alpha \int\limits_{-\infty}^{0} \mathrm{d}t \, \left\langle \widetilde{\mathrm{d}} h^\alpha \wedge \widetilde{\mathrm{d}} h^\alpha_t \right\rangle_{E^\alpha,\mathbf{Q}} \right], \tag{4.40}$$

where $\widetilde{\mathrm{d}}$ denotes the exterior derivative with respect to $\mathbf{Q}$ and the symbol $\wedge$ denotes the wedge product. In components, we have

$$K^{\mathrm{ay}\,\alpha}_{ij} = \frac{1}{2\Sigma^\alpha} \partial_{E^\alpha} \left[\Sigma^\alpha \int\limits_{-\infty}^{0} \mathrm{d}t \, \left\langle \partial_{Q_i} h^\alpha \partial_{Q_j} h^\alpha_t - \partial_{Q_j} h^\alpha \partial_{Q_i} h^\alpha_t \right\rangle_{E^\alpha,\mathbf{Q}} \right]. \tag{4.41}$$

The terms inside the expectation of Eq. (4.41) are reminiscent of the definition of the components of a magnetic field in terms of a vector potential in $\mathbb{R}^2$ or $\mathbb{R}^3$. In Ref. [76], it is shown formally that the exterior derivative of $K^{\mathrm{ay}\,\alpha}$ is identically zero, so $K^{\mathrm{ay}\,\alpha}$ is divergenceless (in a generalized sense) like an ordinary magnetic field. Furthermore, because $K^{\mathrm{ay}\,\alpha}$ is anti-symmetric, the force $\mathbf{F}^\alpha_{GM}$ does no work, just like the ordinary magnetic force:

$$\mathbf{P}^T M^{-1} \mathbf{F}^\alpha_{GM} = -\mathbf{P}^T M^{-1} K^{\mathrm{ay}\,\alpha} M^{-1} \mathbf{P} \tag{4.42}$$

$$= 0.$$

We thus conclude that $\mathbf{F}^\alpha_{GM}$ behaves analogously to an arbitrary dimensional generalization of the magnetic Lorentz force. We note that in Ref. [76], it is shown that if bath α's Hamiltonian is symmetric with respect any anti-canonical symmetry (such as time-reversal), then $K^{\mathrm{ay}\,\alpha}$ will vanish identically. This leads to the interesting observation that anti-

canonical symmetry breaking by the baths can be promoted to the system of interest, even if the system of interest breaks no such symmetries. For example, if the an electrically neutral system of interest is coupled to electrically charged baths, and the if the baths are subject to an external magnetic field, the system of interest can experience a magnetic-like force, even though it produces no electric current. The force $\mathbf{F}^\alpha_{GM}$ is called "geometric" magnetism because the expression for $K^{\text{ay}\,\alpha}$ in Eq. (4.40) is, remarkably, the semi-classical limit of the 2-form which generates the geometric phase (also known as Berry's phase) in chaotic quantum systems with an adiabatically evolving parameters [76, 77]. By "chaotic quantum system," we mean a quantum system whose Hamiltonian generates chaotic motion in the corresponding classical system.

Deterministic friction, first described by Wilkinson [48], causes an irreversible dissipation of energy from the system of interest to the baths, and is given by

$$
\begin{aligned}
\mathbf{F}^\alpha_{DF} &= -\frac{1}{2\Sigma^\alpha}\partial_{E^\alpha}\left[\Sigma\, L^{\text{sy}\,\alpha}\right]M^{-1}\mathbf{P} \\
&= -K^{\text{sy}\,\alpha}M^{-1}\mathbf{P} \\
&= -K^{\text{sy}\,\alpha}\dot{\mathbf{Q}},
\end{aligned}
\tag{4.43}
$$

where $K^{\text{sy}\,\alpha}$ is the symmetric friction tensor, given by

$$
K^{\text{sy}\,\alpha} = \frac{1}{2\Sigma^\alpha}\partial_{E^\alpha}\left[\Sigma^\alpha\, L^{\text{sy}\,\alpha}\right]
\tag{4.44}
$$

This friction is called deterministic because its presence in Eq. (4.51) is a consequence of the deterministic chaos associated with the finite dimensional bath, in contrast to the standard frictional force which arises from an infinite stochastic thermal bath. The rate of dissipation due to this friction is given by $\mathbf{P}^T M^{-1}\mathbf{F}_{DF}$:

$$
\mathbf{P}^T M^{-1}\mathbf{F}^\alpha_{DF} = -\mathbf{P}^T M^{-1}K^{\text{sy}\,\alpha}M^{-1}\mathbf{P}.
\tag{4.45}
$$

As pointed out by Berry and Robbins [77], Eq. (4.45) can honestly be called energy dissipation only if the friction tensor K^{sy} is non-negative. The matrix, $L^{\text{sy}\,\alpha}$ is non-negative (given sufficiently strong chaos), so the quantity $\Sigma^\alpha L^{\text{sy}\,\alpha}$ must be an increasing function of E^α in

order for $K^{\text{sy}\,\alpha}$ to be non-negative. If we expand the derivative in Eq. (4.44) and make use of the identities in Eqs. (3.105) and (3.125), we find

$$K^{\text{sy}\,\alpha} = \frac{1}{2k_B T^\alpha}\left[L^{\text{sy}\,\alpha} + k_B T^\alpha \partial_{E^\alpha} L^{\text{sy}\,\alpha}\right]. \qquad (4.46)$$

The second quantity in the brackets in the above equation represents the first order change in $L^{\text{sy}\,\alpha}$ which occurs when a unit of thermal energy $k_B T^\alpha$ is added to the bath, so this quantity should be negligible in comparison to $L^{\text{sy}\,\alpha}$ in the thermodynamic limit applied to bath α:

$$K^{\text{sy}\,\alpha} \underset{\text{TDlimit}}{=} \frac{L^{\text{sy}\,\alpha}}{2k_B T^\alpha}. \qquad (4.47)$$

Thus, in the thermodynamic limit applied to bath α, $K^{\text{sy}\,\alpha}$ is non-negative.

The fourth force in Eq. (4.30), $\mathbf{F}_F^\alpha$, is the rapidly fluctuating force of bath α on the system of interest:

$$\mathbf{F}_F^\alpha = \sqrt{L^{\text{sy}\,\alpha}} \circ \boldsymbol{\xi}^\alpha. \qquad (4.48)$$

We see that $L^{\text{sy}\,\alpha}$ determines the strength of the fluctuations, and therefore must also determine the rate of diffusion in the system of interest's momentum space due to bath α. The rapidly fluctuating force was first discussed by Jarzynski for the case of a single bath [68]. Although no explicit expression was written for $\mathbf{F}_F^\alpha$ in Ref. [68], the concept of fluctuations facilitated the derivation of the single bath Fokker-Planck equation and its stationary distribution. Comparing Eqs. (4.44) and (4.48), we see that the friction tensor is defined in terms of $L^{\text{sy}\,\alpha}$, so Eq. (4.44) is a fluctuation-dissipation relation [83]. The fluctuations and dissipation are related because they both originate from the same physical source - the chaos associated with the fast motions of the bath. In the thermodynamic limit applied to bath α, Eq. (4.44) reduces to Eq. (4.47), which is the multi-dimensional anaolg of Einstein's fluctuation-dissipation for a Brownian particle diffusing in a thermal bath [1].

The final force in Eq. (4.30), $\mathbf{F}_{ND}^\alpha$, is called the noise induced drift, and is given by

$$\mathbf{F}^{\alpha}_{ND} \quad = \quad \frac{1}{2}\partial_{E^{\alpha}}\left[\sqrt{L^{\text{sy}\,\alpha}}\right]\sqrt{L^{\text{sy}\,\alpha}}M^{-1}\mathbf{P}. \tag{4.49}$$

We are unfortunately at the moment lacking a satisfactory physical interpretation of $\mathbf{F}^{\alpha}_{ND}$. The noise induced drift appeared when converting from the Ito to the Stratonovich calculus, and it is not clear whether or not $\mathbf{F}^{\alpha}_{ND}$ has an intuitive physical explanation, or if it is a mathematical artifact which arises because we are approximating a deterministic process by a random white noise process. Regardless, we see that $\mathbf{F}^{\alpha}_{ND}$ is related to the strength of fluctuations and produces "reverse dissipation," and by multiplying and dividing by $k_B T^{\alpha}$, we see that $\mathbf{F}^{\alpha}_{ND}$ will be negligible in comparison to deterministic friction in the thermodynamic limit applied to bath α.

4.3 Path integrals

In this section, we study transition probabilities implied by our Langevin equation (4.29) using the path integral formalism. Our motivation stems from a fundamental result in stochastic thermodynamics which relates stochastic action along a trajectory to the entropy generation and heat dissipation - we would like to determine whether or not the analog of this result provides useful notions heat and entropy production in the chaos bath framework. For this section and the remainder of chapter 4, we will assume the presence of only a single bath. This assumption will greatly simplify our calculations, and will allow cleaner comparisons with stochastic thermodynamics, which is most simply formulated for the case of single fixed temperature thermal environment.

For a single chaos bath, the energy of the bath, E, is related to the energy of the entire universe, $\mathcal{E}$, and the coordinates of the system of interest by $E = \mathcal{E} - H(\mathbf{X}, \boldsymbol{\lambda})$, so we can equally well consider the mesoscale evolution of $(\mathbf{X}(t), E(t))$ or $(\mathbf{X}(t), \mathcal{E}(t))$ - both trajectories contain the same information. The differential for $\mathcal{E}(t)$, evaluated along mesoscale trajectories, is given by

$$\mathrm{d}\mathcal{E} \;=\; \mathrm{d}E + \partial_{\mathbf{X}}H(\mathbf{X},\boldsymbol{\lambda})\cdot\mathrm{d}\mathbf{X} + \partial_{\boldsymbol{\lambda}}H(\mathbf{X},\boldsymbol{\lambda})\cdot\mathrm{d}\boldsymbol{\lambda} \tag{4.50}$$

$$\;=\; \partial_{\boldsymbol{\lambda}}V\cdot\mathrm{d}\boldsymbol{\lambda}.$$

The first two terms on the right hand side of the first line of the above equation cancel due to the conservation of energy, which can be seen directly from Eq. (4.28). For notational simplicity, we now define the vector function $\mathbf{f}(\mathbf{X}, E, \boldsymbol{\lambda})$ to be the deterministic portion of the total mesoscopic force acting on the system of interest in Eq. (4.28):

$$\mathbf{f}(\mathbf{X}, E, \boldsymbol{\lambda}) = -\partial_{\mathbf{Q}}V + \mathbf{u} - \left(\frac{1}{2\Sigma}\partial_E\left[\Sigma\,L\right] - \frac{1}{2}\partial_E\left[\sqrt{L^{\mathrm{sy}}}\right]\sqrt{L^{\mathrm{sy}}}\right)M^{-1}\mathbf{P}. \tag{4.51}$$

The Langevin equation for the trajectory $(\mathbf{X}(t), \mathcal{E}(t))$ is thus

$$\mathrm{d}\mathbf{Q} \;=\; M^{-1}\mathbf{P}\mathrm{d}t \tag{4.52}$$

$$\mathrm{d}\mathbf{P} \;=\; \mathbf{f}(\mathbf{X}, E, \boldsymbol{\lambda})\mathrm{d}t + \sqrt{L^{\mathrm{sy}}}(\mathbf{Q}, E)\circ\mathrm{d}\boldsymbol{\mathcal{W}}\big|_{E=\mathcal{E}-H(\mathbf{X},\boldsymbol{\lambda})}$$

$$\mathrm{d}\mathcal{E} \;=\; \partial_{\boldsymbol{\lambda}}V(\mathbf{X}, \boldsymbol{\lambda})\cdot\dot{\boldsymbol{\lambda}}\mathrm{d}t.$$

The expression $|_{E=\mathcal{E}-H(\mathbf{X},\boldsymbol{\lambda})}$ indicates that $\mathbf{f}$ and $\sqrt{L^{\mathrm{sy}}}$ are evaluated for $E = \mathcal{E}-H(\mathbf{X},\boldsymbol{\lambda})$ in the equation for $\mathrm{d}\mathbf{P}$. We will use the Langevin equation in the form of Eq. (4.52) throughout the remainder of this chapter.

4.3.1 Stochastic thermodynamics

Before studying path integrals and transition probabilities in the chaos bath framework, we briefly review the corresponding quantities in the simpler framework of stochastic thermodynamics. This will help establish some notation and motivate the work to follow. The important results of stochastic thermodynamics are demonstrated clearly by an overdamped one-dimensional Brownian particle in a single thermal bath. This system has been referred

to as "the paradigm" for the field of stochastic thermodynamics and is reviewed in detail in [13]. The Langevin equation for the position x of an overdamped Brownian particle is given by

$$b\, \mathrm{d}x = -\partial_x V(x,\lambda)\mathrm{d}t + \sqrt{D} \circ \mathrm{d}\mathcal{W}. \tag{4.53}$$

In Eq. (4.53), b is the friction constant, V is a position dependent potential energy which can be manipulated by the scalar control parameter λ, $\mathcal{W}$ is a standard Wiener process, and D is the diffusion constant, which is given by the Einstein relation:

$$D = 2\, b\, k_B\, T, \tag{4.54}$$

where T is the temperature of the environment. If the particle is displaced by $\mathrm{d}x$ and if the control parameter is incremented by $\mathrm{d}\lambda$ during a short time $\mathrm{d}t$, the increment of work done on the particle by the agent manipulating the control protocol is given by [13, 16]

$$\mathrm{d}w = \partial_\lambda V(x,\lambda)\mathrm{d}\lambda, \tag{4.55}$$

The formula in Eq. (4.55) uses the inclusive definition of work, which is the appropriate definition when is done on a system through modification of its internal potential energy [88]. The increment of heat energy transferred from the environment to the particle is given by [13, 16]

$$\begin{aligned} \mathrm{d}q &= -b\frac{(\mathrm{d}x)^2}{\mathrm{d}t} + \sqrt{D}\frac{\mathrm{d}\mathcal{W}\mathrm{d}x}{\mathrm{d}t} \\ &= \partial_x V(x,\lambda)\mathrm{d}x. \end{aligned} \tag{4.56}$$

If $\mathrm{d}q$ denotes the increment of heat energy transferred to the particle, then $-\mathrm{d}q$ gives the increment of heat energy dissipated to the environment. We note that, for infinitesimal $\mathrm{d}t$, we have $\mathrm{d}\mathcal{W} \sim \sqrt{\mathrm{d}t}$ generally, and $\mathrm{d}x \sim \sqrt{\mathrm{d}t}$ for an overdamped Brownian particle, so the first line of Eq. (4.56) does not vanish identically. Equations (4.55) and (4.56) together give

the first law of thermodynamics at the level of individual trajectories [16]:

$$\begin{aligned} \mathrm{d}V &= \partial_\lambda V \mathrm{d}\lambda + \partial_x V \mathrm{d}x \\ &= \mathrm{d}w + \mathrm{d}q. \end{aligned} \tag{4.57}$$

Changes in the particle's kinetic energy are negligible in the overdamped limit, so the change in potential energy $\mathrm{d}V$ is equivalent to the change in total particle energy [13].

Associated with the SDE in Eq. (4.25) are probabilities (or more properly, probability densities) for various transitions and paths the particle may follow as it evolves in time. We denote the probability for the particle to transition from a point x_0 at time t_0 to a point x_f at time t_f, given a pre-specified control protocol $\lambda(\tau)$, as $P_{\lambda(\tau)}(x_f, t_f \,;\, x_0, t_0)$, and we denote the probability to observe a specific path $x(\tau)$ which accomplishes this transition as $P_{\lambda(\tau)}[x(\tau)|x_f, t_f \,;\, x_0, t_0]$. In other words, imagine an experiment where we continuously measure the position of Brownian particle, initially located at position x_0, over the time interval $[t_0, t_f]$ while the implementing the control protocol $\lambda(\tau)$. If we repeat this experiment an infinite number of times, we will generate an ensemble of all possible particle trajectories $x(\tau)$ with various final positions x_f; each member of the ensemble represents one possible realization of the dynamics governed by Eq. (4.53) under the protocol $\lambda(\tau)$, given the initial position x_0, initial time t_0, and final time t_f. The transition probability $P_{\lambda(\tau)}(x_f, t_f \,;\, x_0, t_0)$ is a measure of the fraction of ensemble members which end at position x_f, while the path probability $P_{\lambda(\tau)}[x(\tau)|x_f, t_f \,;\, x_0, t_0]$ is a measure of the fraction of ensemble members which follow a particular trajectory $x(\tau)$ ending at x_f. We will abbreviate the path probability by $P_{\lambda(\tau)}[x(\tau)]$ when the boundary conditions are implied by context. The path and transition probabilities are related by a path integral:

$$P_{\lambda(\tau)}(x_f, t_f \,;\, x_0, t_0) = \int \mathcal{D}[x(\tau)] P_{\lambda(\tau)}[x(\tau)]. \tag{4.58}$$

This path integral is an integral over an infinite dimensional trajectory space, where $\mathcal{D}[x(\tau)]$ is an infinitesimal volume element centred at the trajectory $x(\tau)$. An expression for $P_{\lambda(\tau)}[x(\tau)]$

can be derived using the path and transition probabilities associated with the standard Wiener process. The details of the calculation are given in Ref. [16], with the result

$$P_{\lambda(\tau)}[x(\tau)] = e^{-\mathcal{A}[x(\tau),\lambda(\tau)]}, \tag{4.59}$$

which implies

$$P_{\lambda(\tau)}(x_f, t_f \, ; \, x_0, t_0) = \int \mathcal{D}[x(\tau)] e^{-A[x(\tau),\lambda(\tau)]}, \tag{4.60}$$

where the stochastic action $A[x(\tau), \lambda(\tau)]$ is given in terms of the stochastic Lagrangian $\mathcal{L}(x(t), \dot{x}(t), \lambda(t))$:

$$
\begin{aligned}
A[x(\tau), \lambda(\tau)] &= \int_{t_0}^{t_f} \mathrm{d}t \, \mathcal{L}(x(t), \dot{x}(t), \lambda(t)) \\
&= \int_{t_0}^{t_f} \mathrm{d}t \, \frac{\left(\dot{x}(t) + \frac{1}{b}\partial_x V|_{x(t),\lambda(t)}\right)^2}{2D/b^2} - \frac{1}{2b}\partial_{xx}^2 V|_{x(t),\lambda(t)}.
\end{aligned}
\tag{4.61}
$$

It must be noted that $\dot{x}(t)$ is not well-defined due to the non-differentiable driving noise, so Eq. (4.60) is only a formal expression. The transition probability can be more properly written by discretizing the time interval $t_f - t_0$ and taking a limit. We discretize the time interval into N blocks of length Δt, and we make the following definitions:

$$
\begin{aligned}
t_k &= t_0 + k\Delta t \\
x_k &= x(t_k) \\
x_k^* &= \frac{x_k + x_{k-1}}{2} \\
\Delta x_k &= x_k - x_{k-1} \\
\lambda_k &= \lambda(t_k) \\
\lambda_k^* &= \frac{\lambda_k + \lambda_{k-1}}{2},
\end{aligned}
\tag{4.62}
$$

where $k = 1, ..., N$, and $t_N = t_f$. The transition probability can then be written as

$$P_{\lambda(\tau)}(x_f, t_f \,;\, x_0, t_0) = \int_{-\infty}^{\infty} \prod_{k=1}^{N-1} \left(\frac{\mathrm{d}x_k}{\sqrt{2\pi D \Delta t}} \right) e^{- \sum_{k=1}^{N} \Delta t \, \mathcal{L}\left(x_k^*, \frac{\Delta x_k}{\Delta t}, \lambda_k^* \right)}. \tag{4.63}$$

The midpoint convention for writing x_k^* and λ_k^* was chosen because we are interpreting stochastic noise in the Stratonovich sense. We now denote by Lim the operation of taking limits $N \to \infty$, $\Delta t \to 0$ with $N \Delta t = t_f - t_0$ fixed. The the path integral expression for the transition probability can then be defined by

$$
\begin{aligned}
P_{\lambda(\tau)}(x_f, t_f \,;\, x_0, t_0) &= \mathrm{Lim} \int_{-\infty}^{\infty} \prod_{k=1}^{N-1} \left(\frac{\mathrm{d}x_k}{\sqrt{2\pi D \Delta t}} \right) e^{- \sum_{k=1}^{N} \Delta t \, \mathcal{L}\left(x_k^*, \frac{\Delta x_k}{\Delta t}, \lambda_k^* \right)} \tag{4.64} \\
&= \int \mathcal{D}[x(\tau)] e^{-\mathcal{A}[x(\tau), \lambda(\tau)]}.
\end{aligned}
$$

A fundamental result in stochastic thermodynamics relating heat to stochastic action can be derived by considering the reverse evolution of particle trajectories. For each trajectory $x(\tau)$ which evolves according to Eq. (4.53) under the protocol $\lambda(\tau)$, we associate a conjugate trajectory $x^\dagger(\tau)$ and conjugate protocol $\lambda^\dagger(\tau)$ defined such that, for each $t \in [t_0, t_f]$, we have

$$
\begin{aligned}
x^\dagger(t) &= x\big(t_f - (t - t_0)\big) \tag{4.65} \\
&= x(t^\dagger) \\
\lambda^\dagger(t) &= \lambda\big(t_f - (t - t_0)\big) \\
&= \lambda(t^\dagger),
\end{aligned}
$$

where $t^\dagger = t_f - (t - t_0)$ defines the conjugate time. We will henceforth refer the $x(\tau)$ and $\lambda(\tau)$ as the forward trajectory and forward protocol, respectively. The conjugate trajectories are assumed to be generated by an SDE in the same form as that which generates the forward trajectories:

$$b \, \mathrm{d}x^\dagger = -\partial_{x^\dagger} V(x^\dagger, \lambda^\dagger) \mathrm{d}t + \sqrt{D} \circ \mathrm{d}\mathcal{W}. \tag{4.66}$$

We refer to Eq. (4.66) as the conjugate dynamics, while we refer to Eq. (4.53) as the forward dynamics. Associated with the conjugate dynamics are the conjugate transition and path probabilities, $P^\dagger_{\lambda^\dagger(\tau)}(x^\dagger_f, t_f\,;\, x^\dagger_0, t_0)$ and $P^\dagger_{\lambda(\tau)}[x^\dagger(\tau)]$, defined and interpreted analogously to their forward dynamics counterparts. Note that because x and $x^\dagger$ obey SDE's of the same form, we have

$$P^\dagger_{\lambda(\tau)}[x^\dagger(\tau)] = e^{-\mathcal{A}[x^\dagger(\tau), \lambda^\dagger(\tau)]}, \tag{4.67}$$

where the action functional $\mathcal{A}$ is defined in Eq. (4.61).

We denote the heat absorbed by the system along the forward trajectory $x(\tau)$ by the functional $q[x(\tau)]$. After some manipulation, one can use the definition of the heat increment $\mathrm{d}q$ given in Eq. (4.56) to show the following important relation [13]:

$$
\begin{aligned}
q[x(\tau)] &= \int_{t_0}^{t_f} \mathrm{d}q \\
&= k_B\, T \left(\mathcal{A}[x(\tau), \lambda(\tau)] - \mathcal{A}[x^\dagger(\tau), \lambda^\dagger(\tau)] \right),
\end{aligned}
\tag{4.68}
$$

Using Eq. (4.59), the above relation implies

$$\frac{P_{\lambda^\dagger(\tau)}[x^\dagger(\tau)]}{P_{\lambda(\tau)}[x(\tau)]} = e^{\frac{q[x(\tau)]}{k_B\, T}}. \tag{4.69}$$

Because the thermal bath is at a fixed temperature T, the dissipated heat is transferred to the bath isothermally, so the quantity $-q[x(\tau)]/T$ therefore defines the thermodynamic entropy increase of the bath when the system traverses the path $x(\tau)$, which we denote by $\Delta s[x(\tau)]$. We thus have

$$\Delta s[x(\tau)] = -k_B \left(\mathcal{A}[x(\tau), \lambda(\tau)] - \mathcal{A}[x^\dagger(\tau), \lambda^\dagger(\tau)] \right), \tag{4.70}$$

which implies

$$\frac{P_{\lambda^\dagger(\tau)}[x^\dagger(\tau)]}{P_{\lambda(\tau)}[x(\tau)]} = e^{-\frac{\Delta s[x(\tau)]}{k_B}}. \tag{4.71}$$

By using Eqs. (4.69) or (4.71), and by selecting distributions for the forward and conjugate initial conditions, one can derive the various fluctuation theorems of non-equilibrium thermodynamics [13]. Equations (4.69) and (4.71) are themselves sometimes referred to as detailed fluctuation theorems [27], and are other times referred to a expressions of "microscopic reversibility" [17]. According to Eqs. (4.69) and (4.71) if a particular trajectory $x(\tau)$ is allowed under Eq. (4.53), then the conjugate trajectory is allowed as well, but the probability to observe the conjugate trajectory is exponentially smaller (assuming positive Δs) than the probability to observe the corresponding forward trajectory.

4.3.2 Stochastic action

We will now derive expressions for the path probabilities and stochastic action for our system and bath defined by Eq. (4.52). The diffusion matrix corresponding to Eq. (4.52) is given by

$$\begin{pmatrix} 0 & 0 & \mathbf{0} \\ 0 & L^{\text{sy}} & \mathbf{0} \\ \mathbf{0}^T & \mathbf{0}^T & 0 \end{pmatrix}, \tag{4.72}$$

with corresponding dimensionalities

$$\begin{pmatrix} N_S \times N_S & N_S \times N_S & N_S \times 1 \\ N_S \times N_S & N_S \times N_S & N_S \times 1 \\ 1 \times N_S & 1 \times N_S & 1 \times 1 \end{pmatrix}. \tag{4.73}$$

This diffusion matrix is neither constant nor invertible, so we can not simply apply the stochastic thermodynamics formulas given by Eqs. (4.59) and (4.61) to our system and chaos bath. Formulas for the path probabilities and stochastic action corresponding to a multi-dimensional Stratonovich SDE with a non-constant diffusion matrix is given in Refs. [89] and [90], but an invertible diffusion matrix is assumed. Reference [91] gives a

method to calculate path integrals and stochastic action for Stratonovich SDE's with non-invertible diffusion matrices, but the diffusion matrices are assumed to be constant. We will preform our calculation by combining the techniques of Refs. [89, 90, 91].

The non-invertibility of the diffusion matrix is due to the noiseless equations for $\mathbf{Q}$ and $\mathcal{E}$, which impose constraints on the stochastic evolution of $(\mathbf{X}, \mathcal{E})$. These constraints can be enforced "by hand" by introducing delta functionals in the path integral expression for the transition probability $P_{\boldsymbol{\lambda}(\tau)}(\mathbf{X}_f, \mathcal{E}_f, t_f; \mathbf{X}_0, \mathcal{E}_0, t_0)$. More formally, we can simultaneously enforce the constraints and avoid the non-invertible diffusion matrix by introducing artificial noise terms in Eq. (4.52), calculating the transition probabilities, and then taking the limit of the artificial noise strength going to zero [91]. As we will see, the artificial noises result in additional Gaussian terms present in the discretized expression for $P_{\boldsymbol{\lambda}(\tau)}(\mathbf{X}_f, \mathcal{E}_f, t_f; \mathbf{X}_0, \mathcal{E}_0, t_0)$, and these Gaussians will reduce to delta functions when the noise strength goes to zero. The delta functions signify that the stochastic paths taken by the system (4.52) are constrained to lie on some hypersurface in infinite dimensional path space.

Instead of Eq. (4.52), we consider the SDE

$$
\begin{aligned}
d\mathbf{Q} &= M^{-1}\mathbf{P}dt + \sqrt{D_{\mathbf{Q}}} \circ d\boldsymbol{\mathcal{W}}^{\mathbf{Q}} \qquad\qquad (4.74)\\
d\mathbf{P} &= \mathbf{f}(\mathbf{X}, E, \boldsymbol{\lambda})dt + \sqrt{L^{\mathrm{sy}}}(E, \mathbf{Q}) \circ d\boldsymbol{\mathcal{W}}|_{E=\mathcal{E}-H(\mathbf{X}, \boldsymbol{\lambda})}\\
d\mathcal{E} &= \partial_{\boldsymbol{\lambda}} V(\mathbf{Q}, \boldsymbol{\lambda}) \cdot \dot{\boldsymbol{\lambda}}dt + \sqrt{D_{\mathcal{E}}} \circ d\mathcal{W}^{\mathcal{E}}.
\end{aligned}
$$

The quantities $D_{\mathbf{Q}}$ and $D_{\mathcal{E}}$ are numbers that will eventually be set to zero, and $\boldsymbol{\mathcal{W}}^{\mathbf{Q}}$ and $\mathcal{W}^{\mathcal{E}}$ are N_S and one dimensional Wiener processes, respectively, independent of each other and the N_S dimensional Wiener process $\boldsymbol{\mathcal{W}}$ appearing in the expression for $d\mathbf{P}$. If the matrix $\sqrt{L^{\mathrm{sy}}}$ is invertible, then the diffusion matrix corresponding to Eq. (4.74) will be invertible as well. If $\sqrt{L^{\mathrm{sy}}}$ is not invertible, then diffusion matrix corresponding to Eq. (4.74) will possess one or more vanishing eigenvalues, so an invertible diffusion matrix can be generated by adding additional artificial noises in the corresponding eigendirections. Adding these additional artificial noises will, in the limit of vanishing noise strength, result in additional

constraints imposed on the allowed trajectories through the presence of additional delta functionals in the path integral expression for transition probabilities, but otherwise don't significantly alter our results, so for simplicity, we will assume that $\sqrt{L^{\text{sy}}}$ is invertible for the rest of this chapter.

We now discretize the time interval $t_f - t_0$ into N intervals of length Δt, and we define the following:

$$
\begin{aligned}
t_k &= t_0 + k\Delta t \qquad\qquad\qquad (4.75) \\
\mathbf{X}_k &= \mathbf{X}(t_k) \\
\mathbf{X}_k^* &= \frac{\mathbf{X}_k + \mathbf{X}_{k-1}}{2} \\
\Delta\mathbf{X}_k &= \mathbf{X}_k - \mathbf{X}_{k-1},
\end{aligned}
$$

where $k = 1, ..., N$, and $t_N = t_f$. We also define the quantities $\mathcal{E}_k, \mathcal{E}_k^*, \Delta\mathcal{E}_k, \boldsymbol{\lambda}_k, \boldsymbol{\lambda}_k^*$, and $\Delta\boldsymbol{\lambda}_k$ analogously. We again denote by Lim the limit $N \to \infty, \Delta t \to 0$ with $N\Delta t = t_f - t_0$ fixed. Applying the formulae given in Refs. [89] and [90], we find the transition probabilities $P_{\boldsymbol{\lambda}(\tau)}^{\text{noise}}(\mathbf{X}_f, \mathcal{E}_f, t_f\,;\,\mathbf{X}_0, \mathcal{E}_0, t_0)$ for the Langevin dynamics with artifical noise:

$$
\begin{aligned}
P_{\boldsymbol{\lambda}(\tau)}^{\text{noise}}(\mathbf{X}_f, \mathcal{E}_f, t_f\,;\,\mathbf{X}_0, \mathcal{E}_0, t_0) &= \int_{-\infty}^{\infty} \left[\prod_{k=1}^{N-1} \frac{d\mathbf{P}_k}{[2\pi\Delta t]^{\frac{N_s}{2}} \left| \sqrt{L^{\text{sy}}}(E_k^*, \mathbf{Q}_k^*) \right|} \right. \qquad (4.76) \\
&\qquad\qquad \left. \times \frac{d\mathbf{Q}_k}{[2\pi D_{\mathbf{Q}}\Delta t]^{\frac{N_s}{2}}} \frac{d\mathcal{E}_k}{[2\pi D_{\mathcal{E}}\Delta t]^{\frac{1}{2}}} \right] \\
&\qquad \times e^{-\sum_{k=1}^{N} \Delta t \frac{1}{2D_{\mathcal{E}}} \left(\frac{\Delta\mathcal{E}_k}{\Delta t} - \partial_{\boldsymbol{\lambda}_k^*} V(\mathbf{Q}_k^*, \boldsymbol{\lambda}_k^*) \cdot \frac{\Delta\boldsymbol{\lambda}_k}{\Delta t} \right)^2} \\
&\qquad \times e^{-\sum_{k=1}^{N} \Delta t \frac{1}{2D_{\mathbf{Q}}} \left(\frac{\Delta\mathbf{Q}_k}{\Delta t} - M^{-1}\mathbf{P}_k^* \right)^2} \\
&\qquad \times e^{-\sum_{k=1}^{N} \Delta t \, \mathcal{L}(\mathbf{X}_k^*, \mathcal{E}_k^*, \boldsymbol{\lambda}_k^*, \frac{\Delta\mathbf{P}_k}{\Delta t})},
\end{aligned}
$$

where $\left| \sqrt{L^{\text{sy}}}(E_k^*, \mathbf{Q}_k^*) \right|$ is the magnitude of the determinant of $\sqrt{L^{\text{sy}}}$, evaluated at $\mathbf{Q}_k^*$ and $E_k^* = \mathcal{E}_k^* - H(\mathbf{X}_k^*, \boldsymbol{\lambda}_k^*)$, and the stochastic Lagrangian $\mathcal{L}(\mathbf{X}, \mathcal{E}, \boldsymbol{\lambda}, \dot{\mathbf{P}})$ is given by

$$
\begin{aligned}
\mathcal{L}(\mathbf{X}, \mathcal{E}, \boldsymbol{\lambda}, \dot{\mathbf{P}}) \;=\; & \frac{1}{2} \left(\dot{\mathbf{P}} - \mathbf{f}(\mathbf{X}, E, \boldsymbol{\lambda}) \right)^T L^{\mathrm{sy}}(E, \mathbf{Q})^{-1} \left(\dot{\mathbf{P}} - \mathbf{f}(\mathbf{X}, E, \boldsymbol{\lambda}) \right) \\
& + \frac{1}{2} \partial_{\mathbf{P}} \cdot \mathbf{f}(\mathbf{X}, E, \boldsymbol{\lambda}) - \frac{1}{2} \mathbf{P}^T M^{-1} \partial_E \mathbf{f}(\mathbf{X}, E, \boldsymbol{\lambda}) \\
& - \frac{1}{2} \mathbf{P}^T M^{-1} \partial_E \left[\sqrt{L^{\mathrm{sy}}}(E, \mathbf{Q}) \right] \sqrt{L^{\mathrm{sy}}}(E, \mathbf{Q})^{-1} \left(\dot{\mathbf{P}} - \mathbf{f}(\mathbf{X}, E, \boldsymbol{\lambda}) \right) \\
& + \frac{1}{8} \mathbf{P}^T M^{-1} \partial_E \left[\sqrt{L^{\mathrm{sy}}}(E, \mathbf{Q}) \right] \partial_E \left[\sqrt{L^{\mathrm{sy}}}(E, \mathbf{Q}) \right] M^{-1} \mathbf{P} \Bigg|_{E = \mathcal{E} - H(\mathbf{X}, \boldsymbol{\lambda})}.
\end{aligned}
\tag{4.77}
$$

By taking the limits $D_{\mathcal{E}} \to 0$ and $D_{\mathbf{Q}} \to 0$ in Eq. (4.76), and making use of the identity

$$
\lim_{\sigma \to 0} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x - x')^2}{2\sigma^2}} = \delta(x - x'),
\tag{4.78}
$$

we arrive at the transition probabilities for the actual Langevin (no artificial noise) dynamics:

$$
\begin{aligned}
P_{\boldsymbol{\lambda}(\tau)}(\mathbf{X}_f, \mathcal{E}_f, t_f \,;\, \mathbf{X}_0, \mathcal{E}_0, t_0) \;=\; & \int_{-\infty}^{\infty} \left[\prod_{k=1}^{N-1} \frac{\mathrm{d}\mathbf{P}_k}{[2\pi\Delta t]^{\frac{N_s}{2}} \left| \sqrt{L^{\mathrm{sy}}}(E_k^*, \mathbf{Q}_k^*) \right|} \right. \\
& \times \mathrm{d}\mathbf{Q}_k \, \delta \left(\Delta\mathbf{Q}_k - M^{-1} \mathbf{P}_k^* \Delta t \right) \\
& \left. \times \mathrm{d}\mathcal{E}_k \, \delta \left(\Delta\mathcal{E}_k - \partial_{\boldsymbol{\lambda}_k^*} V(\mathbf{Q}_k^*, \boldsymbol{\lambda}_k^*) \cdot \frac{\Delta\boldsymbol{\lambda}_k}{\Delta t} \Delta t \right) \right] \\
& \times e^{- \sum\limits_{k=1}^{N} \Delta t \, \mathcal{L}(\mathbf{X}_k^*, \mathcal{E}_k^*, \boldsymbol{\lambda}_k^*, \frac{\Delta\mathbf{P}_k}{\Delta t})}.
\end{aligned}
\tag{4.79}
$$

Taking the Lim of Eq. (4.79), we arrive at the path integral expression for the transition probability $P_{\boldsymbol{\lambda}(\tau)}(\mathbf{X}_f, \mathcal{E}_f, t_f \,;\, \mathbf{X}_0, \mathcal{E}_0, t_0)$:

$$
\begin{aligned}
P_{\boldsymbol{\lambda}(\tau)}(\mathbf{X}_f, \mathcal{E}_f, t_f \,;\, \mathbf{X}_0, \mathcal{E}_0, t_0) \;=\; & \int \mathcal{D}_{\boldsymbol{\lambda}(\tau)}[\mathbf{X}(\tau), \mathcal{E}(\tau)] P_{\boldsymbol{\lambda}(\tau)}[\mathbf{X}(\tau), \mathcal{E}(\tau)] \\
\;=\; & \int \mathcal{D}_{\boldsymbol{\lambda}(\tau)}[\mathbf{X}(\tau), \mathcal{E}(\tau)] e^{-\mathcal{A}[\mathbf{X}(\tau), \mathcal{E}(\tau), \boldsymbol{\lambda}(\tau)]}
\end{aligned}
\tag{4.80}
$$

where the stochastic action $\mathcal{A}[\mathbf{X}(\tau), \mathcal{E}(\tau), \boldsymbol{\lambda}(\tau)]$ is given by

124

$$\mathcal{A}[\mathbf{X}(\tau), \mathcal{E}(\tau), \boldsymbol{\lambda}(\tau)] = \int_{t_0}^{t_f} \mathrm{d}t \, \mathcal{L}(\mathbf{X}(t), \mathcal{E}(t), \boldsymbol{\lambda}(t), \dot{\mathbf{P}}(t)), \tag{4.81}$$

with the stochastic Lagrangian $\mathcal{L}(\mathbf{X}, \mathcal{E}, \boldsymbol{\lambda}, \dot{\mathbf{P}})$ is defined in Eq. (4.77). We note that the volume element $\mathcal{D}_{\boldsymbol{\lambda}(\tau)}[\mathbf{X}(\tau), \mathcal{E}(\tau)]$ contains two delta functionals:

$$\delta \left[\mathbf{Q}(\tau) - \left(\mathbf{Q}_0 + \int_{t_0}^{\tau} \mathrm{d}t \, M^{-1}\mathbf{P}(t) \right) \right] = \mathrm{Lim} \prod_{k=1}^{N-1} \delta \left(\Delta\mathbf{Q}_k - M^{-1}\mathbf{P}_k^*\Delta t \right), \tag{4.82}$$

and

$$\delta \left[\mathcal{E}(\tau) - \left(\mathcal{E}_0 + \int_{t_0}^{\tau} \mathrm{d}t \, \partial_{\boldsymbol{\lambda}}(\mathbf{Q}, \boldsymbol{\lambda})|_{\mathbf{Q}(t),\boldsymbol{\lambda}(t)} \, \dot{\boldsymbol{\lambda}}(t) \right) \right] = \tag{4.83}$$
$$\mathrm{Lim} \prod_{k=1}^{N-1} \delta \left(\Delta\mathcal{E}_k - \partial_{\boldsymbol{\lambda}_k^*} V(\mathbf{Q}_k^*, \boldsymbol{\lambda}_k^*) \cdot \frac{\Delta\boldsymbol{\lambda}_k}{\Delta t}\Delta t \right).$$

In contrast to ordinary stochastic thermodynamics, the volume element in path space depends on the control protocol $\boldsymbol{\lambda}(t)$ through the $\mathcal{E}(\tau)$ delta functional.

4.3.3 Conjugate dynamics

Our goal now is to use the path integral formalism to find the chaos bath analog of the fundamental thermodynamic relations given in Eqs. (4.70) and (4.71). In stochastic thermodynamics, the fundamental relations appear when one considers the conjugate trajectories of a Brownian particle immersed in the thermal bath, so we must likewise consider the conjugate trajectories of our system interest immersed in the chaos bath. However, because the mesoscopic forces appearing in the chaos bath Langevin equation (4.52) are defined in terms of the microscopic forces appearing in Hamilton's equations (3.3), we will additionally need to consider the microscopic conjugate trajectories of the bath particles, as well as the microscopic equations of motion governing the dynamics of the conjugate trajectories.

Suppose that during the time interval $[t_0, t_f]$, the system and bath together follow some specific trajectory $(\mathbf{Q}(\tau), \mathbf{q}(\tau), \mathbf{P}(\tau), \mathbf{p}(\tau))$ under Hamilton's equations for some control protocol $\boldsymbol{\lambda}(\tau)$. We will refer to this trajectory as the forward trajectory. We denote the corresponding conjugate trajectory by $(\mathbf{Q}^\dagger(\tau), \mathbf{q}^\dagger(\tau), \mathbf{P}^\dagger(\tau), \mathbf{p}^\dagger(\tau))$, defined such that, for each $t \in [t_0, t_f]$, we have

$$
\begin{aligned}
\mathbf{Q}^\dagger(t) &= \mathbf{Q}(t^\dagger) \\
\mathbf{q}^\dagger(t) &= \mathbf{q}(t^\dagger) \\
\mathbf{P}^\dagger(t) &= -\mathbf{P}(t^\dagger) \\
\mathbf{p}^\dagger(t) &= -\mathbf{p}(t^\dagger),
\end{aligned}
\tag{4.84}
$$

where $t^\dagger = t_f - (t - t_0)$ denotes the conjugate time. Taking time derivatives of the conjugate trajectories, and noting that $\frac{\mathrm{d}t^\dagger}{\mathrm{d}t} = -1$, we find

$$
\begin{aligned}
\dot{\mathbf{Q}}^\dagger(t) &= -\dot{\mathbf{Q}}(t^\dagger) \\
\dot{\mathbf{q}}^\dagger(t) &= -\dot{\mathbf{q}}(t^\dagger) \\
\dot{\mathbf{P}}^\dagger(t) &= \dot{\mathbf{P}}(t^\dagger) \\
\dot{\mathbf{p}}^\dagger(t) &= \dot{\mathbf{p}}(t^\dagger).
\end{aligned}
\tag{4.85}
$$

We also define the conjugate control protocol $\boldsymbol{\lambda}^\dagger(\tau)$ such that, for all $t \in [t_0, t_f]$, we have

$$
\boldsymbol{\lambda}^\dagger(t) = \boldsymbol{\lambda}(t^\dagger).
\tag{4.86}
$$

Finally, if the bath and universe have energies $E(\tau)$ and $\mathcal{E}(\tau)$, respectively, along the forward trajectory, then the bath's and universe's energies along the conjugate trajectory, denoted by $E^\dagger(\tau)$ and $\mathcal{E}^\dagger(\tau)$, respectively, are given by

$$E^\dagger(t) = E(t^\dagger) \tag{4.87}$$

$$\mathcal{E}^\dagger(t) = \mathcal{E}(t^\dagger).$$

In terms of the forward Hamiltonian (3.1), the conjugate energies are given by

$$
\begin{aligned}
E^\dagger(t) &= h(\mathbf{q}(t^\dagger), \mathbf{p}(t^\dagger); \mathbf{Q}(t^\dagger)) \\
&= h(\mathbf{q}^\dagger(t), -\mathbf{p}^\dagger(t); \mathbf{Q}^\dagger(t)) \\
\mathcal{E}^\dagger(t) &= \mathcal{H}(\mathbf{Q}(t^\dagger), \mathbf{q}(t^\dagger), \mathbf{P}(t^\dagger), \mathbf{p}(t^\dagger), \boldsymbol{\lambda}(t^\dagger)) \\
&= \mathcal{H}(\mathbf{Q}^\dagger(t), \mathbf{q}^\dagger(t), -\mathbf{P}^\dagger(t), -\mathbf{p}^\dagger(t), \boldsymbol{\lambda}^\dagger(t)),
\end{aligned}
\tag{4.88}
$$

and the dynamics associated with the conjugate trajectories are given by

$$
\begin{aligned}
\dot{\mathbf{Q}}^\dagger(t) &= -\partial_{\mathbf{P}}\mathcal{H}\big|_{\mathbf{Q}(t^\dagger),\mathbf{q}(t^\dagger),\mathbf{P}(t^\dagger),\mathbf{p}(t^\dagger),\boldsymbol{\lambda}(t^\dagger)} \\
&= -\partial_{\mathbf{P}}\mathcal{H}\big|_{\mathbf{Q}^\dagger(t),\mathbf{q}^\dagger(t),-\mathbf{P}^\dagger(t),-\mathbf{p}^\dagger(t),\boldsymbol{\lambda}^\dagger(t)} \\
\dot{\mathbf{q}}^\dagger(t) &= -\partial_{\mathbf{p}}h\big|_{\mathbf{q}(t^\dagger),\mathbf{p}(t^\dagger);\mathbf{Q}(t^\dagger)} \\
&= -\partial_{\mathbf{p}}h\big|_{\mathbf{q}^\dagger(t),-\mathbf{p}^\dagger(t);\mathbf{Q}^\dagger(t)} \\
\dot{\mathbf{P}}^\dagger(t) &= -\partial_{\mathbf{Q}}\mathcal{H}\big|_{\mathbf{Q}(t^\dagger),\mathbf{q}(t^\dagger),\mathbf{P}(t^\dagger),\mathbf{p}(t^\dagger),\boldsymbol{\lambda}(t^\dagger)} \\
&= -\partial_{\mathbf{Q}}\mathcal{H}\big|_{\mathbf{Q}^\dagger(t),\mathbf{q}^\dagger(t),-\mathbf{P}^\dagger(t),-\mathbf{p}^\dagger(t),\boldsymbol{\lambda}^\dagger(t)} \\
\dot{\mathbf{p}}^\dagger(t) &= -\partial_{\mathbf{q}}h\big|_{\mathbf{q}(t^\dagger),\mathbf{p}(t^\dagger);\mathbf{Q}(t^\dagger)} \\
&= -\partial_{\mathbf{q}}h\big|_{\mathbf{q}^\dagger(t),-\mathbf{p}^\dagger(t);\mathbf{Q}^\dagger(t)}.
\end{aligned}
\tag{4.89}
$$

Motivated by Eqs. (4.88) and (4.89), we define conjugate bath Hamiltonian $h^\dagger$ and the conjugate Hamiltonian of the universe $\mathcal{H}^\dagger$ by

$$h^\dagger(\mathbf{q}, \mathbf{p}; \mathbf{Q}) = h(\mathbf{q}, -\mathbf{p}; \mathbf{Q}) \tag{4.90}$$

$$\mathcal{H}^\dagger(\mathbf{Q}, \mathbf{q}, \mathbf{P}, \mathbf{p}, \boldsymbol{\lambda}) = \mathcal{H}(\mathbf{Q}, \mathbf{q}, -\mathbf{P}, -\mathbf{p}, \boldsymbol{\lambda}).$$

In terms of the conjugate Hamiltonians, the conjugate energies along the conjugate trajectories are given by

$$E^\dagger(t) = h^\dagger(\mathbf{q}^\dagger(t), \mathbf{p}^\dagger(t); \mathbf{Q}^\dagger(t)) \tag{4.91}$$

$$\mathcal{E}^\dagger(t) = \mathcal{H}^\dagger(\mathbf{Q}^\dagger(t), \mathbf{q}^\dagger(t), \mathbf{P}^\dagger(t), \mathbf{p}^\dagger(t), \boldsymbol{\lambda}^\dagger(t)),$$

and the dynamics associated with the conjugate trajectories are given by

$$\dot{\mathbf{Q}}^\dagger(t) = \partial_{\mathbf{P}} \mathcal{H}^\dagger \Big|_{\mathbf{Q}^\dagger(t), \mathbf{q}^\dagger(t), \mathbf{P}^\dagger(t), \mathbf{p}^\dagger(t), \boldsymbol{\lambda}^\dagger(t)} \tag{4.92}$$

$$\dot{\mathbf{q}}^\dagger(t) = \partial_{\mathbf{p}} h^\dagger \Big|_{\mathbf{q}^\dagger(t), \mathbf{p}^\dagger(t); \mathbf{Q}^\dagger(t)}$$

$$\dot{\mathbf{P}}^\dagger(t) = -\partial_{\mathbf{Q}} \mathcal{H}^\dagger \Big|_{\mathbf{Q}^\dagger(t), \mathbf{q}^\dagger(t), \mathbf{P}^\dagger(t), \mathbf{p}^\dagger(t), \boldsymbol{\lambda}^\dagger(t)}$$

$$\dot{\mathbf{p}}^\dagger(t) = -\partial_{\mathbf{q}} h^\dagger \Big|_{\mathbf{q}^\dagger(t), \mathbf{p}^\dagger(t); \mathbf{Q}^\dagger(t)}.$$

We thus see that the conjugate trajectories obey Hamiltonian dynamics under the conjugate Hamiltonian $\mathcal{H}^\dagger$. If $\mathcal{H}^\dagger(\mathbf{Q}, \mathbf{q}, \mathbf{P}, \mathbf{p}, \boldsymbol{\lambda}) = \mathcal{H}(\mathbf{Q}, \mathbf{q}, \mathbf{P}, \mathbf{p}, \boldsymbol{\lambda})$, then $\mathcal{H}$ is said to possess time-reversal symmetry, and if $\mathcal{H}^\dagger(\mathbf{Q}, \mathbf{q}, \mathbf{P}, \mathbf{p}, \boldsymbol{\lambda}) = \mathcal{H}(\mathbf{Q}', \mathbf{q}', \mathbf{P}', \mathbf{p}', \boldsymbol{\lambda})$, where $(\mathbf{Q}', \mathbf{q}', \mathbf{P}', \mathbf{p}')$ and $(\mathbf{Q}, \mathbf{q}, \mathbf{P}, \mathbf{p})$ are related by a canonical transformation, then $\mathcal{H}$ is said to possess an anti-canonical symmetry. For future reference, we note the following symmetric relation between the microscopic expression of force on the system of interest in the forward and conjugate dynamics:

$$\dot{\mathbf{P}}^\dagger(t) = -\partial_\mathbf{Q} V|_{\mathbf{Q}^\dagger(t),\boldsymbol{\lambda}^\dagger(t)} - \partial_\mathbf{Q} h^\dagger|_{\mathbf{x}^\dagger(t);\mathbf{Q}^\dagger(t)} \tag{4.93}$$

$$= -\partial_\mathbf{Q} V|_{\mathbf{Q}(t^\dagger),\boldsymbol{\lambda}(t^\dagger)} - \partial_\mathbf{Q} h|_{\mathbf{x}(t^\dagger);\mathbf{Q}(t^\dagger)}$$

$$= \dot{\mathbf{P}}(t^\dagger). \tag{4.94}$$

Using the expressions for the conjugate dynamics at the microscopic level, we can find expressions for the conjugate dynamics at the mesoscopic level. To begin, we denote the microcanonical average over the bath of an arbitrary function $G(\mathbf{X}, \mathbf{x}, t)$ in the conjugate dynamics by $\langle G \rangle_{E,\mathbf{Q}}^\dagger$:

$$\langle G \rangle_{E,\mathbf{Q}}^\dagger = \frac{1}{\Sigma^\dagger(E, \mathbf{Q})} \int \mathrm{d}\mathbf{x}\, G(\mathbf{X}, \mathbf{x}, t) \delta(E - h^\dagger(\mathbf{x}; \mathbf{Q})), \tag{4.95}$$

where $\Sigma^\dagger(E, \mathbf{Q})$ denotes the conjugate microcanonical partition function of the bath. Using the definition of $h^\dagger$, we find

$$\Sigma^\dagger(E, \mathbf{Q}) = \int \mathrm{d}\mathbf{q}\mathrm{d}\mathbf{p}\, \delta(E - h^\dagger(\mathbf{q}, \mathbf{p}; \mathbf{Q})) \tag{4.96}$$

$$= \int \mathrm{d}\mathbf{q}\mathrm{d}\mathbf{p}\, \delta(E - h(\mathbf{q}, -\mathbf{p}; \mathbf{Q}))$$

$$= \int \mathrm{d}\mathbf{q}\mathrm{d}\mathbf{p}\, \delta(E - h(\mathbf{q}, \mathbf{p}; \mathbf{Q}))$$

$$= \Sigma(E, \mathbf{Q}).$$

The second to last line of Eq. (4.96) follows from the fact that the magnitude of the Jacobian determinant of the transformation $\mathbf{p} \to -\mathbf{p}$ is unity. By similar reasoning, the conjugate microcanonical partition function of the universe is given by

$$\Sigma^{U\,\dagger}(\mathcal{E}, \boldsymbol{\lambda}) = \int d\mathbf{q} d\mathbf{p} d\mathbf{Q} d\mathbf{P} \, \delta(\mathcal{E} - \mathcal{H}^{\dagger}(\mathbf{Q}, \mathbf{q}, \mathbf{P}, \mathbf{p}, \boldsymbol{\lambda})) \qquad (4.97)$$

$$= \int d\mathbf{q} d\mathbf{p} d\mathbf{Q} d\mathbf{P} \, \delta(\mathcal{E} - \mathcal{H}(\mathbf{Q}, \mathbf{q}, \mathbf{P}, \mathbf{p}, \boldsymbol{\lambda}))$$

$$= \Sigma^{U}(\mathcal{E}, \boldsymbol{\lambda}),$$

Thus, from the definitions in Eqs. (3.104) and (3.105), the microcanonical entropies of the bath and universe in the conjugate dynamics are given by

$$s^{\dagger}(E, \mathbf{Q}) = s(E, \mathbf{Q}), \qquad (4.98)$$

and

$$S^{U\,\dagger}(\mathcal{E}, \boldsymbol{\lambda}) = S^{U}(\mathcal{E}, \boldsymbol{\lambda}), \qquad (4.99)$$

respectively. Likewise, the conjugate adiabatic force $\mathbf{u}^{\dagger}(E, \mathbf{Q})$ is given by

$$\mathbf{u}^{\dagger}(E, \mathbf{Q}) = \left\langle -\partial_{\mathbf{Q}} h^{\dagger} \right\rangle_{E,\mathbf{Q}}^{\dagger} \qquad (4.100)$$

$$= \frac{1}{\Sigma^{\dagger}(E, \mathbf{Q})} \int d\mathbf{q} d\mathbf{p} \left(-\partial_{\mathbf{Q}} h^{\dagger}(\mathbf{q}, \mathbf{p}; \mathbf{Q}) \delta(E - h^{\dagger}(\mathbf{q}, \mathbf{p}; \mathbf{Q})) \right)$$

$$= \frac{1}{\Sigma^{\dagger}(E, \mathbf{Q})} \int d\mathbf{q} d\mathbf{p} \left(-\partial_{\mathbf{Q}} h(\mathbf{q}, \mathbf{p}; \mathbf{Q}) \delta(E - h(\mathbf{q}, \mathbf{p}; \mathbf{Q})) \right)$$

$$= \mathbf{u}(E, \mathbf{Q}).$$

The conjugate symmetric and anti-symmetric covariance matrix, $L^{\mathrm{sy}\,\dagger}(E, \mathbf{Q})$ and $L^{\mathrm{ay}\,\dagger}(E, \mathbf{Q})$, respectively, are defined analogously to their forward dynamics counterparts in Eqs.(3.92) and (3.93):

$$L^{\mathrm{sy}\dagger}(E, \mathbf{Q}) = \int_{-\infty}^{\infty} dt' \left\langle \left(-\partial_{\mathbf{Q}} h^{\dagger} - \mathbf{u}^{\dagger} \right) \left(-\partial_{\mathbf{Q}} h_{t'}^{\dagger} - \mathbf{u}^{\dagger} \right)^{T} \right\rangle_{E,\mathbf{Q}}^{\dagger}, \qquad (4.101)$$

and

$$L^{\mathrm{ay}\dagger}(E,\mathbf{Q}) \;=\; \int_{-\infty}^{0} \mathrm{d}t' \left\langle \left(-\partial_{\mathbf{Q}}h^{\dagger} - \mathbf{u}^{\dagger}\right)\left(-\partial_{\mathbf{Q}}h_{t'}^{\dagger} - \mathbf{u}^{\dagger}\right)^{T} \right\rangle^{\dagger}_{E,\mathbf{Q}} \tag{4.102}$$
$$-\int_{0}^{\infty} \mathrm{d}t' \left\langle \left(-\partial_{\mathbf{Q}}h^{\dagger} - \mathbf{u}^{\dagger}\right)\left(-\partial_{\mathbf{Q}}h_{t'}^{\dagger} - \mathbf{u}^{\dagger}\right)^{T} \right\rangle^{\dagger}_{E,\mathbf{Q}}.$$

In order to evaluate the time integrals above, the bath trajectories must be evolved under $h^{\dagger}$ from $t = -\infty$ to $t = \infty$, so we define $t^{\dagger}$ by $t^{\dagger} = -t$ in this context. For the symmetric part of $L^{\dagger}$, we find

$$L^{\mathrm{sy}\,\dagger}(E,\mathbf{Q}) \;=\; \int_{-\infty}^{\infty} \mathrm{d}t \left\langle \left(-\partial_{\mathbf{Q}}h^{\dagger} - \mathbf{u}^{\dagger}\right)\left(-\partial_{\mathbf{Q}}h_{t}^{\dagger} - \mathbf{u}^{\dagger}\right)^{T} \right\rangle^{\dagger}_{E,\mathbf{Q}} \tag{4.103}$$
$$=\; \int_{-\infty}^{\infty} \mathrm{d}t \left\langle \left(-\partial_{\mathbf{Q}}h - \mathbf{u}\right)\left(-\partial_{\mathbf{Q}}h_{-t} - \mathbf{u}\right)^{T} \right\rangle_{E,\mathbf{Q}}$$
$$=\; (L^{\mathrm{sy}})^{T}(E,\mathbf{Q})$$
$$=\; L^{\mathrm{sy}}(E,\mathbf{Q}).$$

The third of Eq. (4.103) follows from the time-translation invariance of the correlation function for a stationary process [11]. The analogous calculation for $L^{\mathrm{ay}\,\dagger}(E,\mathbf{Q})$ gives

$$L^{\mathrm{ay}\,\dagger}(E,\mathbf{Q}) \;=\; (L^{\mathrm{ay}})^{T}(E,\mathbf{Q}) \tag{4.104}$$
$$=\; -L^{\mathrm{ay}}(E,\mathbf{Q}).$$

Just as with an ordinary magnetic field, the magnetic 2-form field $K^{\mathrm{ay}} = \frac{1}{2\Sigma}\partial_{E}[\Sigma L^{\mathrm{ay}}]$ reverses direction when the "source charges" (the bath degrees of freedom) follow their trajectories in reverse.

With the above definitions in hand, we see that if we were repeat all of the work in chapter 3 and Sec. 4.2 using the conjugate microscopic dynamics, we would arrive at the following Langevin equation for the conjugate mesoscopic dynamics:

$$dQ^\dagger = M^{-1}P^\dagger dt \tag{4.105}$$

$$dP^\dagger = f^\dagger(X^\dagger, E^\dagger, \lambda^\dagger)dt + \sqrt{L^{sy}}(E^\dagger, Q^\dagger) \circ dW|_{E^\dagger = \mathcal{E}^\dagger - H(X^\dagger, \lambda^\dagger)}$$

$$d\mathcal{E}^\dagger = \partial_\lambda V(Q^\dagger, \lambda^\dagger) \cdot \dot{\lambda}^\dagger dt,$$

where the conjugate deterministic mesocopic force $f^\dagger(X, E, \lambda)$ is equal to to the deterministic mesoscopic force $f(X, E, \lambda)$ in Eq. (4.51) with the direction of geometric magnetic force reversed:

$$f^\dagger(X, E, \lambda) = -\partial_Q V + u - \left(\frac{1}{2\Sigma} \partial_E \left[\Sigma \left(L^{sy} - L^{ay} \right) \right] - \frac{1}{2} \partial_E \left[\sqrt{L^{sy}} \right] \sqrt{L^{sy}} \right) M^{-1}P. \tag{4.106}$$

Comparing Eqs. (4.52) and (4.105), we see that $\dot{P}^\dagger(t) \neq \dot{P}(t^\dagger)$ due to deterministic friction and the noise induced drift, which is in contradiction with Eq. (4.93). We thus conclude that the microscopic symmetry relation between the forward and conjugate force on the system of interest is broken at the mesoscale. For future reference, we note that the stationary distribution to the conjugate Fokker-Planck equation corresponding to the distribution in Eq. (3.115) is given by

$$\begin{aligned} P_s^\dagger(Q, P; \mathcal{E}, \lambda) &= \frac{\Sigma^\dagger(\mathcal{E} - H^\dagger(Q, P, \lambda), Q)}{\Sigma^{U\dagger}(\mathcal{E}, \lambda)} \\ &= \frac{\Sigma(\mathcal{E} - H(Q, -P, \lambda), Q)}{\Sigma^U(\mathcal{E}, \lambda)} \\ &= P_s(Q, -P; \mathcal{E}, \lambda), \end{aligned} \tag{4.107}$$

where $H^\dagger(Q, P, \lambda) = H(Q, -P, \lambda)$ is the conjugate system of interest Hamiltonian.

4.3.4   Reversibility, entropy, and the first law

We now have all of the pieces in place to derive chaos bath analog of Eqs. (4.68), (4.69), and (4.71). The stochastic action and corresponding stochastic Lagrangian in the forward

dynamics are given in Eqs. (4.81) and (4.77), respectively. In the conjugate dynamics, the stochastic action and Lagrangian are defined analogously using the conjugate Langevin equation given in Eq. (4.105). The conjugate stochastic action is given by

$$\mathcal{A}^\dagger[\mathbf{X}(\tau), \mathcal{E}(\tau), \boldsymbol{\lambda}(\tau))] = \int_{t_0}^{t_f} \mathrm{d}t \, \mathcal{L}^\dagger(\mathbf{Q}(t), \mathbf{P}(t), \mathcal{E}(t), \boldsymbol{\lambda}(t), \dot{\mathbf{P}}(t)), \tag{4.108}$$

where the conjugate stochastic Lagrangian $\mathcal{L}^\dagger$ is given by

$$
\begin{aligned}
\mathcal{L}^\dagger(\mathbf{X}, \mathcal{E}, \boldsymbol{\lambda}, \dot{\mathbf{P}}) \; = \; & \frac{1}{2} \left(\dot{\mathbf{P}} - \mathbf{f}^\dagger(\mathbf{X}, E, \boldsymbol{\lambda}) \right)^T L^{\mathrm{sy}}(E, \mathbf{Q})^{-1} \left(\dot{\mathbf{P}} - \mathbf{f}^\dagger(\mathbf{X}, E, \boldsymbol{\lambda}) \right) \qquad (4.109) \\
& + \frac{1}{2} \partial_{\mathbf{P}} \cdot \mathbf{f}^\dagger(\mathbf{X}, E, \boldsymbol{\lambda}) - \frac{1}{2} \mathbf{P}^T M^{-1} \partial_E \mathbf{f}^\dagger(\mathbf{X}, E, \boldsymbol{\lambda}) \\
& - \frac{1}{2} \mathbf{P}^T M^{-1} \partial_E \left[\sqrt{L^{\mathrm{sy}}}(E, \mathbf{Q}) \right] \sqrt{L^{\mathrm{sy}}}(E, \mathbf{Q})^{-1} \left(\dot{\mathbf{P}} - \mathbf{f}^\dagger(\mathbf{X}, E, \boldsymbol{\lambda}) \right) \\
& + \frac{1}{8} \mathbf{P}^T M^{-1} \partial_E \left[\sqrt{L^{\mathrm{sy}}}(E, \mathbf{Q}) \right] \partial_E \left[\sqrt{L^{\mathrm{sy}}}(E, \mathbf{Q}) \right] M^{-1} \mathbf{P} \bigg|_{E = \mathcal{E} - H(\mathbf{X}, \boldsymbol{\lambda})}.
\end{aligned}
$$

In contrast to stochastic thermodynamics, we have $\mathcal{A} \neq \mathcal{A}^\dagger$ and $\mathcal{L} \neq \mathcal{L}^\dagger$. This occurs because $\mathbf{f} \neq \mathbf{f}^\dagger$ due to the sign reversal of the geometric magnetism 2-form under the conjugate dynamics. The transition and path probabilities in the conjugate dynamics, denoted $P^\dagger_{\boldsymbol{\lambda}(\tau)}(\mathbf{X}_f, \mathcal{E}_f, t_f ; \mathbf{X}_0, \mathcal{E}_0, t_0)$ and $P^\dagger_{\boldsymbol{\lambda}(\tau)}[\mathbf{X}(\tau), \mathcal{E}(\tau)]$, respectively, are related by

$$
\begin{aligned}
P^\dagger_{\boldsymbol{\lambda}(\tau)}(\mathbf{X}_f, \mathcal{E}_f, t_f ; \mathbf{X}_0, \mathcal{E}_0, t_0) \; &= \; \int \mathcal{D}_{\boldsymbol{\lambda}(\tau)}[\mathbf{X}(\tau), \mathcal{E}(\tau)] P^\dagger_{\boldsymbol{\lambda}(\tau)}[\mathbf{X}(\tau), \mathcal{E}(\tau)] \qquad (4.110) \\
&= \; \int \mathcal{D}_{\boldsymbol{\lambda}(\tau)}[\mathbf{X}(\tau), \mathcal{E}(\tau)] e^{-\mathcal{A}^\dagger[\mathbf{X}(\tau), \mathcal{E}(\tau), \boldsymbol{\lambda}(\tau)]}
\end{aligned}
$$

Consider now a particular transition $(\mathbf{Q}_0, \mathbf{P}_0, \mathcal{E}_0, t_0) \to (\mathbf{Q}_f, \mathbf{P}_f, \mathcal{E}_f, t_f)$ in the forward dynamics under a particular pre-specified control protocol $\boldsymbol{\lambda}(\tau)$. The conjugate transition $(\mathbf{Q}_0^\dagger, \mathbf{P}_0^\dagger, \mathcal{E}_0^\dagger, t_0) \to (\mathbf{Q}_f^\dagger, \mathbf{P}_f^\dagger, \mathcal{E}_f^\dagger, t_f)$ (which is equivalent to $(\mathbf{Q}_f, -\mathbf{P}_f, \mathcal{E}_f, t_0) \to (\mathbf{Q}_0, -\mathbf{P}_0, \mathcal{E}_0, t_f)$) under the conjugate protocol $\boldsymbol{\lambda}^\dagger(\tau)$ has a transition probability $P^\dagger_{\boldsymbol{\lambda}^\dagger(\tau)}(\mathbf{Q}_f^\dagger, \mathbf{P}_f^\dagger, \mathcal{E}_f^\dagger, t_f ; \mathbf{Q}_0^\dagger, \mathbf{P}_0^\dagger, \mathcal{E}_0^\dagger, t_0)$ under the conjugate dynamics:

$$
\begin{aligned}
P^\dagger_{\boldsymbol{\lambda}^\dagger(\tau)}(\mathbf{Q}^\dagger_f, \mathbf{P}^\dagger_f, \mathcal{E}^\dagger_f, t_f \,;\, \mathbf{Q}^\dagger_0, \mathbf{P}^\dagger_0, \mathcal{E}^\dagger_0, t_0) &= P^\dagger_{\boldsymbol{\lambda}^\dagger(\tau)}(\mathbf{Q}_0, -\mathbf{P}_0, \mathcal{E}_0, t_f \,;\, \mathbf{Q}_f, -\mathbf{P}_f, \mathcal{E}_f, t_0) \\
&= \int \mathcal{D}_{\boldsymbol{\lambda}^\dagger(\tau)}[\mathbf{Q}^\dagger(\tau), \mathbf{P}^\dagger(\tau), \mathcal{E}^\dagger(\tau)] \\
&\quad \times P^\dagger_{\boldsymbol{\lambda}^\dagger(\tau)}[\mathbf{Q}^\dagger(\tau), \mathbf{P}^\dagger(\tau), \mathcal{E}^\dagger(\tau)] \\
&= \int \mathcal{D}_{\boldsymbol{\lambda}^\dagger(\tau)}[\mathbf{Q}^\dagger(\tau), \mathbf{P}^\dagger(\tau), \mathcal{E}^\dagger(\tau)] \\
&\quad \times e^{-\mathcal{A}^\dagger[\mathbf{Q}^\dagger(\tau), \mathbf{P}^\dagger(\tau)\mathcal{E}^\dagger(\tau), \boldsymbol{\lambda}^\dagger(\tau)]}.
\end{aligned}
\tag{4.111}
$$

Note that the conjugate action along conjugate trajectories can be written in terms of the forward trajectories:

$$
\begin{aligned}
\mathcal{A}^\dagger[\mathbf{Q}^\dagger(\tau), \mathbf{P}^\dagger(\tau), \mathcal{E}^\dagger(\tau), \boldsymbol{\lambda}^\dagger(\tau)] &= \int_{t_0}^{t_f} \mathrm{d}t\, \mathcal{L}^\dagger(\mathbf{Q}^\dagger(t), \mathbf{P}^\dagger(t), \mathcal{E}^\dagger(t), \boldsymbol{\lambda}^\dagger(t), \dot{\mathbf{P}}^\dagger(t)) \\
&= \int_{t_0}^{t_f} \mathrm{d}t\, \mathcal{L}^\dagger(\mathbf{Q}(t^\dagger), -\mathbf{P}(t^\dagger), \mathcal{E}(t^\dagger), \boldsymbol{\lambda}(t^\dagger), \dot{\mathbf{P}}(t^\dagger)) \\
&= -\int_{t_f}^{t_0} \mathrm{d}t^\dagger\, \mathcal{L}^\dagger(\mathbf{Q}(t^\dagger), -\mathbf{P}(t^\dagger), \mathcal{E}(t^\dagger), \boldsymbol{\lambda}(t^\dagger), \dot{\mathbf{P}}(t^\dagger)) \\
&= \int_{t_0}^{t_f} \mathrm{d}t\, \mathcal{L}^\dagger(\mathbf{Q}(t), -\mathbf{P}(t), \mathcal{E}(t), \boldsymbol{\lambda}(t), \dot{\mathbf{P}}(t)).
\end{aligned}
\tag{4.112}
$$

The difference between the conjugate action evaluated along an conjugate trajectory and the forward action along the corresponding forward trajectory can then be written as

$$
\mathcal{A}^\dagger[\mathbf{Q}^\dagger(\tau), \mathbf{P}^\dagger(\tau), \mathcal{E}^\dagger(\tau), \boldsymbol{\lambda}^\dagger(\tau)] - \mathcal{A}[\mathbf{Q}(\tau), \mathbf{P}(\tau), \mathcal{E}(\tau), \boldsymbol{\lambda}(\tau)] = \tag{4.113}
$$
$$
\int_{t_0}^{t_f} \mathrm{d}t\, \left[\mathcal{L}^\dagger(\mathbf{Q}(t), -\mathbf{P}(t), \mathcal{E}(t), \boldsymbol{\lambda}(t), \dot{\mathbf{P}}(t)) - \mathcal{L}(\mathbf{Q}(t), \mathbf{P}(t), \mathcal{E}(t), \boldsymbol{\lambda}(t), \dot{\mathbf{P}}(t)) \right].
$$

Note that $\mathcal{L}$ can be decomposed into parts either even or odd under conjugation:

$$\mathcal{L} = \mathcal{L}^{\mathrm{e}} + \mathcal{L}^{\mathrm{o}} \tag{4.114}$$

$$\mathcal{L}^{\mathrm{e}}(\mathbf{Q}(t), \mathbf{P}(t), \mathcal{E}(t), \boldsymbol{\lambda}(t)) = \mathcal{L}^{\mathrm{e}\,\dagger}(\mathbf{Q}(t), -\mathbf{P}(t), \mathcal{E}(t), \boldsymbol{\lambda}(t))$$

$$\mathcal{L}^{\mathrm{o}}(\mathbf{Q}(t), \mathbf{P}(t), \mathcal{E}(t), \boldsymbol{\lambda}(t)) = -\mathcal{L}^{\mathrm{o}\,\dagger}(\mathbf{Q}(t), -\mathbf{P}(t), \mathcal{E}(t), \boldsymbol{\lambda}(t)).$$

We thus have

$$\mathcal{A}^{\dagger} - \mathcal{A} = -2 \int_{t_0}^{t_f} \mathrm{d}t\, \mathcal{L}^{\mathrm{o}}(\mathbf{Q}(t), \mathbf{P}(t), \mathcal{E}(t), \boldsymbol{\lambda}(t)), \tag{4.115}$$

where

$$\begin{aligned}
\mathcal{L}^{\mathrm{o}}(\mathbf{Q}, \mathbf{P}, \mathcal{E}, \boldsymbol{\lambda}) = {}& -\frac{1}{2}\mathbf{P}^T M^{-1} \partial_E \mathbf{u} \tag{4.116} \\
& + \frac{1}{2}\mathbf{P}^T M^{-1} \left(\dot{\mathbf{P}} + \partial_{\mathbf{Q}} V - \mathbf{u} + \frac{1}{2\Sigma} \partial_E \left[\Sigma L^{\mathrm{ay}} \right] M^{-1} \mathbf{P} \right) \\
& \times \left. \partial_E \ln \Sigma \right|_{E = \mathcal{E} - H(\mathbf{Q}, \mathbf{P}, \boldsymbol{\lambda})}.
\end{aligned}$$

We note that the terms in the parenthesis in the above equation are precisely the terms in the mesoscopic force on the system of interest which break the microscopic symmetry relation $\dot{\mathbf{P}}^{\dagger}(t) = \dot{\mathbf{P}}(t^{\dagger})$. Making use of the expressions for $\mathrm{d}\mathbf{Q}$, $\mathrm{d}\mathbf{P}$, and $\mathrm{d}E$ in Eq. (4.28), we find

$$\begin{aligned}
\mathcal{A}^{\dagger} - \mathcal{A} = {}& \int_{t_0}^{t_f} \partial_E \ln[\Sigma(E, \mathbf{Q})]\mathrm{d}E + \frac{1}{\Sigma(E, \mathbf{Q})} \partial_E[\Sigma(E, \mathbf{Q})\mathbf{u}(E, \mathbf{Q})] \cdot \mathrm{d}\mathbf{Q} \\
= {}& \int_{t_0}^{t_f} \partial_E \ln[\Sigma(E, \mathbf{Q})]\mathrm{d}E + \partial_{\mathbf{Q}} \ln[\Sigma(E, \mathbf{Q})] \cdot \mathrm{d}\mathbf{Q} \\
= {}& \int_{t_0}^{t_f} \mathrm{d}\ln[\Sigma(E, \mathbf{Q})] \\
= {}& \ln[\Sigma(E_f, \mathbf{Q}_f)] - \ln[\Sigma(E_0, \mathbf{Q}_0)] \\
= {}& \frac{s(E_f, \mathbf{Q}_f) - s(E_0, \mathbf{Q}_0)}{k_B}, \tag{4.117}
\end{aligned}$$

where we have employed the identity from Eq. (3.17) and the definition of the microcanonical entropy s of the bath in (3.105), and the constraint $E = \mathcal{E} - H(\mathbf{X}, \boldsymbol{\lambda})$ is implied. We thus have

$$\Delta s[E(\tau), \mathbf{Q}(\tau)] = k_B \bigg(\mathcal{A}^\dagger [\mathbf{Q}^\dagger(\tau), \mathbf{P}^\dagger(\tau), \mathcal{E}^\dagger(\tau), \boldsymbol{\lambda}^\dagger(\tau)] - \mathcal{A}[\mathbf{Q}(\tau), \mathbf{P}(\tau), \mathcal{E}(\tau), \boldsymbol{\lambda}(\tau)] \bigg), \quad (4.118)$$

and

$$\frac{P^\dagger_{\boldsymbol{\lambda}^\dagger(\tau)}[\mathbf{Q}^\dagger(\tau), \mathbf{P}^\dagger(\tau), \mathcal{E}^\dagger(\tau)]}{P_{\boldsymbol{\lambda}(\tau)}[\mathbf{Q}(\tau), \mathbf{P}(\tau), \mathcal{E}(\tau)]} = e^{-\frac{\Delta s[E(\tau), \mathbf{Q}(\tau)]}{k_B}}, \quad (4.119)$$

where $\Delta s[E(\tau), \mathbf{Q}(\tau)]$ denotes the change in the microcaonical entropy of the bath in the forward dynamics when the position of system of interest and the bath energy follow the trajectories $\mathbf{Q}(\tau)$ and $E(\tau)$, respectively, with the constraint $E(t) = \mathcal{E}(t) - H(\mathbf{X}(t), \boldsymbol{\lambda}(t))$ implied. Equations (4.118) and (4.119) are the central results of Sec. 4.3, and Eq. (4.119) is the chaos bath analog of the stochastic thermodynamics microscopic reversibility condition in Eq. (4.71).

Equations (4.118) and (4.119) relate changes in the microcanonical entropy of the bath to stochastic action and transition probabilities, just as stochastic thermodynamics relates changes in the thermodynamic entropy of the environment to transition probabilities and stochastic action. In this sense, the microcanonical entropy of the bath is the chaos bath analog of thermodynamic entropy. In stochastic thermodynamics, the increment of heat flow to the environment at a temperature T is related to the change in thermodynamic entropy s of the environment by $-\mathrm{d}q = T\mathrm{d}s$ (in this thesis, heat flow is defined to be positive when heat energy flows out of the environment, hence the $-$ sign in front of $\mathrm{d}q$), so we will analogously define heat flow in terms of entropy in the chaos bath framework, where T is the bath's microcanonical temperature and s its microcanonical entropy:

$$
\begin{aligned}
- \, \mathrm{d}q \;\; &= \;\; T \mathrm{d}s & (4.120)\\[4pt]
&= \;\; T \partial_E s \, \mathrm{d}E + T \partial_{\mathbf{Q}} s \cdot \mathrm{d}\mathbf{Q}\\[4pt]
&= \;\; \mathrm{d}E + \mathbf{u} \cdot \mathrm{d}\mathbf{Q} + k_B T \, \partial_E \mathbf{u} \cdot \mathrm{d}\mathbf{Q}.
\end{aligned}
$$

Under this definition, heat flow is not equal to the change in energy of the bath. From Eq. (4.28) we see that the first two terms on the last line of Eq. (4.120) together give the increment of energy flow from the system to the bath due to deterministic friction, the noise induced drift, and the rapidly fluctuating force. These quantities together give the *irreversible* energy exchange between the bath and the system. The term $k_B T \, \partial_E \mathbf{u}$ in Eq. (4.120) is the first order change in $\mathbf{u}$ due to adding a single unit of $k_B T$ of energy to the bath, so this term will be negligible in the thermodynamic limit. Thus, for large baths, bath entropy and heat flow are related in a manner analogous to ordinary thermodynamics if heat flow is interpreted as irreversible energy flow.

The increment of work done on the system of interest, $\mathrm{d}w$, due a change $\mathrm{d}\boldsymbol{\lambda}$ in the control protocol is given by

$$
\mathrm{d}w = \partial_{\boldsymbol{\lambda}} V \cdot \mathrm{d}\boldsymbol{\lambda}. \tag{4.121}
$$

By the conservation of energy, we have

$$
\mathrm{d}E + \mathrm{d}H \;\; = \;\; \mathrm{d}w, \tag{4.122}
$$

so Eq. (4.120) can be written as

$$
\mathrm{d}H \;\; = \;\; \mathrm{d}w + \mathrm{d}q + \partial_{\mathbf{Q}} s \cdot \mathrm{d}\mathbf{Q}. \tag{4.123}
$$

We thus have the first law of thermodynamics at the level of individual trajectories for a chaos bath. The last term on the right hand side of Eq. (4.123) accounts for the interaction energy between the system of interest and the bath. This first law generalized straightforwardly for the case of multiple baths:

$$\mathrm{d}H \;\; = \;\; \mathrm{d}w + \sum_\alpha \left(\mathrm{d}q^\alpha + \partial_{\mathbf{Q}} s^\alpha \cdot \mathrm{d}\mathbf{Q} \right). \tag{4.124}$$

4.4 Open loop fluctuation theorems

In this section, we will use the results of Sec. 4.3 to derive various fluctuation theorems in the chaos bath framework. Fluctuation theorems in stochastic thermodynamics provide very general constraints on the manner in which physical systems can be controlled and manipulated in the presence of thermal environments, even if driven arbitrarily far from thermal equilibrium. We will derive analogous expressions using the mesoscale equations of motion for $(\mathbf{X}, \mathcal{E})$, and then use physically well-defined connection between the microscopic and mesoscopic dynamics to show that the fluctuation theorems, although derived and understood at the stochastic mesoscopic level, follow trivially from the deterministic microscopic dynamics and the one-to-one correspondence between forward and conjugate trajectories.

It must be noted that the fluctuation theorems considered here apply only to open loop control protocols, and not feedback control protocols. In stochastic thermodynamics, feedback control introduces conceptual and computational difficulties which result in modified fluctuation theorems involving various measures of mutual information between the system trajectory and feedback controller [13, 27, 24, 30]. Stochastic thermodynamics with feedback control provides a natural setting for discussions and generalizations of Maxwell's demon. We discuss some aspects feedback control in stochastic thermodynamics in chapter 5, where we use feedback to stabilize a diffusing Brownian particle. Feedback control and modified fluctuation theorems would be interesting to examine in the chaos bath framework; the first principles Hamiltonian connection between the micro and mesoscale may

potentially provide a more satisfying physical interpretation of the conceptually abstract mathematical mutual information quantities, but we save this topic for future work.

4.4.1 Derivation

Fluctuation theorems are derived by choosing distributions for initial forward and conjugate conditions and using the reversibility condition in Eq. (4.119) to integrate various quantities over all forward and conjugate transitions [13]. In order to preform such integrations, we will need to know how volume elements in the conjugate trajectory path space relate to volume elements in the forward trajectory path space. Note from Eqs. (4.79) and (4.80) that the volume element in the forward trajectory space is given by

$$
\begin{aligned}
\mathcal{D}_{\boldsymbol{\lambda}(\tau)}[\mathbf{X}(\tau), \mathcal{E}(\tau)] \;\;=\;\; \mathrm{Lim} \prod_{k=1}^{N-1} & \left[\frac{\mathrm{d}\mathbf{P}_k}{[2\pi\Delta t]^{\frac{N_s}{2}} \left| \sqrt{L^{\mathrm{sy}}(E_k^*, \mathbf{Q}_k^*)} \right|} \right. \\
& \times \mathrm{d}\mathbf{Q}_k \, \delta \left(\Delta\mathbf{Q}_k - M^{-1}\mathbf{P}_k^*\Delta t \right) \\
& \left. \times \mathrm{d}\mathcal{E}_k \, \delta \left(\Delta\mathcal{E}_k - \partial_{\boldsymbol{\lambda}_k^*} V(\mathbf{Q}_k^*, \boldsymbol{\lambda}_k^*) \cdot \frac{\Delta\boldsymbol{\lambda}_k}{\Delta t}\Delta t \right) \right].
\end{aligned} \tag{4.125}
$$

The volume element at the evaluated at corresponding conjugate trajectory in the conjugate dynamics is likewise given by

$$
\begin{aligned}
\mathcal{D}_{\boldsymbol{\lambda}^\dagger(\tau)}[\mathbf{X}^\dagger(\tau), \mathcal{E}^\dagger(\tau)] \;\;=\;\; \mathrm{Lim} \prod_{k=1}^{N-1} & \left[\frac{\mathrm{d}\mathbf{P}_k^\dagger}{[2\pi\Delta t]^{\frac{N_s}{2}} \left| \sqrt{L^{\mathrm{sy}}(E_k^{\dagger*}, \mathbf{Q}_k^{\dagger*})} \right|} \right. \\
& \times \mathrm{d}\mathbf{Q}_k^\dagger \, \delta \left(\Delta\mathbf{Q}_k^\dagger - M^{-1}\mathbf{P}_k^{\dagger*}\Delta t \right) \\
& \left. \times \mathrm{d}\mathcal{E}_k^\dagger \, \delta \left(\Delta\mathcal{E}_k^\dagger - \partial_{\boldsymbol{\lambda}_k^{\dagger*}} V(\mathbf{Q}_k^{\dagger*}, \boldsymbol{\lambda}_k^{\dagger*}) \cdot \frac{\Delta\boldsymbol{\lambda}_k^\dagger}{\Delta t}\Delta t \right) \right].
\end{aligned} \tag{4.126}
$$

Because we have chosen to discretize our path integrals using the symmetric Stratonovich convention, we have $t_k^\dagger = t_{N-k}$. This identity, together with Eqs. (4.84) and (4.85), then yields the following:

$$\mathbf{Q}_k^\dagger = \mathbf{Q}_{N-k} \tag{4.127}$$

$$\mathbf{Q}_k^{\dagger*} = \mathbf{Q}_{N-k+1}^*$$

$$\Delta\mathbf{Q}_k^{\dagger*} = -\Delta\mathbf{Q}_{N-k+1},$$

$$\mathcal{E}_k^\dagger = \mathcal{E}_{N-k} \tag{4.128}$$

$$\mathcal{E}_k^{\dagger*} = \mathcal{E}_{N-k+1}^*$$

$$\Delta\mathcal{E}_k^{\dagger*} = -\Delta\mathcal{E}_{N-k+1},$$

$$\boldsymbol{\lambda}_k^\dagger = \boldsymbol{\lambda}_{N-k} \tag{4.129}$$

$$\boldsymbol{\lambda}_k^{\dagger*} = \boldsymbol{\lambda}_{N-k+1}^*$$

$$\Delta\boldsymbol{\lambda}_k^{\dagger*} = -\Delta\boldsymbol{\lambda}_{N-k+1},$$

$$\mathbf{P}_k^\dagger = -\mathbf{P}_{N-k} \tag{4.130}$$

$$\mathbf{P}_k^{\dagger*} = -\mathbf{P}_{N-k+1}^*$$

$$\Delta\mathbf{P}_k^{\dagger*} = \Delta\mathbf{P}_{N-k+1}.$$

From the above relations, and the fact that the determinant of the Jacobian matrix corresponding the transformation $\mathbf{P} \to -\mathbf{P}$ is unity, we find

$$\mathcal{D}_{\boldsymbol{\lambda}(\tau)}[\mathbf{X}(\tau), \mathcal{E}(\tau)] = \mathcal{D}_{\boldsymbol{\lambda}^\dagger(\tau)}[\mathbf{X}^\dagger(\tau), \mathcal{E}^\dagger(\tau)]. \tag{4.131}$$

Suppose now that the initial energy of the universe is given by $\mathcal{E}_0$, and the initial condition of the system of interest are distributed according the stationary distribution

$P_s(\mathbf{Q}, \mathbf{P}; \mathcal{E}_0, \boldsymbol{\lambda}_0)$ in Eq. (3.118). Then, the probability to observe the trajectory $[\mathbf{Q}(\tau), \mathbf{P}(\tau), \mathcal{E}(\tau)]$ which accomplishes the transition $\mathbf{Q}_0, \mathbf{P}_0, \mathcal{E}_0, t_0 \to \mathbf{Q}_f, \mathbf{P}_f, \mathcal{E}_f, t_f$ under the forward dynamics, given the protocol $\boldsymbol{\lambda}(\tau)$, is given by the expression $P_{\boldsymbol{\lambda}(\tau)}[\mathbf{Q}(\tau), \mathbf{P}(\tau), \mathcal{E}(\tau)]P_s(\mathbf{Q}_0, \mathbf{P}_0; \mathcal{E}_0, \boldsymbol{\lambda}_0)$. Likewise, suppose the initial energy of the universe in the conjugate dynamics is given by $\mathcal{E}_f$, and that the initial conjugate conditions are distributed according to the conjugate stationary distribution $P_s^\dagger(\mathbf{Q}, \mathbf{P}; \mathcal{E}_f, \boldsymbol{\lambda}_f)$. The probability to observe the conjugate trajectory $[\mathbf{Q}^\dagger(\tau), \mathbf{P}^\dagger(\tau), \mathcal{E}^\dagger(\tau)]$ which accomplishes the conjugate transition $\mathbf{Q}_f, -\mathbf{P}_f, \mathcal{E}_f, t_0 \to \mathbf{Q}_0, -\mathbf{P}_0, \mathcal{E}_0, t_f$ under the conjugate dynamics, given the protocol $\boldsymbol{\lambda}^\dagger(\tau)$, is given by the expression $P_{\boldsymbol{\lambda}^\dagger(\tau)}^\dagger[\mathbf{Q}^\dagger(\tau), \mathbf{P}^\dagger(\tau), \mathcal{E}^\dagger(\tau)]P_s^\dagger(\mathbf{Q}_f, -\mathbf{P}_f; \mathcal{E}_f, \boldsymbol{\lambda}_f)$. From Eqs. (3.118), (4.107) and (4.119), we find

$$\frac{P_{\boldsymbol{\lambda}^\dagger(\tau)}^\dagger[\mathbf{Q}^\dagger(\tau), \mathbf{P}^\dagger(\tau), \mathcal{E}^\dagger(\tau)]P_s^\dagger(\mathbf{Q}_f, -\mathbf{P}_f; \mathcal{E}_f, \boldsymbol{\lambda}_f)}{P_{\boldsymbol{\lambda}(\tau)}[\mathbf{Q}(\tau), \mathbf{P}(\tau), \mathcal{E}(\tau)]P_s(\mathbf{Q}_0, \mathbf{P}_0; \mathcal{E}_0, \boldsymbol{\lambda}_0)} = e^{-\frac{\Delta S^U[\mathcal{E}(\tau), \boldsymbol{\lambda}(\tau)]}{k_B}} \qquad (4.132)$$

Equation (4.132) expresses a reversibility condition when the initial forward and conjugate states are drawn from the forward and conjugate stationary distributions, respectively.

Making use of reversibility condition in Eq. (4.132) and the volume element relation in Eq. (4.131), as well as the fact that the magnitude of determinant of the Jacobian matrix for the coordinate change $\mathbf{P} \to -\mathbf{P}$ is unity, we have the following equality:

$$\mathcal{D}_{\boldsymbol{\lambda}^\dagger(\tau)}[\mathbf{X}^\dagger(\tau), \mathcal{E}^\dagger(\tau)]d\mathbf{X}_0^\dagger d\mathbf{X}_f^\dagger P_{\boldsymbol{\lambda}^\dagger(\tau)}^\dagger[\mathbf{X}^\dagger(\tau), \mathcal{E}^\dagger(\tau)]P_s^\dagger(\mathbf{X}_0^\dagger; \mathcal{E}_f, \boldsymbol{\lambda}_f) =$$

$$\mathcal{D}_{\boldsymbol{\lambda}(\tau)}[\mathbf{X}(\tau), \mathcal{E}(\tau)]d\mathbf{X}_0 d\mathbf{X}_f P_{\boldsymbol{\lambda}(\tau)}[\mathbf{X}(\tau), \mathcal{E}(\tau)]P_s(\mathbf{X}_0; \mathcal{E}_0, \boldsymbol{\lambda}_0)e^{-\frac{\Delta S^U[\mathcal{E}(\tau), \boldsymbol{\lambda}(\tau)]}{k_B}}. \qquad (4.133)$$

Equation (4.133) will serve as a master equality from which we can derive fluctuation theorems. Integrating both sides and rearranging, we find a statement of reversibility in terms of energy transitions:

$$\frac{P_{\boldsymbol{\lambda}^\dagger(\tau)}^\dagger(\mathcal{E}_0, t_f; \mathcal{E}_f, t_0)}{P_{\boldsymbol{\lambda}(\tau)}(\mathcal{E}_f, t_f; \mathcal{E}_0, t_0)} = e^{-\frac{\Delta S^U(\mathcal{E}_f, \boldsymbol{\lambda}_f; \mathcal{E}_0, \boldsymbol{\lambda}_0)}{k_B}}, \qquad (4.134)$$

where $P_{\boldsymbol{\lambda}(\tau)}(\mathcal{E}_f, t_f; \mathcal{E}_0, t_0)$ denotes the probability to observe the energy transition $\mathcal{E}_0 \to \mathcal{E}_f$ in the forward dynamics under the forward control protocol, given the initial energy $\mathcal{E}_0$ and initial conditions drawn from the stationary state, and $P^{\dagger}_{\boldsymbol{\lambda}^{\dagger}(\tau)}(\mathcal{E}_0, t_f; \mathcal{E}_f, t_0)$ denotes the probability to observe the energy transition $\mathcal{E}_f \to \mathcal{E}_0$ in the conjugate dynamics under the conjugate control protocol, given the initial energy $\mathcal{E}_f$ and initial conditions drawn from the stationary state. Notice that the change in the entropy of the universe in the exponential in Eq. (4.134) is considered a function of the initial and final energies and protocol values, as opposed to a functional of $\mathcal{E}(\tau)$ and $\boldsymbol{\lambda}(\tau)$. If we multiply both sides of Eq. (4.134) by $\delta(\mathcal{E}_f - \mathcal{E}_0 - w)P_{\boldsymbol{\lambda}(\tau)}(\mathcal{E}_f, t_f; \mathcal{E}_0, t_0)$, integrate over all $\mathcal{E}_f$, and rearrange, we have

$$\frac{P^{\dagger}_{\boldsymbol{\lambda}^{\dagger}(\tau)}(\mathcal{E}_0, t_f; \mathcal{E}_0 + w, t_0)}{P_{\boldsymbol{\lambda}(\tau)}(\mathcal{E}_0 + w, t_f; \mathcal{E}_0, t_0)} = \left\langle e^{-\frac{\Delta S^U}{k_B}} \,\middle|\, w \right\rangle_{\mathcal{E}_0, \boldsymbol{\lambda}(\tau)}. \tag{4.135}$$

Equation (4.135) is the microcanonical version of Crooks's fluctuation theorem for work [17, 92]. The quantity $P_{\boldsymbol{\lambda}(\tau)}(\mathcal{E}_0 + w, t_f; \mathcal{E}_0, t_0)$ represents the probability for the external agent to do work w in the forward dynamics, under the protocol $\boldsymbol{\lambda}(t)$, when the initial energy of the universe is $\mathcal{E}_0$, while the quantity $P^{\dagger}_{\boldsymbol{\lambda}^{\dagger}(\tau)}(\mathcal{E}_0, t_f; \mathcal{E}_0 + w, t_0)$ represents the probability for the external agent to do work $-w$ in the conjugate dynamics, under the conjugate protocol $\boldsymbol{\lambda}^{\dagger}(t)$, when the initial energy of the universe is $\mathcal{E}_0 + w$, assuming initial stationary forward and conjugate distributions. The expectation on the right hand side of Eq.(4.135) is an expectation over all realizations of the forward dynamics in which the amount work done under the protocol $\boldsymbol{\lambda}(\tau)$ is equal to w, given an initial stationary distribution and initial energy of the universe $\mathcal{E}_0$. Multiplying both sides of Eq. (4.135) by $P_{\boldsymbol{\lambda}(\tau)}(\mathcal{E}_0 + w, t_f; \mathcal{E}_0, t_0)$ and integrating over all possible work values yields

$$\int \mathrm{d}w \, P^{\dagger}_{\boldsymbol{\lambda}^{\dagger}(\tau)}(\mathcal{E}_0, t_f; \mathcal{E}_0 + w, t_0) = \left\langle e^{-\frac{\Delta S^U}{k_B}} \right\rangle_{\mathcal{E}_0, \boldsymbol{\lambda}(\tau)}. \tag{4.136}$$

Equation (4.136) is the microcanonical version of the integral fluctuation theorem [13, 27, 92]. The average on the right side of the equality is an average over all realizations of the

142

forward dynamics under the protocol $\boldsymbol{\lambda}(\tau)$, given an initial stationary distribution and initial energy of the universe $\mathcal{E}_0$. Note that the integral in Eq. (4.136) does not reduce to unity, so we are being a bit casual with our language in calling the equation an integral fluctuation theorem. This integral is related to the following thought experiment: suppose we are given an ensemble of systems, each prepared in the stationary distribution corresponding to $\boldsymbol{\lambda}_f$, for all possible energies of the universe, and we evolve each ensemble member using the conjugate dynamics under the conjugate protocol. For each ensemble member, there will be some probability to find the universe with energy $\mathcal{E}_0$ after implementing the control protocol, and the integral in Eq. (4.136) represents the sum of all such probabilities. This integral can take values anywhere in the interval $[0, \infty)$. If the transition probability in the integral were to for some reason become independent of $\mathcal{E}_0$, then $P^{\dagger}_{\boldsymbol{\lambda}^{\dagger}(\tau)}(\mathcal{E}_0, t_f; \mathcal{E}_0 + w, t_0)$ would simply represent the probability to do work $-w$ in the conjugate dynamics under the protocol $\boldsymbol{\lambda}^{\dagger}(\tau)$. The integral in Eq. (4.136) would then reduce to unity, and we would have a genuine integral fluctuation theorem. As we will discuss in the next subsection, such an independence is expected to arise in the thermodynamic limit.

We close this subsection by deriving the microcanonical version of Crooks's fluctuation theorem for entropy production [17]. Note that if the change in the total entropy of the universe along a trajectory depends only on the initial and final values of the energy of the universe and control parameter, so we have the equality $\Delta S^U[\mathcal{E}(\tau), \boldsymbol{\lambda}(\tau)] = -\Delta S^U[\mathcal{E}^{\dagger}(\tau), \boldsymbol{\lambda}^{\dagger}(\tau)]$. Using this fact, we can multiply both sides of Eq. (4.133) by $\delta(\omega - \Delta S^U[\mathcal{E}(\tau), \boldsymbol{\lambda}(\tau)])$, and then integrate and rearrange to find

$$\frac{P^{\dagger}_{\boldsymbol{\lambda}^{\dagger}(\tau)}(\Delta S^U = -\omega | \mathcal{E}_0, t_f; \mathcal{E}_f, t_0)}{P_{\boldsymbol{\lambda}(\tau)}(\Delta S^U = \omega | \mathcal{E}_f, t_f; \mathcal{E}_0, t_0)} = e^{-\frac{\omega}{k_B}}. \tag{4.137}$$

The quantity $P_{\boldsymbol{\lambda}(\tau)}(\Delta S^U = \omega | \mathcal{E}_f, t_f; \mathcal{E}_0, t_0)$ is the probability for the entropy of the universe to increase by ω in the forward dynamics under the forward protocol, given the energy transition $\mathcal{E}_0 \to \mathcal{E}_f$ and initial conditions drawn from the stationary distribution, while $P^{\dagger}_{\boldsymbol{\lambda}^{\dagger}(\tau)}(\Delta S^U = -\omega | \mathcal{E}_0, t_f; \mathcal{E}_f, t_0)$ is the probability for the entropy of the universe to increase by $-\omega$ in the conjugate dynamics under the conjugate protocol, given the energy transition

$\mathcal{E}_f \to \mathcal{E}_0$ and initial conditions drawn from the stationary distribution.

4.4.2 The thermodynamic limit and the second law

We now consider our fluctuation theorems in the thermodynamic limit as defined in Sec. 3.4.5. The number of degrees of freedom of the bath will be denoted by N_B, with the number of degrees of freedom of the system of interest given by N_S. Throughout this subsection, we assume that the work done on the universe scales like the energy of the system, so that we have $w \ll \mathcal{E}_0$ in the thermodynamic limit. Physically, this is a reasonable assumption because work is done on the universe only through changes in the system of interest's potential energy, and the energy of the system of interest is assumed to scale like N_S, while the energy of the universe scales like N_B. We will make use of Eq. (3.146), which shows that in the thermodynamic limit, the microcanonical temperature of the universe is equal to the microcanonical temperature of the bath, and both temperatures become independent of $\boldsymbol{\lambda}$. We will therefore denote both temperatures by the function $T(\mathcal{E})$.

In the thermodynamic limit, the total energy of the universe approaches infinity and is effectively constant. The quantity relavent to the system of interest is the average energy per degree of freedom of the bath (i.e. the bath's temperature), and this quantity is also effectively constant in the thermodynamic limit. Therefore, in the thermodynamic limit, we expect the transition probabilities in Eqs. (4.135) and (4.137) to be independent of the energy of the universe. This argument is used by Cleuren et al. in Ref. [92], where a fluctuation theorem is derived from microcanonical considerations. Under this assumption, the entropy production fluctuation theorem in Eq. (4.137) can be written as

$$\frac{P^{\dagger}_{\boldsymbol{\lambda}^{\dagger}(\tau)}(\Delta S^U = -\omega)}{P_{\boldsymbol{\lambda}(\tau)}(\Delta S^U = \omega)} = e^{-\frac{\omega}{k_B}}, \tag{4.138}$$

where the initial and final times are implied. Equation (4.138) is Crooks's fluctuation theorem for entropy production as it originally appeared in [17]. Multiplying both sides of the above equation by $P_{\boldsymbol{\lambda}(\tau)}(\Delta S^U = \omega) = e^{-\frac{\omega}{k_B}}$ and integrating over all possible ω gives an integral fluctuation theorem:

$$1 = \left\langle e^{-\frac{\omega}{k_B}} \right\rangle_{\boldsymbol{\lambda}(\tau)}. \tag{4.139}$$

Applying Jensen's inequality [84], the integral fluctuation theorem above implies the second law of thermodynamics:

$$0 \leq \langle \omega \rangle_{\boldsymbol{\lambda}(\tau)}. \tag{4.140}$$

Consider now the fluctuation theorem for work Eq. (4.135). Assuming that the transition probabilities become independent of $\mathcal{E}_0$ in the thermodynamic limit, we have

$$\frac{P^{\dagger}_{\boldsymbol{\lambda}(\tau)}(-w)}{P_{\boldsymbol{\lambda}(\tau)}(w)} = \left\langle e^{-\frac{\Delta S^U}{k_B}} \,\middle|\, w \right\rangle_{\boldsymbol{\lambda}(\tau)}. \tag{4.141}$$

Note that the change in the entropy of the universe can be written as

$$
\begin{aligned}
-\Delta S^U \;&=\; -S^U(\mathcal{E}_f, \boldsymbol{\lambda}_f) + S^U(\mathcal{E}_0, \boldsymbol{\lambda}_0) \\
&=\; -S^U(\mathcal{E}_0, \boldsymbol{\lambda}_f) + S^U(\mathcal{E}_0, \boldsymbol{\lambda}_0) - \partial_{\mathcal{E}_0} S^U(\mathcal{E}_0, \boldsymbol{\lambda}_f)(\mathcal{E}_f - \mathcal{E}_0) \\
&\quad\; -\frac{1}{2} \partial^2_{\mathcal{E}_0 \mathcal{E}_0} S^U(\mathcal{E}_0, \boldsymbol{\lambda}_f)(\mathcal{E}_f - \mathcal{E}_0)^2 + ... \\
&=\; -S^U(\mathcal{E}_0, \boldsymbol{\lambda}_f) + S^U(\mathcal{E}_0, \boldsymbol{\lambda}_0) - \frac{w}{T^U(\mathcal{E}_0, \boldsymbol{\lambda}_f)} + O\left(\frac{N_S}{N_B}\right) \\
&\underset{\text{TDlimit}}{=}\; -\Delta_{\boldsymbol{\lambda}} S^U(\mathcal{E}_0, \boldsymbol{\lambda}_f, \boldsymbol{\lambda}_0) - \frac{w}{T(\mathcal{E}_0)} \\
&=\; \frac{\Delta_{\boldsymbol{\lambda}} \mathcal{F} - w}{T},
\end{aligned} \tag{4.142}
$$

where we use the definition of free energy in Eq. (3.144), and where $\Delta_{\boldsymbol{\lambda}}...$ denotes the change in the quantity ... due to changes in $\boldsymbol{\lambda}$. The dependence on $\mathcal{E}_0$, which is a fixed parameter in the thermodynamic limit, as been suppressed on the last line of Eq. (4.142). From Eq. (4.141), we thus have

$$\frac{P^\dagger_{\boldsymbol{\lambda}(\tau)}(-w)}{P_{\boldsymbol{\lambda}(\tau)}(w)} = e^{\frac{\Delta_{\boldsymbol{\lambda}}\mathcal{F}}{k_B T}} \left\langle e^{-\frac{w}{k_B T}} \right\rangle_{\boldsymbol{\lambda}(\tau)}. \tag{4.143}$$

The above equation is Crooks's fluctuation theorem for work as it appears in [17]. Multiplying by $P_{\boldsymbol{\lambda}(\tau)}(w)$ and integrating of all possible work values yields Jarzynski's non-equilibrium work relation for free energy differences [18]:

$$e^{-\frac{\Delta_{\boldsymbol{\lambda}}\mathcal{F}}{k_B T}} = \left\langle e^{-\frac{w}{k_B T}} \right\rangle_{\boldsymbol{\lambda}(\tau)}. \tag{4.144}$$

Applying Jensen's inequality to Eq. (4.144) gives another statement of the second law of thermodynamics:

$$\Delta_{\boldsymbol{\lambda}}\mathcal{F} \leq \langle w \rangle_{\boldsymbol{\lambda}(\tau)}. \tag{4.145}$$

4.4.3 Hamiltonian derivation of fluctuation theorems

The fluctuation theorems presented in the previous two sections were all derived starting from the reversibility condition in Eq. (4.119). We arrived at this reversibility condition by considering differences in the stochastic action associated with forward and conjugate trajectories, and we noted that terms contributing to this difference originated from the portions of the stochastic mesoscopic force on the system of interest which break the deterministic microscopic symmetry relation $\dot{\mathbf{P}}^\dagger(t) = \dot{\mathbf{P}}(t^\dagger)$. Because the fluctuation theorems, and by extension the second law of thermodynamics, appear to be intimately tied to our approximate stochastic framework at the mesoscale, it is not clear whether or not they are valid descriptions of nature at the deterministic microscale. Are fluctuation theorems and the second law of thermodynamics simply consequences of our imperfect mesoscopic description of the universe, or do they bear deeper connections to the underlying microscopic dynamics governed by Hamilton's equations? In this section, we will argue for the

latter case by deriving Eq. (4.119) without appealing to stochastic dynamics or mesoscale approximations.

To derive Eq. (4.119), we must remind ourselves of a two simple but important points regarding the work in chapters 3 and 4. First, as shown explicitly in Sec. 4.3.3, for every Hamiltonian of the universe $\mathcal{H}$ subject to some control protocol $\boldsymbol{\lambda}(\tau)$, there exists a conjugate Hamiltonian $\mathcal{H}^\dagger$ subject to the conjugate protocol $\boldsymbol{\lambda}^\dagger(\tau)$ such that, for each system of interest plus bath forward trajectory generated by $\mathcal{H}$, there is a corresponding unique conjugate trajectory generated by $\mathcal{H}^\dagger$ which traces the forward trajectory in reverse. Second, as shown in Sec. (3.4.3), the mesoscopic stationary distribution $P_s(\mathbf{X}; \mathcal{E}, \boldsymbol{\lambda})$ can be derived by assuming an underlying microscopic microcanonical distribution for the entire universe, without making any reference to the Fokker-Planck equation (3.75). Likewise, the conjugate mesoscopic stationary distribution $P_s^\dagger$, defined in Eq. (4.107), follows from an underlying microscopic microcanonical distribution in the conjugate dynamics. In both the forward and conjugate dynamics, a microscopic microcanonical distribution follows from the assumption of microscopic chaos.

Consider now the microcanonical partition function of the universe $\Sigma^U(\mathcal{E}_0, \boldsymbol{\lambda}_0)$. This quantity gives a measure of the total number of system of interest plus bath states on the energy shell $\mathcal{H}(\mathbf{X}, \mathbf{x}, \boldsymbol{\lambda}_0) = \mathcal{E}_0$ in the full phase space, and the stationary distribution $P_s(\mathbf{X}_0; \mathcal{E}_0, \boldsymbol{\lambda}_0)$ measures the fraction of these states for which the system of interest is located at $\mathbf{X}_0$. The quantity $\Sigma^U(\mathcal{E}_0, \boldsymbol{\lambda}_0) P_s(\mathbf{X}_0; \mathcal{E}_0, \boldsymbol{\lambda}_0)$ thus gives a measure of the total number of states in the full phase space with the system of interest located at $\mathbf{X}_0$ when the energy of the universe is $\mathcal{E}_0$. We now suppose that the control parameter evolves under a particular protocol $\boldsymbol{\lambda}(\tau)$, with initial and final values $\boldsymbol{\lambda}_0$ and $\boldsymbol{\lambda}_f$, respectively, over the time interval $[t_0, t_f]$, while the system of interest and bath evolve according to Hamilton's equations under the Hamiltonian $\mathcal{H}(\mathbf{X}, \mathbf{x}, \boldsymbol{\lambda}(t))$. Of all initial states of the universe counted by the measure $\Sigma^U(\mathcal{E}_0, \boldsymbol{\lambda}_0) P_s(\mathbf{X}_0; \mathcal{E}_0, \boldsymbol{\lambda}_0)$, only a fraction will result in the system of interest and energy of the universe following the particular trajectories $\mathbf{X}(\tau)$ and $\mathcal{E}(\tau)$, respectively, with final values $\mathbf{X}_f$ and $\mathcal{E}_f$. We denote this fraction by $P_{\boldsymbol{\lambda}(\tau)}[\mathbf{X}(\tau), \mathcal{E}(\tau)]$. Thus, the measure of the number of states in the full phase which realize the transition $(\mathbf{X}_0, \mathcal{E}_0) \to (\mathbf{X}_f, \mathcal{E}_f)$ along the trajectory $(\mathbf{X}(\tau), \mathcal{E}(\tau))$ is given by $\Sigma^U(\mathcal{E}_0, \boldsymbol{\lambda}_0) P_s(\mathbf{X}_0; \mathcal{E}_0, \boldsymbol{\lambda}_0) P_{\boldsymbol{\lambda}(\tau)}[\mathbf{X}(\tau), \mathcal{E}(\tau)]$.

Consider now the conjugate Hamiltonian under the conjugate protocol $\boldsymbol{\lambda}^\dagger(\tau)$. By similar arguments, the quantity $\Sigma^{U\,\dagger}(\mathcal{E}_f, \boldsymbol{\lambda}_f) P_s^\dagger(\mathbf{X}_f^\dagger; \mathcal{E}_f, \boldsymbol{\lambda}_f) P_{\boldsymbol{\lambda}^\dagger(\tau)}[\mathbf{X}^\dagger(\tau), \mathcal{E}^\dagger(\tau)]$ is the measure of the number of states in the full phase space which begin on the energy shell $\mathcal{E}_f$ with the system interest initially located at $\mathbf{X}_f^\dagger$, evolve according to the conjugate Hamiltonian under the conjugate control protocol $\boldsymbol{\lambda}^\dagger(\tau)$ over the interval $[t_0, t_f]$ with the system of interest and energy of the universe following the trajectories $\mathbf{X}^\dagger(\tau)$ and $\mathcal{E}^\dagger(\tau)$, respectively, with final values $\mathbf{X}_0^\dagger$ and $\mathcal{E}_0$. As noted, in the full system plus bath phase space, for every forward trajectory which evolves under the forward dynamics, there is exactly one conjugate trajectory which makes the reverse transition under the conjugate dynamics. Therefore, the measure of states which realize the transition $(\mathbf{X}_0, \mathcal{E}_0) \to (\mathbf{X}_f, \mathcal{E}_f)$ along the trajectory $(\mathbf{X}(\tau), \mathcal{E}(\tau))$ under the forward dynamics must be equal to the measure of states which accomplish the conjugate transition under the conjugate dynamics, so we have the somewhat trivial equality

$$\Sigma^{U\,\dagger}(\mathcal{E}_f, \boldsymbol{\lambda}_f) P_s^\dagger(\mathbf{X}_f^\dagger; \mathcal{E}_f, \boldsymbol{\lambda}_f) P_{\boldsymbol{\lambda}^\dagger(\tau)}[\mathbf{X}^\dagger(\tau), \mathcal{E}^\dagger(\tau)] \quad = \quad \text{(4.146)}$$
$$\Sigma^{U}(\mathcal{E}_0, \boldsymbol{\lambda}_0) P_s(\mathbf{X}_0; \mathcal{E}_0, \boldsymbol{\lambda}_0) P_{\boldsymbol{\lambda}(\tau)}^\dagger[\mathbf{X}(\tau), \mathcal{E}(\tau)].$$

Using the definitions Eqs. (3.104), (3.105), (3.118), (4.96), (4.97), and (4.107), the above equality can be rearranged to yield

$$\frac{P_{\boldsymbol{\lambda}^\dagger(\tau)}^\dagger[\mathbf{X}^\dagger(\tau), \mathcal{E}^\dagger(\tau)]}{P_{\boldsymbol{\lambda}(\tau)}[\mathbf{X}(\tau), \mathcal{E}(\tau)]} = e^{-\frac{1}{k_B}\left[s(\mathcal{E}_f - H(\mathbf{X}_f, \boldsymbol{\lambda}_f), \mathbf{Q}_f) - s(\mathcal{E}_i - H(\mathbf{X}_i, \boldsymbol{\lambda}_i), \mathbf{Q}_i)\right]}. \quad \text{(4.147)}$$

We remind the reader here that s denotes the microcanonical entropy of the bath.

Equation (4.147) is precisely the reversibility condition in Eq. (4.119) which was derived from the mesoscopic stochastic perspective. We thus conclude that the reversibility condition, as well as the fluctuation theorems which derive from it, follow directly from the assumption of microscopic chaos and the one to one correspondence between microscopic forward and conjugate trajectories. This is the central result of chapter 4. Note

that throughout chapters 3 and 4, the bath dynamics, for a fixed $\mathbf{Q}$, are left arbitrary, so no explicit time-reversal symmetry or anti-canonical symmetry has been assumed for the Hamiltonian of the universe. We have, on the other hand, assumed that the interaction between the system of interest and bath depends only on $\mathbf{Q}$, and the form assumed for the system of interest's Hamiltonian in inherently time-reversal symmetric. The arguments leading to Eq. (4.147), however, apply equally well to arbitrary Hamiltonians with both position and momentum dependent interactions with the bath, so this reversibility condition, and thus fluctuation theorems, will be valid for arbitrary Hamiltonian systems which exhibit microscopic chaos. We also note that the arguments which led to Eq. (4.147) will apply in the presence of multiple baths and yield an analogous reversibility condition:

$$\frac{P^{\dagger}_{\boldsymbol{\lambda}^{\dagger}(\tau)}[\mathbf{X}^{\dagger}(\tau), \mathbf{E}^{\dagger}(\tau)]}{P_{\boldsymbol{\lambda}(\tau)}[\mathbf{X}(\tau), \mathbf{E}(\tau)]} = e^{-\frac{1}{k_B}\sum_{\alpha}\left[s^{\alpha}(E^{\alpha}_f, \mathbf{Q}_f) - s^{\alpha}(E^{\alpha}_i, \mathbf{Q}_i)\right]}. \tag{4.148}$$

Equations (4.147) and (4.148) follow even without any explicit time-scale separation assumptions, although without such a separation, parametrizing the bath Hamiltonians by the state of the system of interest is unlikely to be a useful way to describe the system-baths interaction.

4.5   Summary and conclusions

In this chapter, we have used the Fokker-Planck equation derived in Chap. 3 to deduce a stochastic Langevin equation for the mesoscale evolution of the system of interest and baths. This Langevin equation facilitated a discussion of the mesoscopic forces exerted on the system of interest by the baths, and we showed that these forces manifest in a first law of thermodynamics for the system of interest at the level of individual trajectories. For the case of a single bath, we employed the stochastic Langevin approach to derive fluctuation theorems, and we showed that these fluctuation theorems contain the essence of the second law of thermodynamics - they imply constraints on work extraction and entropy production when the system of interest's Hamiltonian is manipulated through a control protocol, and

they yield various statements of the second law in the thermodynamic limit. We emphasize that these fluctuation theorems are valid even for low-dimensional baths, so the results of Chaps. 3 and 4 taken together show that the chaos bath framework gives a first principles Hamiltonian derivation of the laws of thermodynamics which is independent of any laws of large numbers, both in and out of equilibrium.

For future study, it would be interesting to allow for arbitrary Hamiltonians in the derivation of the Fokker-Planck equation, and to allow for multiple baths in the derivation of the path integral formalism. These generalizations will most certainly complicate the intermediate calculations, but from the micro-meso connection established in Sec. 4.4.3, we know that the end result of such labors will be the Eq. (4.148), with s^{α} a dependent on $\mathbf{P}$ as well . This point highlights the overreaching theme of this thesis - by examining the foundations of stochastic processes in physics, we learn more about the processes themselves as well as the underlying processes from which they originate. We began with a deterministic Hamiltonian system, and we learned that this system exhibits thermodynamic behavior by studying its stochastic counterpart. Then, by re-examining the system from the deterministic perspective, we showed that the stochastic results, namely the reversibility condition in Eq. (4.119), hold under more general conditions than initially considered and follow from simple and generic properties of the deterministic equations of motion. However, in order to derive the results from the deterministic perspective, we had to use Eq. (4.146), an equality which follows so trivially from the one-to-one correspondence between forward and conjugate trajectories that we would have been unlikely to write it down and examine its consequences without first being motivated by results from the stochastic perspective. In future work, we hope to employ similar micro-meso connections to achieve a deeper understanding of information processing and feedback control in physical systems.

Chapter 5

Stabilizing Mesoscale Thermodynamic Systems with Feedback Control

5.1 Introduction

In control systems, there are many notions of stability, but they generally refer to a systems ability to maintain a specific trajectory or remain at a fixed point when subject to perturbations [93, 94]. Maneuverability quantifies a system's response to control inputs - a system which deviates from a particular trajectory or fixed point for relatively little control effort is considered to be highly maneuverable. Thus, by definition, there exists an inherent trade-off between stability and maneuverability. A clear example of this trade-off is found in aviation: training aircraft are designed to be is aerodynamically stable, while an F-16 fighter jet is designed to be aerodynamically unstable [95]. The unstable fighter jet would crash to the ground were it not for the on-board electronic feedback control system (referred to as a "fly-by-wire" system) which keeps the aircraft in the air by continually measuring the aircraft's state and making corresponding changes to control surfaces on the aircraft's body [95]. As a result, the F-16 is inherently more nimble than the training aircraft, and it can preform advanced combat maneuvers which would be otherwise impossible in an aerodynamically stable system. Clearly, there are advantages to designing inherently

unstable systems maintained by feedback, so we ask if it is feasible for nature to employ such designs in molecular machines and other mesoscale biological systems which function in highly dissipative environments. This work is a preliminary exploration of this question.

In this chapter, we study, by means of an analytical tractable example, the inherent energetic costs and inherent limitations associated with stabilizing mesoscale thermodynamic systems with feedback control. Consider a Brownian particle in one dimension. If the particle is subject to forces from the surrounding thermal environment only, it will exhibit Brownian motion and diffuse away from any given point [1]. Suppose that we wish to counteract this diffusive motion and stabilize the position of this particle about the origin by implementing a feedback control scheme. Our feedback scheme can be conceptualized as a neat-fingered intelligent being, referred to in this chapter as a "control demon," which continuously measures the position of the particle and, based on the outcomes of the measurements, exerts corresponding forces on the particle. This control demon represents a control theoretic take on Maxwell's demon [21, 22, 23]. Roughly speaking, if the control demon observes the particle fluctuate away from the origin in the positive direction, then it will exert a force on the particle in the negative direction (and vice-versa if the particle fluctuates away from the origin in the negative direction). The demon could in principle choose any arbitrarily complicated feedback force law to apply to the particle, but in order to keep the ensuing mathematics tractable, in this chapter, we consider only linear Markov controls [9]. In other words, the demon considered in this paper will stabilize by attempting, to the best of it's ability, to mimic the action of an idealized spring attached to the particle. Does a control demon operating in such a manner enact apparent violations the second law of thermodynamics analogously to a Maxwell demon? As we will show, the answer to this question is a bit subtle, even for the simple set-up considered here.

The subject of this chapter can be thought of as an overdamped analog of feedback cooling [26, 96, 97, 98] . Whereas a classical Brownian particle is feedback cooled by an applied momentum-dependent feedback force, for stabilization, we apply a position-dependent feedback force. The goal of feedback cooling is to reduce the variance of the thermal momentum distribution, while the goal of stabilization is to reduce the variance of the position distribution (which would be infinite for an unstable unbounded particle). We note that

results of feedback cooling have interesting interpretations in terms of information theory, and although we do not pursue such an approach in this chapter, analogous techniques could be applied to the study of stabilization [26]. In our work, we opt for a mechanical approach.

The purpose of this chapter is to determine the average rate at which the control demon must do work on the particle in order to achieve stabilization. This endeavor will yield a seemingly paradoxical result: the power required for stabilization preformed by a demon making perfect observations does not converge to the power required for stabilization preformed by a demon making imperfect observations in the limit of vanishingly small observation error. The resolution of this apparent paradox leads to the central result of this chapter; in order for the demon to stabilize against diffusive motion for free, without doing work or dissipating energy to the environment, it must operate at time-scales smaller than the time-scales over which diffusion is a valid description of the particle's motion. Our work towards this result will proceed in steps of increasing complexity. We begin in Sec. 5.2 by considering a simple uncontrolled autonomous system consisting of a spring attached to a Brownian particle, and we determine the average rate at which the spring does work on the particle while subject to thermal fluctuations. Then, in Sec. 5.3, we compare the autonomous system to a non-autonomous system consisting of a free Brownian particle under the influence of a control demon making perfect observations. In Sec. 5.4, we consider the more realistic case of control preformed by a demon making imperfect observations, and we find our apparently paradoxical result. We resolve the apparent paradox in Sec. 5.5 by considering the effect of short-lived correlations in the fluctuating thermal force acting on the particle, and we discuss physical implications and conclude in Sec. 5.6.

5.2   No control

Consider a one-dimensional system consisting of a particle in a thermal environment attached to an idealized spring. We will model the system dynamics by an overdamped Langevin equation [13]:

$$b\dot{x}(t) = -kx(t) + \sqrt{2bk_BT}\xi(t). \tag{5.1}$$

Here, we denote the position of the particle at time t by $x(t)$, and the overdot notation denotes a time derivative. The constant b is the damping coefficient, k is the spring constant, k_B is Boltzmann's constant, and T is the temperature of the environment. In this chapter, we assume the Stratonovich interpretation of all stochastic differential equations. The quantity $\xi(t)$ represents a Gaussian white noise process with mean and autocorrelation given by

$$\begin{aligned} \langle \xi(t) \rangle &= 0, \\ \langle \xi(t)\xi(t') \rangle &= \delta(t - t'), \end{aligned} \tag{5.2}$$

where $\langle A \rangle$ denotes an average of the quantity A (some functional of the particle's trajectory) over all realizations of the stochastic noise and initial particle conditions. The solution to Eq. (5.1) can be written in closed form, and is given by

$$x(t) = x_0 e^{-\frac{t}{\tau}} + \int_0^t dt'\, e^{-\frac{(t-t')}{\tau}} \sqrt{\frac{2k_BT}{b}}\xi(t'), \tag{5.3}$$

where $x_0 = x(0)$ and τ denotes the system's natural relaxation time scale:

$$\tau = \frac{b}{k}. \tag{5.4}$$

We make no assumptions about the initial distribution of particle positions (aside from the assumption of a finite mean and variance), and will we primarily be concerned with times for which the initial distribution has decayed to the stationary solution of the Fokker-Planck equation associated with Eq. (5.1). The stationary distribution corresponds to the thermal equilibrium distribution - a Gaussian with mean zero and variance k_BT/k [16]. From Eqs. (5.2) and (5.3), the mean, $\langle x(t) \rangle$, and variance, $\sigma_x^2(t)$, in the distribution of

particle positions at any time t can be calculated:

$$\langle x(t) \rangle = \langle x_0 \rangle e^{-\frac{t}{\tau}} \tag{5.5}$$

$$\sigma_x^2(t) = \langle x^2(t) \rangle - \langle x(t) \rangle^2$$
$$= \left(\sigma_x^2(0) - \frac{k_B T}{k} \right) e^{-\frac{2t}{\tau}} + \frac{k_B T}{k}.$$

The force of the spring on the particle at time t, denoted by $F_{sp}(t)$, is simply the standard spring force: $F_{sp}(t) = -kx(t)$. Therefore, the rate of work done by the spring on the particle at time t, denoted by $\dot{w}_{sp}(t)$, is given by

$$\dot{w}_{sp}(t) = F_{sp}(t)\dot{x}(t) \tag{5.6}$$
$$= -kx(t)\dot{x}(t).$$

This work rate fluctuates rapidly due to the fluctuations in the particle's position and velocity, but we are not interested in these momentary fluctuations. Rather, we would like to gain an intuitive sense of the total work done by the spring over long times by considering the time-averaged work rate $\overline{\dot{w}_{sp}}$. That is, we wish to calculate

$$\overline{\dot{w}_{sp}} = \lim_{t \to \infty} \frac{1}{t} \int_0^t \mathrm{d}t'\, \dot{w}_{sp}(t'), \tag{5.7}$$

The white noise process is stationary and ergodic [11], so long time-averages over a single trajectory are equivalent to averages (evaluated at large times) over all initial conditions and all realizations of the noise:

$$\langle \dot{w}_{sp}(t) \rangle \underset{t \to \infty}{=} \overline{\dot{w}_{sp}}. \tag{5.8}$$

The average over initial conditions and stochastic noise can be computed straightforwardly using the statistics of the white noise process, so it will be the average of interest here and

throughout the rest of this paper.

The following physical argument makes clear that $\langle \dot{w}_{sp}(t) \rangle$ must be identically zero at long times. Changes in the particle's kinetic energy are negligible in the overdamped limit, and there are no external influences acting on the spring-particle system aside from those associated with the surrounding thermal environment, so by the first law of thermodynamics at the level of individual trajectories, the rate at which the spring does work on the particle is equal to the rate at which heat is dissipated to the environment [16]. If the system is allowed to evolve long enough to reach thermal equilibrium with the environment, then the ergodicity and stationarity of the white noise process implies that the long-time work rate averaged over all initial particle conditions and all realizations of the stochastic noise will be equivalent to the work rate averaged over the stationary thermal state [11]. Therefore, were the average work rate not identically zero at long times, we would necessarily have a non-zero average heat flow between the system and environment once the system reaches thermal equilibrium, and this contradicts the very definition of thermal equilibrium.

To calculate $\langle \dot{w}_{sp}(t) \rangle$ directly, we first use Eqs. (5.1) and (5.6) to write

$$\langle \dot{w}_{sp}(t) \rangle = \frac{k \langle x^2(t) \rangle}{\tau} - \sqrt{2bk_BT} \frac{\langle x(t)\xi(t) \rangle}{\tau}. \tag{5.9}$$

We compute the above averages using Eqs. (5.2) and (5.3), and we find

$$\langle \dot{w}_{sp}(t) \rangle = \left[\frac{k \langle x_0^2 \rangle}{\tau} - \frac{k_BT}{\tau} \right] e^{-\frac{2t}{\tau}}. \tag{5.10}$$

It should be noted that in order to calculate the term $\langle x(t)\xi(t) \rangle$ in Eq. (5.9), one must apply the following rule associated with the Stratonovich calculus [16]:

$$\begin{aligned} \int_0^t \mathrm{d}t' \, f(t') \langle \xi(t)\xi(t') \rangle &= \int_0^t \mathrm{d}t' \, f(t')\delta \left(t - t' \right) \\ &= \frac{1}{2}f(t), \end{aligned} \tag{5.11}$$

where $f(t)$ is any integrable function. The work rate in Eq. (5.10) decays to zero over times much larger than the system time-scale τ, so as expected, we see that the spring does no work on the particle, on average, over long times:

$$\langle \dot{w}_{sp}(t) \rangle \underset{t \to \infty}{=} 0 \tag{5.12}$$

The exponentially decaying term in Eq. (5.10) is simply a transient effect which arises when the initial distribution of particle positions is not thermal. If we assume that the system begins in thermal equilibrium, so that $\langle x_0^2 \rangle = k_B T / k$, then Eq. (5.10) vanishes for all times.

5.3 Perfect feedback

Consider now a free Brownian particle in a thermal environment at temperature T with damping constant b. Suppose that at time $t = 0$, a control demon begins continuously observing the position of the particle with perfect precision and exerts a feedback force $F_{fb}^o(t)$ on the particle based on the outcome of these observations. The superscript ^{o} is a reminder that we are considering the optimal case of perfect observations. At a time t, this feedback force is a functional of entire observation history of the particle up until t. More formally, in the language of measure theory, we require that $F_{fb}^o(t)$ be a measurable function with respect to the filtration generated by $x(t)$ [9]. The feedback force $F_{fb}^o(t)$ is then said to be $x(t)$ measurable. We again consider an arbitrary distribution of particle positions at $t = 0$, assuming only the existence of a finite mean and variance. After the demon begins making observations at $t = 0$, the particle's dynamics can be modelled by the following overdamped Langevin equation:

$$b\dot{x}(t) = F_{fb}^o(t) + \sqrt{2bk_B T}\xi(t). \tag{5.13}$$

Here, and throughout the rest of this paper, we assume that the demon attempts to stabilize the position of the particle about the origin by, to the best of its ability, mimicking the action of the ideal spring described in Sec. 5.2. The demon has perfect knowledge of the position of the particle, so it can perfectly mimic the spring force by implementing the following feedback protocol:

$$F_{fb}^o(t) = -kx(t). \tag{5.14}$$

When the above feedback force is applied, Eq. (5.13) becomes identical to Eq. (5.1). The only difference between the situations described in Secs. 5.2 and this section is our definition of internal and external forces. In Sec. 5.2, the particle and spring together constitute the system of interest, so the force of the spring on the particle is considered an internal force, and we picture the spring as a component of the system which utilizes the interaction potential to preform work and stabilize the particle about the origin. In this section, there is no potential energy; the system of interested is comprised of only the particle, and the force of the control demon on the particle is considered an external force. In this case, we imagine that the demon utilizes energy from some external source in order to perform work and achieve stabilization.

The rate at which the demon does work on the particle, denoted by $\dot{w}_{ext}^o(t)$, is given by

$$\begin{aligned} \dot{w}_{ext}^o(t) &= F_{fb}^o(t)\dot{x}(t) \\ &= -kx(t)\dot{x}(t). \end{aligned} \tag{5.15}$$

Because there is no potential energy to speak of, and because kinetic energy changes are negligible, the first law of thermodynamics [16] states that any work done on the particle is immediately dissipated to the environment as heat. The work rate in Eq. (5.15) is mathematically identical to the rate at which the spring does work on the particle, so therefore, because the particle's dynamics are identical in Secs. 5.2 and 5.3, we have from Eq. (5.10)

$$\langle \dot{w}^o_{ext}(t) \rangle = \left[\frac{k \langle x_0^2 \rangle}{\tau} - \frac{k_B T}{\tau} \right] e^{-\frac{2t}{\tau}}, \tag{5.16}$$

which gives, after long times,

$$\langle \dot{w}^o_{ext}(t) \rangle \underset{t \to \infty}{=} 0. \tag{5.17}$$

The above equality can easily be verified by direct calculation as in Sec. 5.2. We thus conclude that the demon does no work on average, and likewise dissipates no heat to the environment on average, when stabilizing the position of the particle about the origin. This conclusion is perhaps a bit counter intuitive, but is logically straightforward: the average rate at which the spring does work on the particle must be zero, or else the existence of thermal equilibrium would be a contradiction, so if the demon is able perfectly mimic the force of the spring on the particle, the demon too will do no work on average.

5.4 Noisy feedback

That an external feedback controller can stabilize a mesoscale thermodynamic system without expending energy or dissipating heat is a promising conclusion from an engineering standpoint, but this result only holds under the highly idealized conditions of continuous perfect observations and a continuous perfectly applied force. In in real experiments, observations are prone to errors and occur discretely in time, applied feedback forces are prone to errors, drifts, and change discretely in time, and experimental apparatuses can introduce time-delays [99, 100]. We will now incorporate some this experimental reality into our control demon by considering observation errors in the form of stochastic noise. To keep the mathematics tractable, we will model the observation noise as a simple white noise process. Despite the simplicity of our noise model, the analysis in this section is much more involved than the analyses of Secs. 5.2 and 5.3 and will proceed in steps. After mathematically

defining the noisy observations and system dynamics, we will define an algorithm by which the demon can combine the available observational and dynamical information to estimate the actual state of the system. Next, as a preliminary step to finding a closed form solution for the particle's trajectory, we will determine the statistics of the error between the actual system state and the demon's best estimate of the system state. Finally, after solving the system's equation of motion and discussing the steady state distributions, we will calculate the average work rate required to achieve stabilization.

5.4.1   A model of system observation

Consider again a free Brownian particle in a thermal environment at temperature T with damping constant b, along with a control demon which begins continuously observing the position of the particle at time $t = 0$, but with some observation error, and exerts a feedback force $F_{fb}(t)$ based on the outcomes of the noisy observations. After time $t = 0$, the position of particle can be modelled the following overdamped Langevin equation:

$$b\dot{x}(t) = F_{fb}(t) + \sqrt{2bk_B T}\xi(t). \tag{5.18}$$

We model the noisy observation, $y(t)$, of the particle's actual position at time t with the following equation:

$$y(t) = x(t) + \sigma\eta(t). \tag{5.19}$$

where σ gives the strength of the measurement noise, and $\eta(t)$ is a standard Gaussian white noise, independent of $\xi(t)$, with mean and autocorrelation

$$\langle \eta(t) \rangle = 0,$$
$$\langle \eta(t)\eta(t') \rangle = \delta(t - t'), \tag{5.20}$$

Throughout this section, the angular brackets denote an average over all realizations of $\xi(t)$, $\eta(t)$, and initial particle conditions. We again assume nothing about the initial particle

160

distribution aside from a finite mean and variance. The feedback force at a time t is a functional of the noisy observation history up to time t, so we require $F_{fb}(t)$ to be a measurable function with respect to the filtration generated by $y(t)$ [9]. The feedback force $F_{fb}(t)$ is then said to be $y(t)$ measurable.

5.4.2 A model of state estimation

We denote by $\hat{x}(t)$ the demon's estimate of the position of the particle at time t given the noisy observation history. This implies that $\hat{x}(t)$ is a $y(t)$ measurable function. We choose to define $\hat{x}(t)$ through the following linear differential equation:

$$b\dot{\hat{x}}(t) = F_{fb}(t) + K(t)(y(t) - \hat{x}(t)), \ \hat{x}_0 = \langle x_0 \rangle. \tag{5.21}$$

The function $K(t)$ will be defined below. We note that, a priori, there is no reason to expect these simple linear dynamics to provide a good state estimate - it is simply the model of state estimation which we have chosen to use. For our system, it will turn out that of all possible estimators, even those that obey non-linear dynamics, the linear estimator described by Eq. (5.21), for appropriately chosen $K(t)$, is the unique unbiased minimal mean square error estimator of $x(t)$.

Once $K(t)$ is satisfied, Eq. (5.21) defines a recursive algorithm by which the control demon can construct the state estimate - given the current estimate and current measurement at a time t, Eq. (5.21) tells the demon what the estimate will be a short time dt later. Intuitively, Eq. (5.21) allows the demon to construct a state estimate by weighting the relative importance of two sources of information - the information gained through observation and the information gained through knowledge of the system's dynamics (the demon is choosing the feedback force F_{fb} to apply, so from Eq. (5.18), we see that the demon always has some knowledge of the system's dynamics). For very large σ, $y(t)$ provides very little information about $x(t)$, and the observations act chiefly as an useless source of noise. In this case, in order to minimize the influence of the strong noise on the state estimate, $K(t)$ should be chosen to be very small, and the evolution of $\hat{x}(t)$ will be dominated by the knowledge of the system's dynamics and nearly deterministic (without trustworthy measurements, the

161

demon can only make state estimates based on knowledge of the system's dynamics). On the other hand, if σ is very small, then the observations are very trustworthy, and $y(t)$ contains a comparatively large mount information about $x(t)$. In this case, $K(t)$ should be made very large, and the evolution of $\hat{x}(t)$ will be dominated by the knowledge gained through observations. As $K(t) \to \infty$, Eq. (5.21) shows that $\hat{x}(t)$ will relax to $y(t)$ infinitely fast, as one would hope for extremely accurate observations. We thus see that Eq. (5.21) is a mathematical expression of a noise filter [9]: for appropriately chosen $K(t)$, the filter receives $y(t)$ as input, eliminates some noise, and outputs a better estimate $\hat{x}(t)$.

We denote by $\epsilon(t)$ the error between demon's estimate and the actual system position:

$$\epsilon(t) = x(t) - \hat{x}(t). \tag{5.22}$$

Using Eq. (5.19), Eq. (5.21) can be written as

$$b\dot{\hat{x}}(t) = F_{fb}(t) + K(t)\big(x(t) - \hat{x}(t)\big) + K(t)\sigma\eta(t). \tag{5.23}$$

Subtracting Eq. (5.23) from Eq. (5.18) yields an evolution equation for $\epsilon(t)$:

$$\dot{\epsilon}(t) = -\frac{K(t)}{b}\epsilon(t) + \sqrt{\frac{2k_B T}{b}}\xi(t) - \frac{K(t)}{b}\sigma\eta(t). \tag{5.24}$$

Equation (5.24) has a closed form solution, given by

$$\epsilon(t) = \gamma(t,0)\epsilon_0 + \int_0^t dt'\gamma(t,t')\left(\sqrt{\frac{2k_B T}{b}}\xi(t') - \frac{K(t')}{b}\sigma\eta(t')\right), \tag{5.25}$$

where $\gamma(t,t')$ is defined as

$$\gamma(t,t') = e^{-\int_{t'}^{t} dt'' \frac{K(t'')}{b}}. \tag{5.26}$$

Because $\hat{x}_0 = \langle x_0 \rangle$, we have $\langle \epsilon_0 \rangle = 0$, and from Eq. (5.25) we find,

$$\langle \epsilon(t) \rangle = 0. \tag{5.27}$$

We thus see that the estimate constructed by Eq. (5.21) is an unbiased estimator.

5.4.3   The Kalman-Bucy filter

The mean error of the estimator given by Eq. (5.21) vanishes identically for any $K(t)$, so we now will attempt to find a $K(t)$ which will minimize the mean square error $\mathfrak{E}^2(t)$, defined as

$$\mathfrak{E}^2(t) = \left\langle \epsilon^2(t) \right\rangle. \tag{5.28}$$

An estimate evolving according to Eq. (5.21), with $K(t)$ chosen to minimize $\mathfrak{E}^2(t)$, is called a Kalman-Bucy filter [9]. Note that, because $\hat{x}_0 = \langle x_0 \rangle$, we have

$$
\begin{aligned}
\mathfrak{E}_0^2 &= \left\langle x_0^2 \right\rangle - \langle x_0 \rangle^2 \\
&= \sigma_x^2(0),
\end{aligned}
\tag{5.29}
$$

where we denote $\mathfrak{E}_0^2 = \mathfrak{E}^2(0)$ and $\sigma_x^2(t)$ denotes the variance in the marginal distribution of x at time t. Using Eqs. (5.27) and (5.28), we also find the useful relation

$$\sigma_x^2(t) = \sigma_{\hat{x}}^2(t) + \mathfrak{E}^2(t), \tag{5.30}$$

where $\sigma_{\hat{x}}^2(t)$ is the variance in the marginal distribution of $\hat{x}$ at time t.

By combining Eqs. (5.24) and (5.25), one can find a differential equation for the evolution of $\mathfrak{E}(t)$:

$$\frac{\mathrm{d}\mathfrak{E}^2(t)}{\mathrm{d}t} = \left(\frac{\sigma}{b} K(t) - \frac{\mathfrak{E}^2(t)}{\sigma} \right)^2 - \frac{\mathfrak{E}^4(t)}{\sigma^2} + \frac{2k_B T}{b}. \tag{5.31}$$

If we minimize $\dot{\mathfrak{E}}^2(t)$ for every $t > 0$, then we will have minimized $\mathfrak{E}^2$ for all $t > 0$. From Eq. (5.31), we see that this minimization can be realized by choosing $K(t)$ such that

$$K(t) = \frac{b}{\sigma} \mathfrak{E}^2(t), \tag{5.32}$$

which reduces Eq. (5.31) to

$$\frac{\mathrm{d}\mathfrak{E}^2(t)}{\mathrm{d}t} = -\frac{\mathfrak{E}^4(t)}{\sigma^2} + \frac{2k_B T}{b}. \tag{5.33}$$

The solution to Eq. (5.33) is given by

$$\mathfrak{E}^2(t) = \begin{cases} \sigma\sqrt{\frac{2k_B T}{b}} \tanh\left[\frac{t}{\tau_\epsilon} + \operatorname{arctanh}\left(\frac{\mathfrak{E}_0^2}{\sigma\sqrt{\frac{2k_B T}{b}}}\right)\right], & \mathfrak{E}_0^2 < \sigma\sqrt{\frac{2k_B T}{b}} \\ \sigma\sqrt{\frac{2k_B T}{b}}, & \mathfrak{E}_0^2 = \sigma\sqrt{\frac{2k_B T}{b}} \\ \sigma\sqrt{\frac{2k_B T}{b}} \coth\left[\frac{t}{\tau_\epsilon} + \operatorname{arccoth}\left(\frac{\mathfrak{E}_0^2}{\sigma\sqrt{\frac{2k_B T}{b}}}\right)\right], & \mathfrak{E}_0^2 > \sigma\sqrt{\frac{2k_B T}{b}}, \end{cases} \tag{5.34}$$

where τ_ϵ is the natural error relaxation timescale:

$$\tau_\epsilon = \frac{\sigma}{\sqrt{\frac{2k_B T}{b}}}. \tag{5.35}$$

The initial value of $\mathfrak{E}^2$ is determined by the initial variance in the marginal distribution of x, but after times much larger than τ_ϵ, the mean square error will relax to the steady state value $\sigma\sqrt{2k_B T/b}$. From Eq. (5.32), we find

$$K(t) = \begin{cases} \frac{\sqrt{2bk_B T}}{\sigma} \tanh\left[\frac{t}{\tau_\epsilon} + \operatorname{arctanh}\left(\frac{\mathfrak{E}_0^2}{\sigma\sqrt{\frac{2k_B T}{b}}}\right)\right], & \mathfrak{E}_0^2 < \sigma\sqrt{\frac{2k_B T}{b}} \\ \frac{\sqrt{2bk_b T}}{\sigma}, & \mathfrak{E}_0^2 = \sigma\sqrt{\frac{2k_B T}{b}} \\ \frac{\sqrt{2bk_B T}}{\sigma} \coth\left[\frac{t}{\tau_\epsilon} + \operatorname{arccoth}\left(\frac{\mathfrak{E}_0^2}{\sigma\sqrt{\frac{2k_B T}{b}}}\right)\right], & \mathfrak{E}_0^2 > \sigma\sqrt{\frac{2k_B T}{b}}. \end{cases} \tag{5.36}$$

A $K(t)$ chosen according to Eq. (5.36) will thus yield the best (in the minimal mean square error sense) estimator out of all linear estimators of the form of Eq. (5.21), but we still do not know whether or not there exists an even better, possibly non-linear, estimator which

does not take the from of Eq. (5.21).

5.4.4 Solution trajectories

The control demon would like to exert a feedback force on the Brownian particle which mimics the force of an ideal spring, but the demon only knows its best linear estimate $\hat{x}(t)$ and not the actual position $x(t)$. Thus, the demon's best course of action is to implement the following feedback protocol:

$$F_{fb}(t) = -k\hat{x}(t),$$

With this feedback protocol selected, the state and state estimator are described by two coupled linear stochastic differential equations:

$$
\begin{aligned}
b\dot{x}(t) &= -k\hat{x}(t) + \sqrt{2bk_BT}\xi(t) & (5.37) \\
b\dot{\hat{x}}(t) &= -k\hat{x}(t) + K(t)\big(x(t) - \hat{x}(t)\big) + K(t)\sigma\eta(t), \ \hat{x}_0 = \langle x_0 \rangle
\end{aligned}
$$

with $K(t)$ defined by Eq. (5.36). The closed form solution trajectories to these equations are

$$x(t) = e^{-\frac{t}{\tau}}x_0 + \int_0^t \mathrm{d}t' \, e^{-\frac{(t-t')}{\tau}} \left[\frac{\epsilon(t')}{\tau} + \sqrt{\frac{2k_BT}{b}}\xi(t') \right], \qquad (5.38)$$

and

$$\hat{x}(t) = e^{-\frac{t}{\tau}}\langle x_0 \rangle + \int_0^t \mathrm{d}t' \, e^{-\frac{(t-t')}{\tau}} \left[\frac{K(t')}{b}\epsilon(t') + \frac{\sigma}{b}K(t')\eta(t') \right], \qquad (5.39)$$

with $\epsilon(t)$ given by Eq. (5.25) and $K(t)$ given by Eq. (5.36).

Due to the linearity of the evolution equations for $x(t)$, a classical theorem from control theory asserts that the Kalman-Bucy filter we have constructed is in fact, out of all possible (even non-linear) estimators, the minimum mean square error estimator, given the model

165

of observation specified by Eq. (5.19) (see Ref. [9] for a detailed explanation). Stated more formally, we have

$$\hat{x}(t) = \underset{z}{\arg\min} \left\{ \left\langle \left(x(t) - z \right)^2 \right\rangle | z \text{ is } y(t) \text{ measurable} \right\}. \tag{5.40}$$

One can also show, in general, that for any time t, the set of all $x(t)$ measurable functions with finite mean square forms a Hilbert space which contains the set of all $y(t)$ measurable functions with finite mean square as a Hilbert subspace [9]. The inner product of two functions $a(t)$ and $b(t)$ is defined as the expectation $\langle a(t)b(t) \rangle$. The definition in Eq. (5.40) then implies that, for any time t, $\hat{x}(t)$ is the orthogonal projection of $x(t)$ from the space of $x(t)$ measurable function onto the subspace of $y(t)$ measurable functions. As a consequence, we have

$$\langle \epsilon(t)\hat{x}(t) \rangle = \langle [x(t) - \hat{x}(t)] \, \hat{x}(t) \rangle \tag{5.41}$$
$$= 0,$$

for all t, and we find

$$\langle x(t)\hat{x}(t) \rangle = \left\langle \hat{x}^2(t) \right\rangle. \tag{5.42}$$

This relation, along with Eq. (5.27), implies that

$$\sigma_{x\hat{x}}^2(t) = \sigma_{\hat{x}}^2(t), \tag{5.43}$$

where $\sigma_{x\hat{x}}^2(t)$ is the covariance between x and $\hat{x}$ at time t:

$$\sigma_{x\hat{x}}^2(t) = \langle x(t)\hat{x}(t) \rangle - \langle x(t) \rangle \langle \hat{x}(t) \rangle \tag{5.44}$$

We note that the above relations can all be verified by direct calculation using Eqs. (5.25), (5.38), and (5.39).

5.4.5 Distributions

We are ultimately interested in the average rate at which the demon does work at long times, so we now discuss the statistics of the particle and estimate in the long time limit. We assume that x is distributed initially according to some distribution P_0^x, and $\hat{x}$ is initially fixed by $\hat{x}_0 = \langle x_0 \rangle$, so the initial joint distribution $P_0^{x\hat{x}}$ is given by

$$P_0^{x\hat{x}} = P_0^x \, \delta(\hat{x} - \langle x_0 \rangle). \tag{5.45}$$

After time $t = 0$, the dynamics of the particle and estimate become time-dependent due to the time-dependence of $K(t)$, and the distribution evolves in a complicated manner. The time-dependent means and variances for the marginal distributions, however, can be calculated analytically. From Eqs. (5.38) and (5.39), we see immediately that

$$
\begin{aligned}
\langle x(t) \rangle &= \langle x_0 \rangle \, e^{-\frac{t}{\tau}} \\
\langle \hat{x}(t) \rangle &= \langle x_0 \rangle \, e^{-\frac{t}{\tau}}.
\end{aligned}
\tag{5.46}
$$

In Appendix B, we calculate the time-dependent variances of the marginal distributions of x and $\hat{x}$, and we find

$$
\begin{aligned}
\sigma_x^2(t) &= \left(\sigma_x^2(0) - \frac{k_B T}{k} \right) e^{-\frac{2t}{\tau}} + \frac{k_b T}{k} + 2 \int_0^t \frac{dt'}{\tau} e^{-\frac{2(t-t')}{\tau}} \mathfrak{E}^2(t') \\
\sigma_{\hat{x}}^2(t) &= \left(\sigma_x^2(0) - \frac{k_B T}{k} \right) e^{-\frac{2t}{\tau}} + \frac{k_b T}{k} + 2 \int_0^t \frac{dt'}{\tau} e^{-\frac{2(t-t')}{\tau}} \mathfrak{E}^2(t') - \mathfrak{E}^2(t).
\end{aligned}
\tag{5.47}
$$

In the long-time limit, $\mathfrak{E}^2(t)$ relaxes to the constant value $\sigma \sqrt{2 k_B T / b}$, and the variances of the marginal distributions relax to the constant values

$$
\begin{aligned}
\sigma_x^2 &\underset{t \to \infty}{=} \frac{k_B T}{k} + \sigma \sqrt{\frac{2 k_B T}{b}} \\
\sigma_{\hat{x}}^2 &\underset{t \to \infty}{=} \frac{k_B T}{k}.
\end{aligned}
\tag{5.48}
$$

From Eq. (5.46), it is clear that the marginal means relax to the constant value of zero, and by Eq. (5.43), we see the covariance between x and $\hat{x}$ relaxes to the same constant to which $\sigma_{\hat{x}}^2(t)$ relaxes.

From Eq. (5.36), we see that for times much larger than τ_ϵ, $K(t)$ relaxes to the constant value $\sqrt{2bk_BT}/\sigma$, and the stochastic differential equations describing x and $\hat{x}$ reduce to

$$
\begin{aligned}
b\dot{x}(t) &\underset{t\to\infty}{=} -k\hat{x}(t) + \sqrt{2bk_BT}\xi(t) && (5.49) \\
b\dot{\hat{x}}(t) &\underset{t\to\infty}{=} -k\hat{x}(t) + \frac{\sqrt{2bk_BT}}{\sigma}(x(t) - \hat{x}(t)) + \sqrt{2bk_BT}\eta(t).
\end{aligned}
$$

Equation (5.49) is a linear stochastic system with constant coefficients, so if a steady state distribution exists, we expect it to be a Gaussian [11]. In Appendix C, we verify directly that the steady state distribution is given by

$$
P_s^{x\hat{x}} = \sqrt{\frac{1}{2\pi\sigma\sqrt{\frac{2k_BT}{b}}}\frac{k}{2\pi k_BT}}\,e^{-\frac{(x-\hat{x})^2}{2\sigma\sqrt{\frac{2k_BT}{b}}}-\frac{k\hat{x}^2}{2k_BT}}. \tag{5.50}
$$

From Eq. (5.50), the steady state marginal distributions of x and $\hat{x}$ can easily be calculated; $\hat{x}$ relaxes to a Gaussian with mean zero and variance k_BT/k, and x relaxes to a Gaussian with mean zero and variance $k_BT/k + \sigma\sqrt{2k_BT/b}$. We thus see that the estimate $\hat{x}$ relaxes to the same thermal distribution to which the uncontrolled particle attached to the spring in Sec. 5.2 relaxes. The reason for this is clear - if the demon is mimicking the action of the spring to the best of its ability, then the demon's best estimate of the position of the particle ought to relax to the thermal distribution described in Sec. 5.2. Experimental evidence for this prediction has been observed in Brownian feedback traps [99, 101]. The actual position of the particle, is distributed as a Gaussian with a larger variance than the thermal variance due to the observation noise feeding into and corrupting the system. We close this subsection by noting that in the limit of vanishingly small observation noise, the joint steady state distribution reduces to

$$P_s^{x\hat{x}} \underset{\sigma \to 0}{=} \sqrt{\frac{k}{2\pi k_B T}} \delta(x - \hat{x}) e^{-\frac{kx^2}{2k_B T}}. \tag{5.51}$$

As expected, for zero observation error, both x and $\hat{x}$ are distributed thermally in the steady state, and are perfectly correlated with each other.

5.4.6 Work

We now calculate the rate of work done by the demon on the particle. The demon is an external agent exerting a force $F_{fb}(t) = -k\hat{x}(t)$ on the particle (as opposed to a force on the particle which arises from a time-dependent potential), so the proper definition of work in this context is the exclusive definition [88]. The rate at which the demon does work on the particle, denoted by $\dot{w}_{ext}(t)$, is thus

$$\begin{aligned} \dot{w}_{ext}(t) &= F_{fb}(t)\dot{x}(t) \\ &= -k\hat{x}(t)\dot{x}(t). \end{aligned} \tag{5.52}$$

From Eq. (5.37) and the definition $\epsilon(t) = x(t) - \hat{x}(t)$, we have

$$\dot{w}_{ext}(t) = -kx(t)\dot{x}(t) - \frac{k}{\tau}\epsilon(t)\hat{x}(t) + \frac{1}{\tau}\sqrt{2bk_B T}\epsilon(t)\xi(t). \tag{5.53}$$

Taking averages and rewriting the first term on the left hand side yields

$$\langle \dot{w}_{ext}(t) \rangle = -\frac{d}{dt}\left\langle \frac{1}{2}kx^2(t) \right\rangle - \frac{k}{\tau}\langle \epsilon(t)\hat{x}(t) \rangle + \frac{1}{\tau}\sqrt{2bk_B T}\langle \epsilon(t)\xi(t) \rangle \tag{5.54}$$

Before actually calculating the above work rate, we comment on the meaning of each term appearing in the equation. In the long time limit, when the particle and estimate have reached the steady state, we have $\langle x^2(t) \rangle = \sigma_x^2(t)$, and we know that $\sigma_x^2(t)$ is constant in time by the very definition of steady state, so the first term on the right hand side of

169

Eq. (5.54) will vanish. This first quantity thus represents the rate of energy consumption associated with evolving the particle's distribution towards the steady state. The second term on the right hand side of Eq. (5.54) vanishes for all times due to Eq. (5.30). This term will always vanish when the $\hat{x}$ is constructed to be the optimal estimate, so it represents the extra energetic price the demon would have had to pay had it sub-optimally estimated the position of the particle. The last term on the right hand side of Eq. (5.54) represents the energetic penalty rate paid by demon due to its noisy observations.

The error penalty rate $\sqrt{2bk_BT} \langle \epsilon(t)\xi(t)\rangle /\tau$ can be calculated straightforwardly by using Eq. (5.25):

$$\langle \epsilon(t)\xi(t)\rangle = \frac{1}{2}\sqrt{\frac{2k_BT}{b}}. \tag{5.55}$$

The cost associated with approaching the steady state $\frac{\mathrm{d}}{\mathrm{d}t}\langle kx^2(t)\rangle /2$ can be calculated by taking the time derivatives of $\langle x(t)\rangle^2$ and $\sigma_x^2(t)$ in Eqs. (5.46) and (5.47), respectively, and we find the average work rate at any time t to be

$$
\begin{aligned}
\langle \dot{w}_{ext}(t)\rangle \;=\; & \left[\left(\frac{k\langle x_0^2\rangle}{\tau} - \frac{k_BT}{\tau}\right)e^{-\frac{2t}{\tau}} \right. \\
& \left. +\frac{k}{\tau}\left(2\int_0^t \frac{\mathrm{d}t'}{\tau}e^{-\frac{2(t-t')}{\tau}}\mathfrak{E}^2(t') - \mathfrak{E}^2(t)\right)\right] + \frac{k_BT}{\tau}.
\end{aligned}
\tag{5.56}
$$

The term in the square brackets in Eq. (5.56) is due to the term $\frac{\mathrm{d}}{\mathrm{d}t}\langle kx^2(t)\rangle /2$ in Eq. (5.54), and vanishes as expected in the long time limit when $\mathfrak{E}^2(t)$ reaches its steady state value. Ignoring this term, we finally have the long time average rate of work preformed by an imperfect demon when stabilizing a Brownian particle:

$$\langle \dot{w}_{ext}(t)\rangle \underset{t\to\infty}{=} \frac{k_BT}{\tau}. \tag{5.57}$$

The quantity k_BT/τ is a natural unit of power associated with our system - the natural thermal energy unit k_BT divided by the system's natural time scale τ.

Equation (5.57) is the central result of this section, and it presents an immediate problem. According to Eq. (5.57), the average work rate is independent of the observation noise strength σ, and this implies that in the limit where σ approaches zero, the rate of work preformed by the demon making noisy observations from Sec. 5.4 does not converge to the rate of work preformed by the demon making perfect observations from Sec. 5.3:

$$\lim_{\sigma \to 0} \langle \dot{w}_{ext}(t) \rangle \underset{t \to \infty}{=} \frac{k_B T}{\tau} \neq 0 \underset{t \to \infty}{=} \langle \dot{w}^0_{ext}(t) \rangle. \tag{5.58}$$

We are thus presented with an apparent paradox. When σ approaches 0, we see from Eq. (5.25) that $\epsilon(t)$ vanishes, so $x(t) = \hat{x}(t)$, and thus, given the same realizations of $\xi(t)$ and initial particle conditions, the particles from Secs. 5.3 and 5.4 will experience the same force and follow the same trajectory, so Eq. (5.58) can not be correct. This means that either Eq. (5.17) or Eq. (5.57) must be incorrect. However, we know that Eq. (5.17) must be correct, or else the existence of thermal equilibrium for a Brownian particle attached to a spring would be a contradiction, and we know that Eq. (5.57) is correct due to our careful derivation. We now focus our efforts on resolving this paradox.

5.5   Correlated thermal force

The constant work rate $k_B T/\tau$ in Eq. (5.57) originates from the term $\sqrt{2b k_B T} \langle \epsilon(t) \xi(t) \rangle /\tau$ in Eq. (5.54), so we immediately suspect the white noise process $\xi(t)$ (which models the fluctuating thermal force in the evolution equation for $x(t)$) to be the source of our paradox. This suspicion is not entirely surprising - the Gaussian white noise process is uncorrelated down to infinitely fine time-scales and produces noise uniformly across all frequencies, so it is somewhat pathological. Any real force, no matter how seemingly irregular, will have a non-zero (possibly immeasurably small) correlation time. We will therefore attempt to resolve our paradox by considering a more realistic model of the fluctuating thermal force.

Consider an Ornstein-Uhlenbeck process $z(t)$, defined by [11]

$$\tau_z \dot{z}(t) = -z(t) + \xi(t), \tag{5.59}$$

where $\xi(t)$ is the white noise process defined by Eq. (5.2), τ_z is the process's effective correlation time, and $z(0) = z_0$ is assumed to be distributed as a Gaussian with mean zero and variance $1/2\tau_z$. The closed form solution to Eq. (5.59) is given by

$$z(t) = z_0 e^{-\frac{t}{\tau_z}} + \int_0^t \frac{\mathrm{d}t'}{\tau_z} e^{-\frac{(t-t')}{\tau_z}} \xi(t'). \tag{5.60}$$

Using Eqs. (5.2) and (5.60), we find the mean and auto correlation of $z(t)$ to be

$$
\begin{aligned}
\langle z(t) \rangle &= 0 \\
\langle z(t)z(t') \rangle &= \frac{1}{2\tau_z} e^{-\frac{|t-t'|}{\tau_z}}.
\end{aligned}
\tag{5.61}
$$

In the limit $\tau_z \to 0$, Eq. (5.61) reduces to Eq. (5.2), and Eq. (5.59) reduces to $z(t) = \xi(t)$. We thus see that for very a small correlation time, an Onstein-Uhlenbeck process is well approximated by a white noise process.

A classical thermal bath is ultimately a deterministic Hamiltonian system, so the fluctuating thermal force acting on a Brownian particle must be correlated over short time scales (even if these time scales are immeasurably small) and is only approximately described by a white noise process. Therefore, we now consider a free Brownian particle under the influence of a control demon with the fluctuating thermal force modelled by $z(t)$, where τ_z is assumed to be much smaller than the natural system time-scale τ. The equation of motion for the position of particle, $x(t)$, is given by

$$b\dot{x}(t) = -k\hat{x}(t) + \sqrt{2bk_B T}\, z(t), \tag{5.62}$$

where $\hat{x}(t)$ again denotes the demon's estimate of the position of the particle. Again, we assume nothing about the initial distribution of x aside from a finite mean and variance. We model the demon's noisy observations, $y(t)$, with a white noise process as in Eq. (5.19):

$$y(t) = x(t) + \sigma\eta(t). \tag{5.63}$$

Finding the minimal mean square error estimator of $x(t)$, given Eqs. (5.62) and (5.63), is not a straightforward task as in Sec. 5.4. We will therefore model the demon's estimate in the same manner as in Eq. (5.21), but with $K(t)$ chosen to be equal to the optimal steady state value $\sqrt{2bk_BT}/\sigma$ calculated in Sec. 5.4:

$$
\begin{aligned}
b\dot{\hat{x}}(t) &= -k\hat{x}(t) + \frac{\sqrt{2bk_BT}}{\sigma}(y(t) - \hat{x}(t)) \tag{5.64} \\
&= -k\hat{x}(t) + \frac{\sqrt{2bk_BT}}{\sigma}(x(t) - \hat{x}(t)) + \sqrt{2bk_BT}\eta(t),
\end{aligned}
$$

where we set $\hat{x}(0) = \langle x_0 \rangle$. This model of state estimation is not optimal in the minimal mean square sense (as in Sec. 5.4), but for τ_z much smaller than τ, Eq. (5.64) should provide a reasonable estimator in the long time limit. We note that the joint stochastic process described Eqs. (5.62) and (5.64) does not obey the fluctuation-dissipation theorem, so the existence of a thermal steady state in the long time limit is dubious [83]. Regardless, the long time average work rate will turn out to be constant and finite.

The equation of motion for the error between the actual particle position and the demon's estimate, $\epsilon(t) = x(t) - \hat{x}(t)$, is given by

$$\dot{\epsilon}(t) = -\frac{\epsilon(t)}{\tau_\epsilon} + \sqrt{\frac{2k_BT}{b}}z(t) - \sqrt{\frac{2k_BT}{b}}\eta(t), \tag{5.65}$$

where τ_ϵ is the natural error time-scale defined in Eq. (5.35). The closed form solution of Eq. (5.65) is given by

$$\epsilon(t) = \epsilon_0 e^{-\frac{t}{\tau_\epsilon}} + \int_0^t dt' \, e^{-\frac{(t-t')}{\tau_\epsilon}} \sqrt{\frac{2k_B T}{b}} \left(z(t') - \eta(t') \right), \qquad (5.66)$$

where $\epsilon_0 = x_0 - \langle x_0 \rangle$. Using Eq. (5.66), the closed form solutions to Eqs. (5.62) and (5.64) can be written as

$$x(t) = x_0 e^{-\frac{t}{\tau}} + \int_0^t dt' \, e^{-\frac{(t-t')}{\tau}} \left[\frac{\epsilon(t')}{\tau} + \sqrt{\frac{2k_B T}{b}} z(t') \right], \qquad (5.67)$$

and

$$\hat{x}(t) = \langle x_0 \rangle e^{-\frac{t}{\tau}} + \int_0^t dt' \, e^{-\frac{(t-t')}{\tau}} \left[\frac{\epsilon(t')}{\tau_\epsilon} + \sqrt{\frac{2k_B T}{b}} \eta(t') \right], \qquad (5.68)$$

respectively.

The rate at which the demon does work on the particle at time t, denoted by $\dot{w}_{ext}^z(t)$, is defined as in Sec. 5.4:

$$\dot{w}_{ext}^z(t) = -k\hat{x}(t)\dot{x}(t). \qquad (5.69)$$

Taking expectations, and using the definition $\epsilon(t) = x(t) - \hat{x}(t)$ along with Eq. (5.62), we have

$$\langle \dot{w}_{ext}^z(t) \rangle = -\frac{\mathrm{d}}{\mathrm{d}t} \left\langle \frac{1}{2} k x^2(t) \right\rangle - \frac{k}{\tau} \langle \epsilon(t)\hat{x}(t) \rangle + \frac{1}{\tau} \sqrt{2bk_B T} \langle \epsilon(t) z(t) \rangle. \qquad (5.70)$$

The expectations on the right hand side of Eq.(5.70) can be computed by substituting in the expressions for the solution trajectories in Eqs. (5.60), (5.66), (5.67), and (5.68) and then making repeated use of the identities in Eqs. (5.20) and (5.61), the result of which is a series of nested integrals of exponentials and delta functions which are tedious but straightforward to compute. The resulting expressions are time-dependent and cumbersome, but simplify substantially in the long-time limit. The first term vanishes:

$$-\frac{\mathrm{d}}{\mathrm{d}t}\left\langle\frac{1}{2}kx^2(t)\right\rangle \underset{t\to\infty}{=} 0 \tag{5.71}$$

The second term does not vanish at large times due to the demon's sub-optimal method of estimation:

$$-\frac{k}{\tau}\left\langle\epsilon(t)\hat{x}(t)\right\rangle \underset{t\to\infty}{=} \frac{k_BT}{\tau}\left(\frac{\tau_z}{\tau}\right)^2\frac{1}{\left(1+\frac{\tau_\epsilon}{\tau}\right)\left(1+\frac{\tau_z}{\tau}\right)\left(1+\frac{\tau_z}{\tau_\epsilon}\right)} \tag{5.72}$$

The error time-scale τ_ϵ, defined in Eq. (5.35), is proportional the observation noise strength, and we see that the sub-optimal estimation cost correspondingly vanishes in the limit of perfect observations (that is, the limit $\tau_\epsilon \to 0$). In the limit $\tau_z \to 0$, the system reduces to the system described in Sec. 5.4, so the estimator becomes optimal in the long time limit, and Eq. (5.72) correspondingly vanishes. The sub-optimal estimation cost, being a second order quantity in τ_z/τ, is considered small and is not the central quantity of interest. The last term on the right hand side of Eq. (5.70), whose counterpart in Sec. 5.4 produced the paradox, is given at large times by

$$\frac{1}{\tau}\sqrt{2bk_BT}\left\langle\epsilon(t)z(t)\right\rangle \underset{t\to\infty}{=} \frac{k_BT}{\tau}\frac{1}{1+\frac{\tau_z}{\tau_\epsilon}}. \tag{5.73}$$

Equations (5.70), (5.71), (5.72), and (5.73) together give

$$\left\langle\dot{w}_{ext}^z(t)\right\rangle \underset{t\to\infty}{=} \frac{k_BT}{\tau}\frac{1}{1+\frac{\tau_z}{\tau_\epsilon}}\left[1+O\left(\frac{\tau_z}{\tau}\right)^2\right]. \tag{5.74}$$

Ignoring the small $O\left(\tau_z/\tau\right)^2$ contribution from the sub-optimal estimation cost, we arrive at the central result of this chapter:

$$\langle \dot{w}_{ext}^z(t) \rangle \underset{t \to \infty}{=} \frac{k_B T}{\tau} \frac{1}{1 + \frac{\tau_z}{\tau_\epsilon}}. \tag{5.75}$$

Equation (5.75) clarifies and resolves the paradox discovered in Sec. 5.4. We see that for small but non-zero τ_z, in the limit of vanishing observation noise strength, the average work rate vanishes:

$$\lim_{\sigma \to 0} \langle \dot{w}_{ext}^z(t) \rangle \underset{t \to \infty}{=} 0 \underset{t \to \infty}{=} \langle \dot{w}_{ext}^0(t) \rangle. \tag{5.76}$$

Thus, given perfect observations, the demon can perfectly mimic the force of a spring on the particle and stabilize without doing work or dissipating heat, in agreement with the results of Secs. 5.2 and 5.3. More generally, if the observation noise is small but non-zero, so that $\tau_\epsilon \ll \tau_z$, we have

$$\langle \dot{w}_{ext}^z(t) \rangle \underset{\substack{t \to \infty \\ \tau_\epsilon \ll \tau_z}}{=} O\left(\frac{\tau_\epsilon}{\tau_z}\right). \tag{5.77}$$

On the other hand, if the observation noise is large, so that $\tau_\epsilon \gg \tau_z$, we have

$$\langle \dot{w}_{ext}^z(t) \rangle \underset{\substack{t \to \infty \\ \tau_\epsilon \gg \tau_z}}{=} \frac{k_B T}{\tau} + O\left(\frac{\tau_z}{\tau_\epsilon}\right). \tag{5.78}$$

Thus, if the demon wants to perform work at a rate much less than the natural thermal power $k_B T / \tau$, the demon's observations must be accurate enough such that the error time scale is much smaller than the thermal force's correlation time.

From Eq. (5.75), we see that the paradoxical result of Sec. 5.4 is related to the ordering of two limits - the perfect observation limit and the uncorrelated thermal force limit. If one takes the uncorrelated thermal force limit first, Eq. (5.75) evaluates to $k_B T / \tau$ as in Sec. 5.4, and if one takes the perfect observation limit first, Eq. (5.75) evaluates to zero as

5.6 Summary and conclusions

In this chapter we have shown that, given the existence of short lived correlations in the fluctuating thermal force, a control demon can stabilize a diffusing particle without doing work or dissipating heat by perfectly mimicking the force of an ideal spring. Applying the same thermal equilibrium arguments from Sec. 5.2, we conclude that this result should also hold if the demon stabilizes by perfectly mimicking any conservative force. We have also shown that in order for the demon to perform work at a rate much less than the natural thermal power unit $k_B T/\tau$, the natural time scale associated with the demon's errors must be much smaller than the thermal force's correlation time. This result, however, calls into question the validity of the overdamped limit in the presence of an extremely accurate demon.

The overdamped limit is only valid when the particle's inertia m is negligible such that the momentum relaxation time scale $\tau_p = m/b$ is much smaller than all other relevant time scales associated with the problem [16]. The motion of the particle is ballistic over time scales smaller than τ_p, and diffusive over larger time scales [102, 103]. Intuitively, the Brownian particle is subject to erratic, seemingly random collisions with the microscopic particles of surrounding medium, and due to the Brownian particle's inertia, it takes time intervals of order τ_p for the cumulative force of these tiny collisions to have a noticeable effect on the particle's motion. When the demon's error relaxation time scale approaches zero, it will estimate the position of the Brownian particle and exert corresponding feedback forces by making computations which operate over time scales far smaller than the ballistic time scale, so using the overdamped Langevin equation is not a priori justified. A complete treatment of the problem using the full underdamped equations of motion would address this issue, but because the equation for $\dot{x}(t)$ becomes non-local in time, the problem becomes much more complicated and is no longer analytically tractable. Some insight into this issue may gained by considering lowest order corrections to the overdamped limit, as is done in Refs. [82] and [104], and this is a topic we hope to explore in future work. When

the full underdamped equations of motion are used, the fluctuating thermal force appearing in the equation for $\dot{x}(t)$ is endowed with a correlation time of τ_p, so we suspect that the correlations that were introduced artificially in Sec. 5.5 may be related physically to the ballistic time-scale. Regardless, we conclude that if a control demon is to counteract the diffusive destabilization of a Brownian particle without doing work or dissipating heat, then it needs the ability to take advantage of correlations in the fluctuating thermal force, and it must have control over time-scales smaller than those over which diffusion is an accurate description of the particle's motion.

The control demon results presented in this chapter can be framed in terms of the second law of thermodynamics and Maxwell's demon. A Maxwell's demon utilizes the information gained through observation of a thermodynamic system in order to rectify thermal fluctuations and enact apparent violations of the second law, such as decreasing entropy without doing work or extracting work from a single temperature heat bath without increasing entropy [21, 22, 23]. A control demon observes a Brownian particle and attempts to rectify thermal fluctuations in order to counter-act diffusive motion and maintain a concentrated particle distribution centered at a fixed location. Were it not for the control demon's efforts, the particle's distribution would expand outward forever, continually increasing the total thermodynamic entropy of the universe. If the control demon is present and operates at time-scales above the thermal force's correlation time-scale, there is no entropy increase associated with the particle's diffusive motion, but the demon continually dissipates heat at a rate of $k_B T/\tau$ and thus continually increases the thermodynamic entropy of the universe, so there is no apparent second law violation. It is only when the demon operates far below the thermal force correlation time-scale that it is able to apparently violate the second law and counteract the entropic increase associated with diffusive destabilization without increasing entropy elsewhere in the universe. Thus, in order to rectify thermal fluctuations and apparently violate the second law, the demon must be able to take advantage of short-lived correlations in the fluctuating thermal force acting on the particle.

We close by discussing the experimental implications of our work. Experimental realizations of a control demon exist in the form of Brownian feedback traps [99, 100, 101, 105]. A Brownian feedback trap functions to confine a small charged Brownian particle suspended

in solution by imaging the position of the particle and using surrounding macroscopic electrodes to exert corresponding feedback forces, in effect creating a virtual trapping potential. This set-up contrasts with an optical tweezers experiment, where a focused laser is used to confine a Brownian particle by creating a genuine trapping potential [102, 103]. Our results predict constraints on the rate of heat dissipated by Brownian feedback trap in terms of imaging resolution and the time-delay associated with calculating the estimated the particle position. In practice, other experimental realities, such time-delays associated with the experimental apparatuses, discrete-time observations, discretely applied feedback forces, and errors in applied feedback forces, will be present and may lead to further dissipation comparable to or greater than the dissipation due to observation error [99]. Furthermore, a particle suspended in solution may be subject to hydrodynamic forces, and these effects may also influence the observed rate of dissipation [103]. These experimental realities warrant further theoretical investigation using the control demon framework.

Chapter 6

Summary and outlook

The universe is a complicated chaotic environment comprised of many, many interacting degrees of freedom. Scientists only observe and experiment on small subsets of these degrees of freedom at any one time, and the unobserved degrees of freedom can potentially, sometimes drastically, affect the evolution of the observed degrees of freedom. There will thus always exist a need for scientists to describe the effects of the unobserved on the observed, and stochastic processes function as modelling tools which are useful for precisely this purpose. When used for modelling in the physical sciences in particular, a stochastic process always represents a simplified or effective description of some underlying complicated physical process which is ultimately deterministic in origin. The central message of this thesis is that, by understanding the details of how stochastic processes in mesoscale physical models naturally emerge from the underlying complicated microscale physical processes, we stand to obtain useful new results as well as the means to address inconsistencies and conceptual difficulties for which the alternative approach would have proven insufficient - the alternative approach being to simply accept stochastic processes at face value as modelling tools, without regard for their underlying physical origins. In this thesis, we have presented a survey of results from the fields of time-dependent billiards, the dynamical foundations of thermodynamics, and the feedback control of mesoscale thermodynamic systems, all of which illustrate our central message. We now summarize our work, highlight our new and original contributions, and discuss potential directions for future research.

In Chap. 2, we constructed the quivering limit as a description of time-dependent billiard motion in the limit of infinitely small boundary displacements. The original motivation for introducing the quivering limit was to find a physically consistent but simplified description of the microscopic dynamics associated with a particle evolving inside a time-dependent billiard. The resulting system turned out to be stochastic, with particle energy universally evolving under a mesoscopic diffusion equation, and this universal stochastic mesoscopic description allowed us to address some long-standing problems in the deterministic time-dependent billiard literature. The confusion associated with these long-standing problems can, in some sense, be attributed to the universal presence of the previously undiscovered quivering limit. The quivering limit is an original and important contribution to the field of time-dependent billiards because it is a *physically consistent* fixed-wall simplification of billiard motion. In contrast to the physically inconsistent fixed-wall simplifications introduced in previous literature [38, 50, 63], the quivering limit explicitly incorporates the physical fact that, no matter how small the amplitude of wall oscillations, collisions between particles and walls with large relative speeds of approach are statistically more likely to occur than collisions with small relative speeds of approach.

Currently, we are employing quivering billiards to study the energy dynamics of time-dependent billiards with finitely massive walls. In these systems, the billiard shape itself is considered to be a physical degree of freedom, with the Hamiltonian notions of energy and inertia corresponding to the billiard wall's kinetic energy and mass, respectively. The finite wall mass allows the billiard to gain and loose energy due to collisions with an enclosed non-interacting gas, and the gas can effectively interact with itself through these energy exchanges. For chaotic billiard shapes whose wall's mass is much, much greater than that of the enclosed particles, the billiard and gas can together be described using the chaos bath formalism established in Chap. 3, with the system of interest corresponding to the billiard shape and the enclosed non-interacting gas functioning as the baths. The quivering limit allows us to consider the energy dynamics without worrying about the complications associated with the changing billiard shape, and we hope to use this framework to make analytical progress on the adiabatic piston problem [106] and a microscopic derivation of Fourier's law of heat conduction [107].

In Chaps. 3 and 4, we developed and employed the chaos bath framework to explore the microscopic foundations of thermodynamics. We derived the chaos bath equivalent of important results from non-equilibrium thermodynamics (results which are well-established within the framework of stochastic thermodynamics), and by using the micro-meso connection, we showed that these results have simple, somewhat trivial, Hamiltonian origins. To the best of our knowledge, our work represents the first time that the ergodic adiabatic Hamiltonian framework of Berry and Robbins [77, 78], Wilkinson [48], and Jarzynski [68] has been explicitly connected to the framework stochastic thermodynamics. Specifically, the relationship between stochastic action, path probabilities, entropy production, the adiabatic reaction forces, and heat flow at the level of individual mesoscale trajectories, as demonstrated by Eqs. (4.118), (4.119), and (4.120), is new result for ergodic adiabatic Hamiltonian systems. The starting point for our work was the derivation of the Fokker-Planck equation (4.7) for the mesoscale evolution of the system of interest and bath energies. This equation was previously derived by Jarzynski [68] for the special case of a single chaos bath with no potential energy associated with the system of interest - our work is a generalization of Jarzynski's derivation to the case of multiple baths and a time-dependent potential energy function. The associated Langevin equation in Eq. (4.29), is, to the best of our knowledge, the first time that a stochastic differential equation the has been used to describe the time evolution of individual mesoscale trajectories obeying ergodic adiabatic Hamiltonian dynamics.

A notable omission from our work in Chaps. 3 and 4 is any mention of non-equilibrium steady states, but we suspect that allowing for multiple baths in the thermodynamic limit may rectify this deficiency. The presence of two fixed temperature baths should allow for a continual energy flow through the system of interest from the warmer bath to the colder bath, but deriving such effects from the chaos bath framework will likely require a subtle handling of limits, and may require the introduction of additional separated time scales. This is a topic for future research.

In Chap. 5, we employed mesoscopic stochastic thermodynamics to study the energetic costs of using an external feedback controller to stabilize the diffusive motion of an overdamped Brownian particle obeying Langevin dynamics. Our feedback control scheme

was conceptualized as a control demon - a hypothetical being which continuously makes possibly noisy observations of the Brownian particle's position, continuously estimates the actual position of the particle, and continuously exerts corresponding feedback forces. We showed that, as a somewhat trivial consequence of the first law of thermodynamics and the notion of thermal equilibrium, a demon which makes perfect observations is able stabilize the Brownian particle without doing any work on average. Paradoxically, however, we also showed that a demon which makes noisy observations, even when employing a Kalman-Bucy filter to optimally estimate the Brownian particle's position, must continuously do work on average in order to achieve stabilization, where the average work rate is independent of the observation noise strength and does not vanish in the limit of perfect observations. Although the overdamped Langevin equation is a frequently used model in stochastic thermodynamics which has been referred to as "a paradigm for the field" [13], to the best of our knowledge, the paradox presented in Chap. 5 is an original new result. In order to resolve this paradox, we had to recognize that the fluctuating thermal force, which is modelled as an uncorrelated Gaussian white noise process in the overdamped Langevin equation, is ultimately deterministic in origin and must therefore be correlated over short (possibly immeasurably short) time scales which are much smaller than the Brownian particle's natural diffusion time scale. We showed that in order to stabilize while doing work at a vanishingly small rate on average, the demon's error time scale (which is proportional to the observation noise strength) must be much, much smaller than the thermal force's correlation time scale, and that the paradox arises because the perfect observation limit and the uncorrelated noise limit do not commute - this is also a new and original result.

In future work, we would like to employ control demon set-up to the problem of stabilizing a Brownian particle modelled by an underdamped Langevin equation. The underdamped Langevin description of Brownian motion lies somewhere between the microscopic and mesoscopic descriptions, and the white noise which enters the equation of motion for the particle's momentum results in a correlated Ornstein-Uhlenbeck noise entering the equation of motion for the particle's position. Because these correlations are inherent to the underdamped model, it is not clear whether or not a paradox analogous to that of Chap. 5 and the corresponding to need endow the thermal force with even finer correlations will arise.

Considering all of the work presented in this thesis, we conclude that the microscopic foundations of stochastic processes in physics can be equal in importance to the applications of stochastic processes in physics. Just as stochastic thermodynamics yields insights and results by relating the physics of the macroscale to the physics of the mesoscale, our work yields insights and results by relating the physics of the mesoscale to the physics of the microscale. Overall, an interesting next step in our research will be to develop the micro-meso connection in a framework which can describe feedback control in thermodynamic systems at both the microscale and mesoscale. Such an avenue of research will likely present significant mathematical challenges, and will involve combining and extending the ideas presented in Chaps. 3, 4, and 5. If such a framework can be established, we suspect that interesting new results related to the physical Hamiltonian limitations of controlling thermodynamic systems will follow.

Appendix A

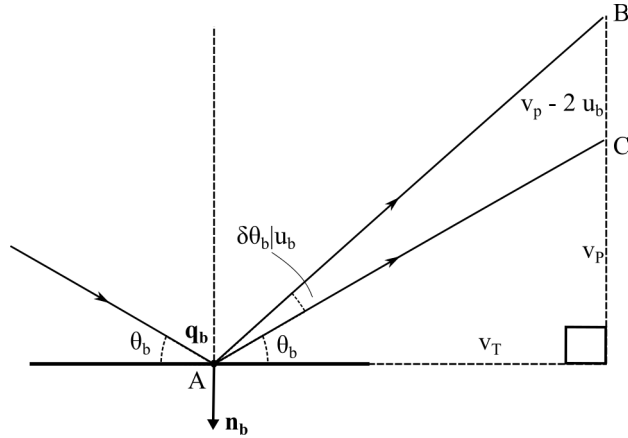# Collision location perturbation

Figure A.1: Incoming and outgoing particle trajectories at the bth collision location $\mathbf{q}_b$ in the full and frozen dynamics, assuming a collision wall velocity u_b. The full dynamics trajectory is perturbed by an angle $\delta\theta|u_b$ relative to the frozen dynamics trajectory. $\mathbf{n}_b$ is the outward normal to the boundary at $\mathbf{q}_b$.

Here, we find $\|\delta\mathbf{q}_{b+1}|u_b\|$, the magnitude of the perturbation to the frozen dynamics $(b+1)^{th}$ collision location due to the energy gained or lost at the bth collision in the full dynamics. In the frozen dynamics, the collision angle θ_b is equal to the angle of reflection. Let $\theta_b + \delta\theta|u_b$ be the reflected angle in the full dynamics, assuming a wall velocity of u_b at the bth collision. We denote v_{b-1} as the incoming particle speed at the bth collision, v_T as the velocity component tangent to the wall, v_P as the reflected particle's velocity component perpendicular to the wall in the frozen dynamics, and $v_P|u_b$ as the reflected perpendicular
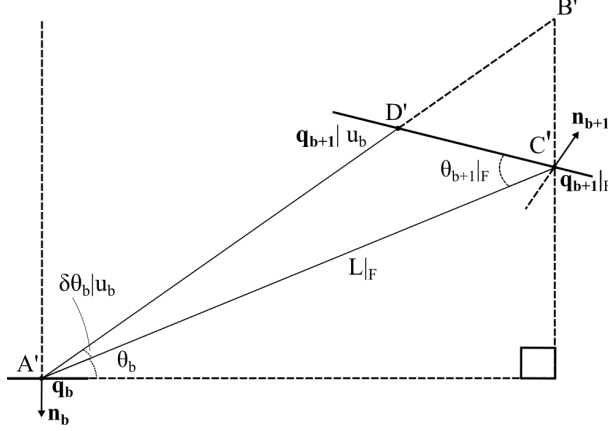
185

Figure A.2: The geometrical relationship between $\mathbf{q}_b$, $\mathbf{q}_{b+1}|_F$, and $\mathbf{q}_{b+1}|_{u_b}$. $\mathbf{q}_b$ and $\mathbf{q}_{b+1}|_F$ denote the bth and $(b+1)^{th}$ collision locations in the frozen dynamics, respectively, while $\mathbf{q}_{b+1}|_{u_b}$ denotes the $(b+1)^{th}$ collision location in the full dynamics. $\mathbf{n}_b$ and $\mathbf{n}_{b+1}$ are the outward normals to the boundary at $\mathbf{q}_b$ and $\mathbf{q}_{b+1}|_F$, respectively.

velocity component in the full dynamics. The collision kinematics give $v_P|_{u_b} = v_P - 2u_b$. The perturbation $\delta\theta|_{u_b}$ can be found using the geometry in Fig. A.1. Note that $\tan(\theta_b) = \frac{v_P}{v_T}$ and $\tan(\theta_b + \delta\theta|_{u_b}) = \frac{v_P|_{u_b}}{v_T}$. Expanding $\tan(\theta_b + \delta\theta|_{u_b})$ to first order in $\delta\theta|_{u_b}$, we find

$$
\begin{aligned}
\tan(\theta_b + \delta\theta|_{u_b}) &= \frac{v_P|_{u_b}}{v_T} \\
&= \tan(\theta_b) + \frac{1}{\cos^2(\theta_b)}\delta\theta|_{u_b} \\
&= \frac{v_P}{v_T} + \frac{1}{\cos^2(\theta_b)}\delta\theta|_{u_b}.
\end{aligned}
\tag{A.1}
$$

Noting that $v_P|_{u_b} = v_P - 2u_b$ and $v_T = v_{b-1}\cos\theta_b$, we solve for $\delta\theta|_{u_b}$ to find

$$
\delta\theta|_{u_b} = 2\cos(\theta_b)\frac{u_b}{v_{b-1}}.
\tag{A.2}
$$

Figure A.2 shows the geometry of the bth and $(b+1)^{th}$ collisions in both the full and frozen dynamics, where $\|\delta\mathbf{q}_{b+1}|_{u_b}\|$ is the length of the line segment $C'D'$. We assume that $\delta\theta|_{u_b}$ is small enough such that the wall appears flat between the frozen and full dynamics' $(b+1)^{th}$ collision locations. The triangle $A'B'C'$ in Fig. A.2 is similar to the triangle ABC in Fig. A.1, so we have $\frac{|BC|}{|AC|} = \frac{|B'C'|}{|A'C'|} = \frac{2|u_b|}{v_{b-1}}$. We note that $|A'C'|$ is the distance between the bth and $(b+1)^{th}$ collision locations in the frozen dynamics, so we denote $|A'C'| = L_b|_F$

186

and find

$$|B'C'| = \frac{2|u_b|}{v_{b-1}} L_b|_F.$$

(A.3)

All angles in Fig. A.1 can be found in terms of θ_b, $\theta_{b+1}|_F$, and $\delta\theta|u_b$. By applying the Law of Sines to the triangle $B'C'D'$, we find

$$|C'D'| = 2L_b|_F \frac{\cos(\theta_b)}{\sin(\theta_{b+1}|_F)} \frac{|u_b|}{v_{b-1}}.$$

(A.4)

We thus have

$$\|\delta\mathbf{q}_{b+1}|u_b\| = 2L_b|_F \frac{\cos(\theta_b)}{\sin(\theta_{b+1}|_F)} \frac{|u_b|}{v_{b-1}}.$$

(A.5)

Appendix B

System and estimate variance

Here, we calculate the expressions for $\sigma_x^2(t)$ and $\sigma_{\hat{x}}^2(t)$ given in Eq. (5.47). First, denote by $r(t)$ the quantity

$$r(t) = \int_0^t \frac{\mathrm{d}t'}{\tau} e^{-\frac{(t-t')}{\tau}} \epsilon(t'), \tag{B.1}$$

with $\epsilon(t)$ given in Eq. (5.25). Note that $r(t)$ obeys the following differential equation:

$$\dot{r}(t) = \frac{\epsilon(t)}{\tau} - \frac{r(t)}{\tau}, \ r(0) = 0. \tag{B.2}$$

Using the above equation, and the fact that $\frac{\mathrm{d}r^2(t)}{\mathrm{d}t} = 2r(t)\dot{r}(t)$, we find

$$r^2(t) = 2\int_0^t \frac{\mathrm{d}t'}{\tau} e^{-\frac{2(t-t')}{\tau}} r(t')\epsilon(t') \tag{B.3}$$

Note that the particle trajectory in Eq. (5.38) can be written as

$$x(t) = x_0 e^{-\frac{t}{\tau}} + r(t) + \int_0^t \mathrm{d}t' \, e^{-\frac{(t-t')}{\tau}} \sqrt{\frac{2k_B T}{b}} \xi(t'). \tag{B.4}$$

Define the trajectory $\widetilde{x}(t)$ by

$$
\begin{aligned}
\widetilde{x}(t) &= x(t) - r(t) \\
&= x_0 e^{-\frac{t}{\tau}} + \int_0^t \mathrm{d}t'\, e^{-\frac{(t-t')}{\tau}} \sqrt{\frac{2k_B T}{b}} \xi(t').
\end{aligned}
\tag{B.5}
$$

This trajectory represents the path the particle would have taken, given the same initial condition and realization of $\xi(t)$, had the demon been able to make perfect observations. Using the orthogonality relation in Eq. (5.41) along with the definition of $\widetilde{x}(t)$, we find

$$
\langle r(t)\epsilon(t)\rangle = \mathfrak{E}^2(t) - \langle \widetilde{x}(t)\epsilon(t)\rangle,
\tag{B.6}
$$

and the expectation of $r^2(t)$ can be written as

$$
\langle r^2(t)\rangle = 2 \int_0^t \frac{\mathrm{d}t'}{\tau} e^{-\frac{2(t-t')}{\tau}} \left[\mathfrak{E}^2(t') - \langle \widetilde{x}(t')\epsilon(t')\rangle\right]
\tag{B.7}
$$

Upon squaring Eq. (B.5) and taking expectations, we have

$$
\langle x^2(t)\rangle = \langle \widetilde{x}^2(t)\rangle + \langle r^2(t)\rangle + 2\langle \widetilde{x}(t)r(t)\rangle.
\tag{B.8}
$$

The expectation $\langle \widetilde{x}^2(t)\rangle$ can be calculated straightforwardly by using Eqs. (5.2) and (B.5):

$$
\langle \widetilde{x}^2(t)\rangle = \left(\langle x_0^2\rangle - \frac{k_B T}{k}\right) e^{-\frac{2t}{\tau}} + \frac{k_B T}{k}
\tag{B.9}
$$

Using Eqs. (B.1) and (B.7), we also have

189

$$\langle r^2(t) \rangle + 2 \langle \widetilde{x}(t) r(t) \rangle \quad = \quad 2 \int_0^t \frac{dt'}{\tau} e^{-\frac{2(t-t')}{\tau}} \mathfrak{E}^2(t') \tag{B.10}$$

$$+ \quad 2 \int_0^t \frac{dt'}{\tau} e^{-\frac{(t-t')}{\tau}} \left[\langle \widetilde{x}(t) \epsilon(t') \rangle - e^{-\frac{(t-t')}{\tau}} \langle \widetilde{x}(t') \epsilon(t') \rangle \right]$$

From Eqs. (5.25) and (B.7), it is straightforward to show the relation

$$\langle \widetilde{x}(t) \epsilon(t') \rangle \underset{t \geq t'}{=} e^{-\frac{(t-t')}{\tau}} \langle \widetilde{x}(t') \epsilon(t') \rangle, \tag{B.11}$$

which implies that the second integral in Eq. (B.10) vanishes. Taking Eqs. (B.8), (B.9), and (B.10) together, along with the equality $\langle x(t) \rangle^2 = \langle x_0 \rangle^2 e^{-\frac{2t}{\tau}}$, we have

$$\sigma_x^2(t) \quad = \quad \left(\sigma_x^2(0) - \frac{k_B T}{k} \right) e^{-\frac{2t}{\tau}} + \frac{k_b T}{k} + 2 \int_0^t \frac{dt'}{\tau} e^{-\frac{2(t-t')}{\tau}} \mathfrak{E}^2(t'). \tag{B.12}$$

The variance relation in Eq. (5.30) then implies

$$\sigma_{\hat{x}}^2(t) \quad = \quad \left(\sigma_x^2(0) - \frac{k_B T}{k} \right) e^{-\frac{2t}{\tau}} + \frac{k_b T}{k} + 2 \int_0^t \frac{dt'}{\tau} e^{-\frac{2(t-t')}{\tau}} \mathfrak{E}^2(t') - \mathfrak{E}^2(t). \tag{B.13}$$

Appendix C

Gaussian distribution

In this appendix, we show that the system state and estimate have a steady state distribution given by Eq. (5.50). The Fokker-Plank equation associated with the system defined by Eq. (5.49) is given by [9]

$$
\begin{aligned}
\partial_t P^{x\hat{x}}(t) &= -\partial_x \left[-\frac{\hat{x}}{\tau} P^{x\hat{x}}(t) \right] - \partial_{\hat{x}} \left[\left(-\frac{\hat{x}}{\tau} + \frac{x - \hat{x}}{\tau_\varepsilon} \right) P^{x\hat{x}}(t) \right] \\
&+ \partial_{xx}^2 \left[\frac{k_B T}{b} P^{x\hat{x}}(t) \right] + \partial_{\hat{x}\hat{x}}^2 \left[\frac{k_B T}{b} P^{x\hat{x}}(t) \right],
\end{aligned}
\tag{C.1}
$$

where $P^{x\hat{x}}(t)$ is the time-dependent joint probability distribution of x and $\hat{x}$. If a steady state distribution $P_s^{x\hat{x}}$ exists, then we must have

$$
\begin{aligned}
0 &= -\partial_x \left[-\frac{\hat{x}}{\tau} P_s^{x\hat{x}} \right] - \partial_{\hat{x}} \left[\left(-\frac{\hat{x}}{\tau} + \frac{x - \hat{x}}{\tau_\varepsilon} \right) P_s^{x\hat{x}} \right] \\
&+ \partial_{xx}^2 \left[\frac{k_B T}{b} P_s^{x\hat{x}} \right] + \partial_{\hat{x}\hat{x}}^2 \left[\frac{k_B T}{b} P_s^{x\hat{x}} \right],
\end{aligned}
\tag{C.2}
$$

Due to the linearity of Eq. (5.49), if a steady state exist, we expect it to be a Gaussian. From Eq. (5.46), we expect the steady state have mean zero, and from Eqs. (5.43) and (5.48), we expect the covariance matrix Σ to be given by

$$\Sigma = \begin{pmatrix} \frac{k_B T}{\tau} + \sigma\sqrt{\frac{2k_B T}{b}} & \frac{k_B T}{\tau} \\ \frac{k_B T}{\tau} & \frac{k_B T}{\tau} \end{pmatrix}. \tag{C.3}$$

We thus expect the steady state distribution to be of the form

$$P_s^{x\hat{x}} = \mathcal{N} e^{-\frac{1}{2}\begin{pmatrix} x & \hat{x} \end{pmatrix}\Sigma^{-1}\begin{pmatrix} x \\ \hat{x} \end{pmatrix}}, \tag{C.4}$$

where $\mathcal{N}$ is a normalization constant. It is a simple matter to check that Eq. (C.4) indeed satisfies Eq. (C.2). After normalizing and factoring the terms in the exponential, Eq. (C.4) can be brought into the form of Eq. (5.50).

Bibliography

[1] A. Einstein, *Investigations on the Theory of the Brownian Movement* (Dover, New York, 1956).

[2] P. Pearle, B. Collett, K. Bart, D. Bilderback, D. Newman, and S. Samuels, Am J. Phys. **78**, 1278 (2010).

[3] P. F. Gora, Foton **92 SE**, 52 (2006).

[4] I. M. Sokolov and J. Klafter, Chaos **15**, 026103 (2005).

[5] A. G. Peeters and D. Strintzi, Ann. Phys. (Berlin) **17**, 142 (2008).

[6] N. Wiener, Stud. Appl. Math **2**, 131 (1923).

[7] A. N. Kolmogorov, *Foundations of the Theory of Probability*, 2nd ed. (Chelsea Publishing Company, New York, 1956).

[8] K. Itô, Proc. Imp. Acad. **20**, 519 (1944).

[9] B. Oksendal, *Stochastic Differential Equations*, 6th ed. (Springer, Heidelberg Dordrecht London New York, 2010).

[10] S. Chandrasekhar, Rev. Mod. Phys. **15**, 1 (1943).

[11] N. G. V. Kampen, *Stochastic Processes in Physics and Chemistry*, 3rd ed. (Elsevier, Amsterdam, 2007).

[12] A. B. Kolomeisky and M. E. Fisher, Annu. Rev. Phys. Chem. **58**, 675 (2007).

[13] U. Seifert, Rep. Prog. Phys. **75**, 126001 (2012).

[14] B. Altaner, *Foundations of Stochastic Thermodynamics*, Ph.D. thesis, Georg-August-Universitt Gttingen (2014).

[15] U. Seifert, Phys. Rev. Lett. **95**, 040602 (2005).

[16] K. Sekimoto, *Stochastic Energetics*, Lect. Notes Phys., Vol. 799 (Springer, Berlin Heidelberg, 2010).

[17] G. E. Crooks, Phys. Rev. E **60**, 2721 (1999).

[18] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997).

[19] C. Jarzynski, Phys. Rev. E **56**, 5018 (1997).

[20] C. Jarzynski, Annu. Rev. Condens. Matter Phys. **2**, 329 (2010).

[21] J. C. Maxwell, *The Theory of Heat*, 3rd ed. (Longmans, Green, and Co., London, 1872).

[22] L. Szilard, Z. Phys. **53**, 840 (1929).

[23] C. Bennett, Int. J. Theor. Phys. **21**, 905 (1982).

[24] J. M. R. Parrondo, J. M. Horowitz, and T. Sagawa, Nature Phys. **11**, 131 (2015).

[25] R. Landauer, IBM J. Res. Dev. , 183 (1961).

[26] J. M. Horowitz and H. Sandberg, New J. Phys. **16**, 125007 (2014).

[27] T. Sagawa and M. Ueda, Phys. Rev. E **85**, 012204 (2012).

[28] S. Deffner and C. Jarzynski, Phys. Rev. X **3**, 041003 (2013).

[29] H. Sandberg, J. Delvenne, N. J. Newton, and S. K. Mitter, Phys. Rev. E **90**, 042119 (2014).

[30] A. C. Barato and U. Seifert, Phys. Rev. Lett. **112**, 090601 (2014).

[31] Y. Jun, M. Gravilov, and J. Bechhoefer, Phys. Rev. Lett. **113**, 190601 (2014).

[32] S. K. Mitter and N. J. Newton, J. Stat. Phys. **118**, 145 (2005).

[33] P. Gaspard and X. Wang, Phys. Rep. **235**, 291 (1993).

[34] M. H. Partovi, Phys. Rev. Lett. **89**, 144101 (2002).

[35] J. Demers and C. Jarzynski, Phys. Rev. E **95**, 042911 (2015).

[36] E. Ott, *Chaos in Dynamical Systems* (Cambridge University Press, Cambridge, England, 2002).

[37] A. Brahic, Astron. and Astrophys. **12**, 98 (1970).

[38] M. A. Lieberman and A. J. Lichtenberg, Phys. Rev. A **5**, 1852 (1972).

[39] A. J. Lichtenberg, M. A. Lieberman, and R. H. Cohen, Physica D **1**, 291 (1980).

[40] C. Jarzynski and W. J. Swiatecki, Nucl. Phys. **A552**, 1 (1993).

[41] E. Fermi, Phys. Rev. **75**, 1169 (1949).

[42] G. M. Zaslavskii and B. V. Chirikov, Sov. Phys. Dokl. **9**, 989 (1965).

[43] S. M. Ulam, in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 3 (California UniversityPress, Berkeley, 1961) p. 315.

[44] J. M. Hammersley, in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 3 (California University Press, Berkeley, 1961) p. 79.

[45] V. Gelfriech, V. Rom-Kedar, and D. Turaev, Chaos **22**, 033116 (2012).

[46] B. Batistic, Phys. Rev. E **89**, 022912 (2014).

[47] B. Batistic, Phys. Rev. E **90**, 032909 (2014).

[48] M. Wilkinson, J. Phys. A **23**, 3603 (1990).

[49] C. Jarzynski, Phys. Rev. E **48**, 4340 (1993).

[50] A. K. Karlis, F. K. Diakonos, V.Constantoudis, and P. Schmelcher, Phys. Rev. E **78**, 046213 (2008).

[51] C. Jarzynski, Phys. Rev. A **46**, 7498 (1992).

[52] P. Revesz, *The Laws of Large Numbers* (Academic Press, 1968).

[53] A. V. Kargovsky, E. I. Anashkina, O. A. Chichigina, and A. K. Krasnova, Phys. Rev. E **87**, 042133 (2013).

[54] A. Loskutov, A. B. Ryabov, and L. G. Akinshin, J. Phys A **33**, 7973 (2000).

[55] S. O. Khamporst and S. P. de Carvalho, Nonlinearity **12**, 1363 (1999).

[56] F. Lenz, C. Petri, F. R. N. Koch, F. K. Diakonos, and P. Schmelcher, New J. Phys. **11**, 083035 (2011).

[57] K. Shah, Phys. Rev. E **83**, 046215 (2011).

[58] K. Shah, D. Turaev, and V. Rom-Kedar, Phys. Rev. E **81**, 056205 (2010).

[59] V. Gelfreich, V. Rom-Kedar, K. Shah, and D. Turaev, Phys. Rev. Lett. **106**, 074101 (2011).

[60] V. Gelfreich, V. Rom-Kedar, and D. Turaev, J. Phys. A **47**, 395101 (2014).

[61] A. Loskutov, A. B. Ryabov, and L. G. Akinshin, J. Exp. Theor. Phys. **89**, 966 (1999).

[62] E. D. Leonel, P. V. E. McClintock, and J. K. da Silva, Phys. Rev. Lett. **93**, 014101 (2004).

[63] A. K. Karlis, P. K. Papachristou, F. K. Diakonos, V. Constantoudis, and P. Schmelcher, Phys. Rev. Lett. **97**, 194102 (2006).

[64] A. K. Karlis, P. K. Papachristou, F. K. Diakonos, V. Constantoudis, and P. Schmelcher, Phys. Rev. E **76**, 016214 (2007).

[65] J. Dorfman, *An Introduction to Chaos in Nonequilibrium Statistical Mechanics*, Cambridge Lecture Notes in Physics (Cambridge University Press, Cambridge, 1999).

[66] P. Gaspard, *Chaos, Scattering, and Statistical Mechanics* (Cambridge University Press, Cambridge, 1998).

[67] A. I. Kinchin, *Mathematical Foundations of Statistical Mechanics* (Dover, New York, 1949).

[68] C. Jarzynski, Phys. Rev. Lett. **74**, 2937 (1995).

[69] Y. A. Sinai, Russ. Math Surveys **25**, 137 (1970).

[70] R. K. Pathria and P. D. Beale, *Statistical Mechanics*, 3rd ed. (Elsevier, Oxford, 2011).

[71] N. Chernov, J. Stat. Phys **88**, 1 (1996).

[72] J. L. Lebowitz and E. Rubin, Phys. Rev. **131**, 2381 (1963).

[73] P. Resibois and H. T. Davis, Physica **30**, 1077 (1964).

[74] C. Maes and H. Tasaki, Lett. Math Phys. **79**, 251 (2007).

[75] E. Ott, Phys. Rev. Lett **42**, 1628 (1979).

[76] M. V. Berry and J. M. Robbins, Proc. R. Soc. Lond. A **436**, 631 (1992).

[77] M. V. Berry and J. M. Robbins, Proc. R. Soc. Lond. A **442**, 659 (1993).

[78] M. V. Berry and J. M. Robbins, Proc. R. Soc. Lond. A **442**, 641 (1993).

[79] M. V. Berry, Proc. R. Soc. Lond. A **392**, 45 (1984).

[80] V. I. Arnaold, *Mathematical Methods of Classical Mechanics*, 2nd ed., Graduate Texts in Mathematics, Vol. 60 (Springer, New York, 1989).

[81] S. Sasa and Y. Yokokura, Phys. Rev. Lett **116**, 140601 (2106).

[82] M. san Miguel and J. M. Sancho, J. Stat. Phys. **22**, 605 (1980).

[83] R. Kubo, Rep. Prog. Phys. **29**, 255 (1966).

[84] T. Cover and J. Thomas, *Elements of Information Theory*, 2nd ed. (Wiley, Hoboken, 2006).

[85] R. L. Stratonovich, J. SIAM Control **4**, 362 (1966).

[86] H. J. Sussmann, Annal. of Prob. **6**, 19 (1978).

[87] F. Schwabl, *Quantum Mechanics*, 4th ed. (Springer, Berlin Heidelberg New York, 2007).

[88] C. Jarzynski, C. R. Physique **8**, 495 (2007).

[89] P. Arnold, Phys. Rev. E **61**, 6099 (2000).

[90] H. Calisto and E. Tirapegui, Phys. Rev. E **65**, 038101 (2002).

[91] M. Chaichian and A. Demichev, *Path Integrals in Physics Volume I: Stochastic Processes and Quantum Mechanics*, Series in Mathematical and Computational Physics (Institute of Physics Publishing, Bristol and Philadelphia, 2001).

[92] B. Cleuren, C. van den Broeck, and R. Kawai, Phys. Rev. Lett. **96**, 050601 (2006).

[93] R. Brockett, *Finite Dimensional Linear Systems* (John Wiley and Sons, New York, 1970).

[94] J. Bechhoefer, Rev. Mod. Phys. **77**, 783 (2005).

[95] T. R. Yechout, S. L. Morris, D. E. Bossert, W. F. Hallgren, and J. K. Hall, *Introduction to Aircraft Flight Mechanics: Preformance, Static Stability, Dynamic Stability, Feedback Control, and State Space Foundadtions*, 2nd ed., AIAA Education (American Institute of Aeronatics and Astronautics, Reston, VA, 2014).

[96] K. H. Kim and H. Qian, Phys. Rev. E **75**, 022102 (2007).

[97] T. Munakata and M. L. Rosinberg, J. Stat. Mech. **2012**, P05018 (2012).

[98] T. Munakata and M. L. Rosinberg, J. Stat. Mech. **2013**, P06014 (2013).

[99] Y. Jun and J. Bechhoefer, Phys. Rev. E. **86**, 061106 (2012).

[100] M. Gravrilov, Y. Jun, and J. Bechhoefer, Rev. Sci. Instrum. **85**, 095102 (2014).

[101] A. E. Cohen, Phys. Rev. Lett. **94**, 118102 (2005).

[102] T. Li, S. Kheifets, D. Medellin, and M. G. Raizen, Science **328**, 1673 (2010).

[103] S. Kheifets, A. Simha, K. Melin, T. Li, and M. G. Raizen, Science **343**, 1493 (2014).

[104] J. M. Sancho, M. san Miguel, and D. Dürr, J. Stat. Phys **28**, 291 (1982).

[105] A. E. Cohen and W. E. Moerner, Appl. Phys. Lett. **86**, 093109 (2005).

[106] E. Kestemont, C. van den Broeck, and M. M. Mansour, EPL **49**, 143 (2000).

[107] M. Buchanan, Nature Phys. **1**, 71 (2005).