The Edge of Thermodynamics: Driven Steady States in Physics and Biology

by

Robert Alvin Marsland III

Submitted to the Department of Physics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2017

Author . . . .
Department of Physics
May 24, 2017

Certified by . . . .
Jeremy L. England
Cabot Career Development Associate Professor of Physics
Thesis Supervisor

Accepted by . . . .
Nergis Mavalvala
Curtis and Kathleen Marble Professor of Astrophysics
Associate Department Head of Physics

The Edge of Thermodynamics: Driven Steady States in Physics and Biology

by

Robert Alvin Marsland III

Submitted to the Department of Physics
on May 24, 2017, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

From its inception, statistical mechanics has aspired to become the link between biology and physics. But classical statistical mechanics dealt primarily with systems in thermal equilibrium, where detailed balance forbids the directed motion characteristic of living things. Formal variational principles have recently been discovered for nonequilibrium systems that characterize their steady-state properties in terms of generalized thermodynamic quantities. Concrete computations using these principles can usually only be carried out in certain limiting regimes, including the near-equilibrium regime of linear response theory. But the general results provide a solid starting point for defining these regimes, demarcating the extent to which system's behavior can be understood in thermodynamic terms.

I use these new results to determine the range of validity of a variational procedure for predicting the properties of near-equilibrium steady states, illustrating my conclusions with a simulation of a sheared Brownian colloid. The variational principle provides a good prediction of the average shear stress at arbitrarily high shear rates, correctly capturing the phenomenon of shear thinning. I then present the findings of an experimental collaboration, involving a specific example of a nonequilibrium structure used by living cells in the process of endocytosis. I first describe the mathematical model I developed to infer concentrations of signaling molecules that control the state of this structure from existing microscopy data. Then I show how I performed the inference, with special attention to the quantification of uncertainty, accounting for the possibility of "sloppy modes" in the high-dimensional parameter space. In the final chapter I identify a trade-off between the strength of this kind of structure and its speed of recovery from perturbations, and show how nonequilibrium driving forces can accelerate the dynamics without sacrificing mechanical integrity.

Thesis Supervisor: Jeremy L. England
Title: Cabot Career Development Associate Professor of Physics

Acknowledgments

This interdisciplinary work would not have been possible without the generous collaboration of a large number of people with a wide range of expertise. First of all, I would like to thank Prof. Jeremy England for his mentorship over these past five years, and for nurturing such a vibrant intellectual environment in our lab. It has been wonderful to see the group grow and mature almost from its very beginnings, and I owe a great deal of my own personal growth during this time to its rich atmosphere of genuine friendship and serious scholarship. I am also grateful for the hard work of the administrative staff – especially to Catherine Modica, who has been so helpful and encouraging from the beginning of my degree to the end.

I need to express my deep appreciation for the hospitality of Prof. Tomas Kirchhausen and his group, who have patiently brought me up to speed on some of the most exciting topics in cell biology, and transformed me from a physicist curious about biology into a biophysicist. Special thanks go to Kangmin He, who welcomed me into his core postdoctoral project, and performed all the experiments that form the basis of my fourth chapter. I also had many valuable conversations with Ilja Kusters, Benjamin Capraro, Gokul Upadhyayula, and Joe Sarkis about scientific issues related to this project. I want to thank Catherine McDonald as well, for always being ready to help with any logistical issue.

I am grateful to Tal Kachman for introducing me to Python and the iPython Notebook. The numerical work and data analysis in Chapters 3 - 5 would have been much more challenging and time-consuming without these helpful tools. For the simulations of Chapter 3, I want to acknowledge the contribution of Benjamin Harpt, an undergraduate research assistant who helped me finish that project and start exploring some possible extensions. The model of Chapter 5 matured through discussions with another undergraduate researcher, Arsen Vasilyan, and with postodoc Sumantra Sarkar.

Finally, I want to thank my friends who have generously given me feedback on this manuscript and related papers. Jordan Horowitz, Todd Gingrich, Gili Bisker and

Zachary Slepian have each dedicated time to reading and commenting on my drafts, and have dramatically improved the quality of the final product.

Contents

Chapter 1

Introduction

> The general struggle for existence of animate beings is therefore not a struggle for raw materials - these, for organisms, are air, water and soil, all abundantly available - nor for energy which exists in plenty in any body in the form of heat (albeit unfortunately not transformable), but a struggle for entropy, which becomes available through the transition of energy from the hot sun to the cold earth.
>
> – Ludwig Boltzmann, 1886 [10, p.24]

From the standpoint of statistical mechanics, as Boltzmann points out in this quotation, the characteristic activities of living things are part of the process by which the universe tends towards a state of maximum entropy. Schrödinger came back to this point in his lecture series *What is Life?* [89], and since then it has remained a constant theme in biophysics.

Both Boltzmann and Schrödinger place the connection to statistical mechanics in a "no-go theorem": the activities of living things are impossible in thermal equilibrium, and necessarily depend on harnessing a pre-existing disequilibrium like the temperature difference between earth and sun. This is simply an extension of the most basic statement of the Second Law of Thermodynamics, which implies that engines can only run in an environment that has not yet reached its maximum entropy. Most successful applications of statistical physics to biology are quantitative elaborations on this claim, investigating how much entropy is actually generated in particular biological activities such as sensing, DNA replication and reproduction [59, 73, 72, 58, 85, 76, 82, 23].

But classical statistical mechanics is also *predictive*. Given a closed physical system that has been left alone for a sufficiently long time, we can predict the most likely value of any observable quantity by maximizing the system's entropy. Living things are clearly not in a state of maximum entropy, since they exist precisely as processes *on the way* to maximum entropy. But perhaps there is some other quantity that is maximized or minimized in a living system's most likely state, and can be estimated from knowledge of the organism's constituent parts. This hope has generated and continues to generate considerable excitement (cf. [80]). If such a principle could be found, it would go a long way to explaining the origin of life on earth, and would pave the way for a predictive physical theory of biology.

Advances in the theory of stochastic processes and in the foundations of thermodynamics have made it possible to write down formal variational principles characterizing the most probable state of any observable in a nonequilibrium system, and to express these principles in terms of generalized thermodynamic quantities [19, 36, 7]. But the quantities involved in these exact results lack any direct empirical meaning, and can in general be computed only if the equations of motion have already been solved [86, 7].

The real value of these theoretical advances lies in their power to clarify the limits of thermodynamic reasoning [68]. For driven systems near equilibrium, considerations of work, energy and entropy lead to definite predictions of wide applicability such as the Einstein relation and more general fluctuation-dissipation theorems [21, 66]. These theorems are usually derived in the limit of vanishing nonequilibrium driving force, and the factors that determine their range of validity at finite driving force remain only vaguely specified. The new exact results for arbitrary driving now provide the necessary tools for defining this boundary. Knowing the boundary will not only steer us away from dead ends in our pursuit of biological understanding, but also help us mine all of the relevant insight from "near-equilibrium" results.

This thesis is a tour of the edge of thermodynamics. It is organized around two poles: a general theoretical result I derived concerning the range of linear response theory, and a set of experimental results I obtained in collaboration with the Kirchhausen Laboratory at Harvard Medical School. In the final chapter, I bring these poles together by investigating the thermodynamic properties of the experimental system through the lens of a simplified stochastic model.

I begin Chapter 2 with a simple example: a piston of gas kept out of equilibrium by an oscillatory driving force. I use this example to illustrate a result from linear response theory, which shows how the probability of a fluctuation is controlled by its free energy minus the average work done by a nonequilibrium driving force on the way there. I then derive a general fluctuation theory for nonequilibrium steady states in terms of work statistics. Finally, I construct a novel perturbative analysis of this general formula that contains the linear response result as a limiting case, and provides a framework for determining precisely where and why it breaks down.

As becomes clear in the course of the derivation, the distinction between my analysis and the traditional linear response approach appears when the relaxation time of the system to its stationary state depends on the strength of the driving force. In my driven piston example, it is hard to identify a simple mechanism by which the amplitude of the pressure fluctuations could affect the rate of relaxation. But in many physical systems – especially systems of interacting particles – driving forces can "stir" the particle configuration and thereby accelerate the relaxation dynamics. In Chapter 3, I apply the theory of Chapter 2 to a concrete example of such a system, showing how the nonlinear response remains well-described by my near-equilibrium approximation.

Chapter 4 contains the results of my experimental collaboration. I start by introducing the biological process I studied, focusing on the features of particular physical interest. This system involves a self-assembled structure that is switched "on" and "off" by modulation of the coupling to a nonequilibrium environment. The structure has to be switched off at a certain stage in the process, and for the past decade my collaborators have been advancing an interesting hypothesis about how this timing is regulated. I translated this hypothesis into a quantitative model, and combined the model with some new data from their laboratory to assess its physical plausibility.

In Chapter 5, I use a model inspired by the biological system of Chapter 4 to illustrate the novel material properties that become possible in a nonequilibrium steady state. The presence of an "active" disassembly pathway powered by an externally supplied free energy source allows mechanically robust structures to rapidly recover from large perturbations, something difficult to achieve in a passive material.

The thesis finishes with Chapter 6, where I sum up the conclusions of these four chapters, and point out what I believe to be the most promising future directions.

Chapter 2

The Edge of Linear Response

2.1 Introduction: A Driven Ideal Gas

Consider the cylinder depicted in Figure 2-1. It is filled with a dilute gas of N particles, and immersed in an environment of uniform temperature T and pressure $P_{\text{ext}} = P_0$. One wall of the cylinder is formed by a movable piston, so that the volume is free to vary, and eventually relaxes to its equilibrium value $V = V_0$.

In thermal equilibrium, one can compute V_0 in terms of the other parameters by maximizing the entropy S_{tot} of the whole setup – including the environment – which is equal to minus the Gibbs free energy G (up to an additive constant). At fixed temperature, the internal energy of an ideal gas is constant, as is the contribution to the system entropy from the momentum degrees of freedom, so G is determined by the configurational entropy $Nk_B \ln V$:

$$G = E - TS + PV = -k_B T N \ln V + P_0 V + g(T) \tag{2.1}$$

where $g(T)$ is a function independent of V.

The Gibbs free energy is extremized at

$$0 = \frac{\partial G}{\partial V} = -\frac{k_B T N}{V_0} + P_0, \tag{2.2}$$

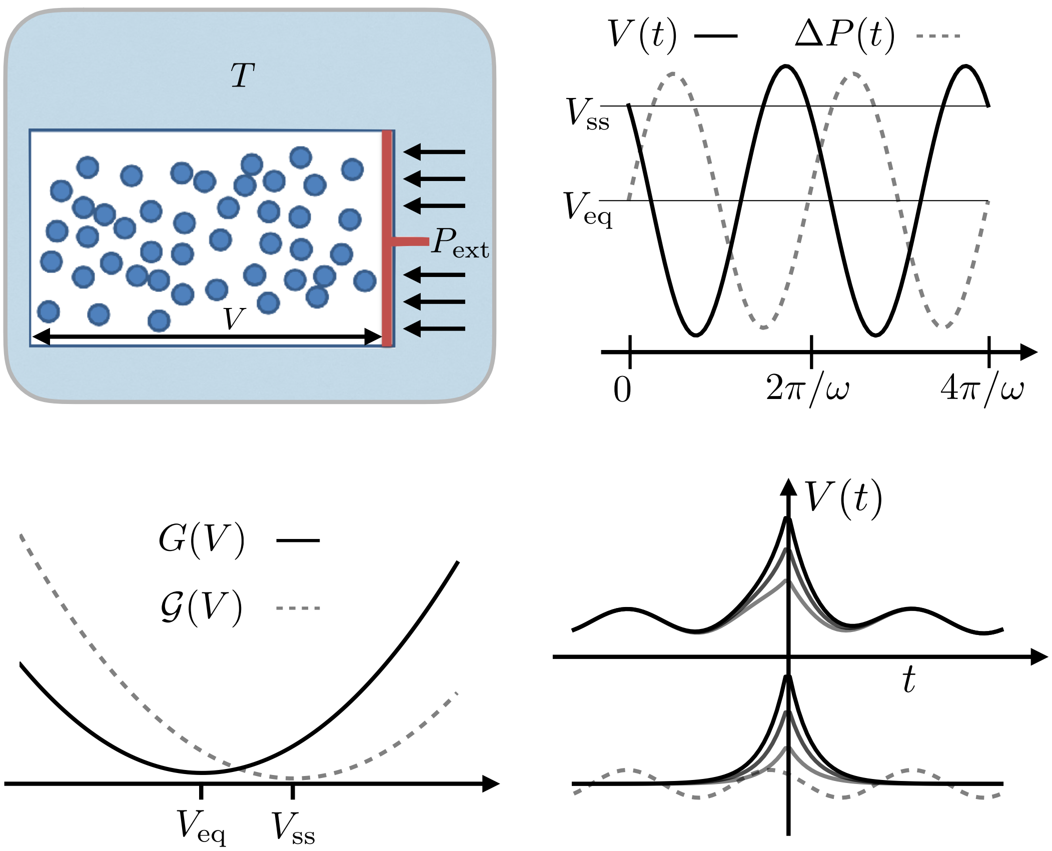

Figure 2-1: Color. Top left: Cylinder filled with N non-interacting particles. The volume V available to the gas varies in response to changes in an externally controlled pressure P_{ext} applied to the movable piston on the right side of the container. The cylinder is immersed in an equilibrated environment of temperature T and pressure $P_{\mathrm{ext}} = P_0$. Top right: Adding a small periodic perturbation $\Delta P(t) = \Delta P_0 \sin(\omega t)$ to the original constant pressure $P_{\mathrm{ext}} = P_0$ drives the system away from its equilibrium volume $V_{\mathrm{eq}} = V_0$. I define a "steady state" by measuring the volume at discrete time intervals separated by the driving period $2\pi/\omega$. Bottom left: The equilibrium volume V_{eq} can be found by minimizing the Gibbs free energy $G(V)$. The steady-state volume V_{ss} minimizes $\mathcal{G} = G - W$, where $W(V)$ is the typical work done by the periodic perturbation ΔP during the fluctuation to volume V. Bottom right: The typical trajectory to generate a given volume fluctuation away from the steady state for this linear system (top set of solid lines) is the sum of an exponential part that is independent of the drive properties (bottom set of solid lines), and a sinusoidal part that is independent of the size of the fluctuation (dotted line).

which agrees with the prediction of the ideal gas law

$$V_0 = \frac{Nk_BT}{P_0}. \tag{2.3}$$

2.1.1 Overdamped Model

The goal of this chapter is to determine the regime of validity of a generalization of this variational procedure for driven systems, which has traditionally been derived in linear response theory using the limit of vanishing driving force. I introduce the key concepts in this section by explicitly computing the steady-state volume V_{ss} in a model of a driven dilute gas, and showing that the linear response result leads to the same answer.

I drive the gas out of equilibrium by causing the external pressure P_{ext} to vary in time. The response of the volume will depend on the details of the system dynamics. I consider the case where friction between the piston and the cylinder is strong enough that the inertia of the piston is negligible, so the rate of change of volume is simply proportional to the pressure difference across the piston:

$$\frac{dV}{dt} = -\gamma \left[P_{\text{ext}}(t) - \frac{Nk_BT}{V}\right] \tag{2.4}$$

where γ is a friction coefficient that controls the timescale of relaxation.

For small enough variations of $P_{\text{ext}} = \Delta P(t) + P_0$ about P_0, the volume change $\Delta V = V - V_0$ from the equilibrium volume at P_0 will be much smaller than V_0. In this regime, the right-hand side can be approximated by the first two terms of a Taylor expansion in ΔV:

$$\begin{aligned}
\frac{d}{dt}\Delta V &= -\gamma \left[P_{\text{ext}}(t) - \frac{Nk_BT}{V_0}\left(1 - \frac{1}{V_0}\Delta V\right)\right] && (2.5)\\
&= -\gamma P_0 \left(\frac{\Delta P(t)}{P_0} + \frac{\Delta V}{V_0}\right) && (2.6)\\
&= -\frac{V_0}{\tau}\left(\frac{\Delta P(t)}{P_0} + \frac{\Delta V}{V_0}\right) && (2.7)
\end{aligned}$$

where $\tau = V_0/(P_0\gamma)$ is the relaxation timescale governing the exponential decay of ΔV to its equilibrium state at fixed ΔP near P_0.

2.1.2 Periodic Steady State

The homogeneous solution of Equation (2.7), for $\Delta P = 0$, can be read off immediately:

$$\Delta V = \Delta V_0 e^{-t/\tau} \tag{2.8}$$

For a sinusoidal driving force $\Delta P(t) = \Delta P_0 \sin \omega t$, the particular solution is

$$\Delta V(t) = -\frac{V_0}{\sqrt{1+\omega^2\tau^2}} \frac{\Delta P_0}{P_0} \sin(\omega t - \phi) \tag{2.9}$$

where the phase shift is $\phi = \tan^{-1} \omega\tau$.

The general solution for an arbitrary initial condition can be written as the sum of the particular solution and a scaled copy of the homogeneous solution:

$$\Delta V(t) = -\frac{V_0}{\sqrt{1+\omega^2\tau^2}} \frac{\Delta P_0}{P_0} \sin(\omega t - \phi) + \delta V e^{-t/\tau} \tag{2.10}$$

where the constant δV controls the initial volume.

The periodic driving prevents ΔV from relaxing to any fixed value. To define a stationary state, I observe the system stroboscopically, taking snapshots at discrete times $t_n = \frac{2\pi}{\omega} n$ with $n = 0, 1, 2, 3 \ldots$. These are the times when the pressure returns to P_0 from below. This choice implies $\sin(\omega t_n - \phi) = -\sin\phi = -\omega\tau/\sqrt{1+\omega^2\tau^2}$ for all n, so that the distance from the equilibrium volume V_0 at observation time t_n is given by:

$$\Delta V(t_n) = V_0 \frac{\omega\tau}{1+\omega^2\tau^2} \frac{\Delta P_0}{P_0} + \delta V e^{-t_n/\tau}. \tag{2.11}$$

In the limit $t_n \to \infty$, these stroboscopic measurements relax to a stationary value:

$$\Delta V_{\text{ss}} = V_0 \frac{\omega\tau}{1+\omega^2\tau^2} \frac{\Delta P_0}{P_0}. \tag{2.12}$$

This is the steady-state volume I set out to compute, given in terms of the nonequilibrium drive parameters ω and ΔP_0.

2.1.3 A Variational Principle

I will now obtain the same answer from a thermodynamic perspective.

The physical intuition for generalizing free energy minimization is most accessible in the context of fluctuation theory. While I began by considering a macroscopic system whose dynamics are effectively deterministic, a finite thermal system is necessarily subject to fluctuations. The probability $p_{\text{eq}}(V)$ of a given fluctuation away from the deterministic volume V^* in a large system at thermal equilibrium is determined by the decrease in entropy $\Delta S_{\text{tot}} = S_{\text{tot}}(V) - S_{\text{tot}}(V^*)$:

$$p_{\text{eq}}(V) \propto e^{\Delta S_{\text{tot}}(V)/k_B} \propto e^{-\Delta G(V)/k_B T} \tag{2.13}$$

where S_{tot} is the entropy of the whole setup, and $G(V)$ is the Gibbs free energy of the ideal gas, as defined in Equation (2.1). This relationship is essentially a coarse-grained version of the Boltzmann distribution. It was first identified by Einstein [22] and has since been confirmed more rigorously in the large system size limit, using the methods of large deviation theory [95].

In 1959, James McLennan showed that the correction to the Boltzmann distribution for near-equilibrium steady states is related in a simple way to the work done by the external driving forces [71, 65]. The coarse-grained version of his finding gives:

$$p_{\text{ss}}(V) \propto e^{-[\Delta G(V) - W_{\text{ex}}(V)]/k_B T} \tag{2.14}$$

where $W_{\text{ex}}(V) = W(V) - W(V^*)$ is the "excess work" done on the way to the fluctuation. For my driven piston, $W(V)$ is the work done by the pressure perturbation $\Delta P(t)$ over a trajectory of duration $\mathcal{T} \gg \tau$ that ends in a state with volume V.

The most likely volume in a near-equilibrium steady state thus minimizes the quantity

$$\mathcal{G}(V) = G(V) - W(V). \tag{2.15}$$

Evaluating the work on the way to a fluctuation $W(V)$ requires knowing the most likely trajectory that leads to the fluctuation. Detailed balance requires that the most likely fluctuation trajectories in thermal equilibrium are mirror images of the corresponding relaxation trajectories. In this linearized nonequilibrium model, the homogeneous part of the fluctuation trajectory (which is the solution in the absence of a driving force) will thus be the mirror image of the homogeneous part of the relaxation trajectory. Since only the homogeneous part of the solution depends on the initial condition ΔV, the mean work done on the way to a fluctuation is given by

$$\begin{aligned} W(\Delta V) &= \int_{-\infty}^{0} \Delta P(t) \dot{V} dt & (2.16) \\ &= \frac{1}{\tau} \int_{-\infty}^{0} \Delta P_0 \sin(\omega t) \Delta V e^{t/\tau} + \text{const.} & (2.17) \\ &= \Delta P_0 \Delta V \frac{\omega\tau}{1+\omega^2\tau^2} + \text{const.} & (2.18) \end{aligned}$$

where I have suppressed constant terms that are independent of ΔV.

The steady-state volume is now found by minimizing

$$\mathcal{G} = G - W = \frac{P_0 V_0}{2} \left(\frac{\Delta V}{V_0} \right)^2 - \Delta P_0 \Delta V \frac{\omega\tau}{1+\omega^2\tau^2} + \text{const.}, \tag{2.19}$$

where I have approximated the free energy G by the lowest-order non-vanishing term in a Taylor expansion about V_0. This new quantity is minimized when

$$0 = \frac{\partial \mathcal{G}}{\partial V} = \frac{P_0}{V_0} \Delta V - \Delta P_0 \frac{\omega\tau}{1+\omega^2\tau^2}. \tag{2.20}$$

Solving for ΔV yields the steady-state volume difference

$$\Delta V_{\text{ss}} = V_0 \frac{\omega\tau}{1+\omega^2\tau^2} \frac{\Delta P_0}{P_0}. \tag{2.21}$$

This agrees with the direct calculation of Equation (2.12), in accord with McLennan's general result (2.14).

2.1.4 Range of Validity

I performed the above derivation in the $\Delta P_0 \to 0$ limit, following the standard approach of linear response theory. But Equation (2.21) can remain valid far from this limit, depending on the size of τ relative to $1/\omega$. Figure 2-2 compares the prediction of Equation (2.21) with the actual steady-state value of ΔV (measured at observation times $t_n = \frac{2\pi}{\omega} n$) from the full nonlinear equation (2.4). At small τ, the two values show good agreement even when ΔP_0 is large enough to make the gas expand more than ten times its original volume at the top of the drive cycle and the dynamics are clearly nonlinear.

The reason for this is shown in the bottom-left panel of Figure 2-2. The state at time t_n is fully determined by the portion of the $\Delta P(t)$ protocol contained in the preceding time interval of order τ. As τ decreases, the system loses memory of its past states more quickly. When τ is sufficiently small, the nonlinear regions of the trajectory with large $\Delta P(t)/P_0$ can no longer have any effect on the state of the system at the observation times t_n (where $\Delta P = 0$), and so the linearized results can still provide an adequate prediction. If τ were somehow coupled to ΔP_0 as described in the top right panel of Figure 2-2, so that τ asymptotically decreased as $1/\Delta P_0$, the "linear response" result (2.14) would predict the full nonlinear response of ΔV_{ss} to ΔP_0 for arbitrarily large values of the forcing ΔP_0. It is hard to envision such a mechanism in the driven piston, but the sheared suspension I will analyze in Chapter 3 has this behavior naturally built in to the dynamics.

The fluctuation theory of (2.14) also points the way towards a thermodynamic version of this analysis, which can be more readily generalized. Since Equation (2.14) is exact for when the dynamics are linear, it should start to fail only when nonlinearities significantly affect the amount of work done during the time τ immediately before the end of the trajectory. Since $W_{\text{ex}}(V)$ appears in the exponent of (2.14) divided by $k_B T$, it seems that the nonlinearities become important when their contribution to

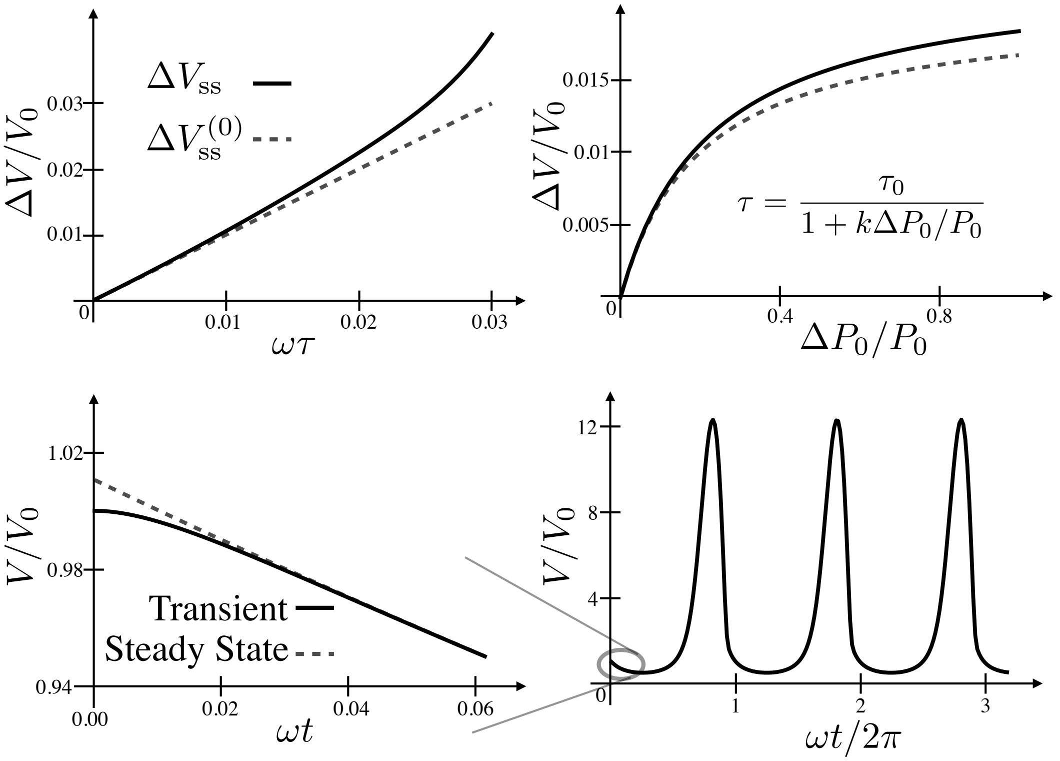

Figure 2-2: Top left: Steady-state volume difference $\Delta V_{\rm ss}$ vs. relaxation time τ for $\Delta P_0 = P_0$. The near-equilibrium prediction $V_{\rm ss}^{(0)}$ from Equation (2.21) agrees well with the exact solution $V_{\rm ss}$ up to about $\omega\tau \sim 0.01$. Top right: Nonlinear response for imaginary scenario where τ depends on ΔP_0. If τ decreases asymptotically as $1/\Delta P_0$, then the near-equilibrium result can remain valid for arbitrarily large choices of ΔP_0. Bottom right: $V(t)$ in the periodic steady state with this large ΔP_0 value and $\omega\tau = 0.01$. The sinusoidal pressure changes produce a volume response with a very different shape, due to the nonlinearity of the dynamical equation (2.4). Bottom left: Zoomed-in plot of $V(t)$ showing the relaxation dynamics to the periodic steady state. The system loses memory of its initial condition well before it leaves the region where the small-ΔV approximation is valid.

$W_{\mathrm{ex}}(V)$ during a typical fluctuation becomes comparable to $k_B T$. The limit of vanishing driving force guarantees that this contribution is small, but Figure 2-2 indicates that this assumption is by no means necessary. The factor determining the validity of "near-equilibrium" results like (2.14) appears not to be the strength of the driving, but *the strength of nonlinearities* measured in these thermodynamic terms. Over the course of this chapter, I will employ the tools of contemporary nonequilibrium statistical mechanics to confirm this conjecture for a broad class of systems.

2.2 Theoretical Framework

Almost 60 years after the publication of McLennan's result, we are finally in a position to see where it comes from, and thereby determine its full range of validity. This is primarily due to the conceptual reorganization of nonequilibrium statistical mechanics that has taken place over the past two decades, which highlights the role of time-reversal symmetry as the central physical principle of the theory [17, 47]. In this section I summarize this new way of looking at things, setting up the derivations of Section 2.4 where I obtain the conditions of validity for McLennan's variational principle. All the important claims in this background section are standard results in the stochastic thermodynamics literature, but arranged and described in an original way. My novel results are contained in Sections 2.4-2.6, where I determine the boundaries of the near-equilibrium regime with a novel expansion in the degree of nonlinearity.

2.2.1 Microscopic Reversibility

The foundation of this new point of view on nonequilibrium statistical mechanics is the insight that statistical irreversibility enters time-symmetric dynamics via the distribution over environmental initial conditions. In this subsection I will present the core equation that captures this insight (2.24), after setting up the basic concepts required to write it down.

Consider an isolated chunk of classical matter with Hamiltonian H_{tot}, whose state is described by a set of N coordinates q_i with conjugate momenta p_i. The dynamics are given by Hamilton's equations

$$\begin{aligned} \dot{q}_i &= \frac{\partial H_{\text{tot}}}{\partial p_i} \\ \dot{p}_i &= -\frac{\partial H_{\text{tot}}}{\partial q_i}. \end{aligned} \tag{2.22}$$

These equations of motion are deterministic, and symmetric under the time reversal operation $t \to -t, p_i \to -p_i$ (and also $\mathbf{B} \to -\mathbf{B}$ in the presence of a magnetic field $\mathbf{B}$). Any trajectory allowed by these equations of motion is therefore also allowed to happen in reverse. As illustrated in Figure 2-3, I will keep track of some subset of the degrees of freedom, which I will call the "system" and denote by $\mathbf{x}$, and refer to the rest of the degrees of freedom $\mathbf{y}$ as the "environment." The Hamiltonian can now be split into three parts: one that depends on the system degrees of freedom alone, one that depends on the environment alone, and one that couples the two sets together:

$$H_{\text{tot}}(\mathbf{x}, \mathbf{y}, t) = H_{\text{sys}}(\mathbf{x}, \lambda_t) + H_{\text{env}}(\mathbf{y}) + h_{\text{int}}(\mathbf{x}, \mathbf{y}) \tag{2.23}$$

Importantly, I have allowed for an explicit time-dependence in the system Hamiltonian H_{sys} (but not in the other terms) via a control parameter λ_t. This generalizes the extra external pressure $\Delta P(t)$ from my introductory example, and can be used to do work on the system.

If I now choose the initial condition $\mathbf{y}_0$ of the environment at random, the trajectory of the system degrees of freedom $\mathbf{x}$ from a given initial condition $\mathbf{x}_0$ becomes stochastic. I will write the probability that $\mathbf{x}$ takes a given trajectory from this initial condition as $p[\mathbf{x}_0^{\mathcal{T}}|\mathbf{x}_0, \lambda_0^{\mathcal{T}}]$. The symbol $\mathbf{x}_0^{\mathcal{T}}$ denotes the whole trajectory from its beginning at time $t = 0$ to some chosen end time $\mathcal{T}$. I have also explicitly included the dependence on the variation of the control parameter λ_t, using $\lambda_0^{\mathcal{T}}$ to represent its trajectory over the observation time. Rigorously defining this probability requires providing a suitable measure on the infinite-dimensional space of trajectories $\mathbf{x}_0^{\mathcal{T}}$. This is easy to do in the present context, since the trajectory $\mathbf{x}_0^{\mathcal{T}}$ from a given system initial

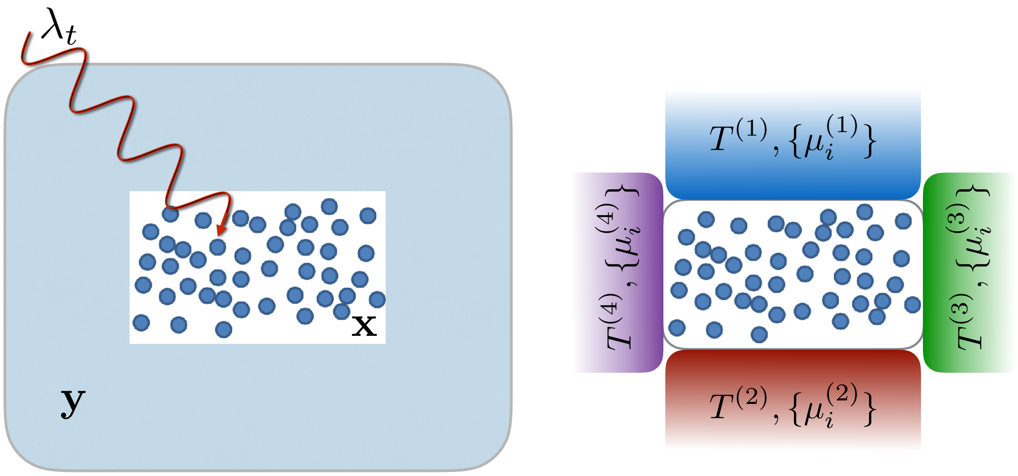
λ_t
$\mathbf{x}$
$\mathbf{y}$
$T^{(1)}, \{\mu_i^{(1)}\}$
$T^{(4)}, \{\mu_i^{(4)}\}$
$T^{(3)}, \{\mu_i^{(3)}\}$
$T^{(2)}, \{\mu_i^{(2)}\}$

Figure 2-3: Color. Left: The N degrees of freedom in a piece of isolated matter are partitioned into two sets, a system with microstate $\mathbf{x}$ and an environment with microstate $\mathbf{y}$. The system is illustrated here as discrete particles, since I keep track of the full trajectories $\mathbf{x}_0^{\mathcal{T}}$. The environment is just a solid color, because I integrate it out to obtain an effective stochastic dynamics for $\mathbf{x}$. Work can be done on the system by externally imposed variations in a set of parameters λ, which affect only the system part of the Hamiltonian H_{sys}. Right: I will focus on situations where the environment can be modeled as a set of ideal thermal and chemical reservoirs with temperatures $T^{(\alpha)}$ and chemical potentials $\mu_i^{(\alpha)}$.

condition $\mathbf{x}_0$ is fully determined by the environment state $\mathbf{y}_0$. The path measure can thus be taken as the ordinary Liouville measure $\Pi_i dp_i \, dq_i$ on the environment phase space.

This random choice of environment state can break time-reversal symmetry, introducing statistical irreversibility into the effective stochastic dynamics. To measure the extent to which this symmetry is broken, we can compare the probability $p[\mathbf{x}_0^{\mathcal{T}}|\mathbf{x}_0, \lambda_0^{\mathcal{T}}]$ with the probability of the reverse trajectory $p[\hat{\mathbf{x}}_0^{\mathcal{T}}|\mathbf{x}_{\mathcal{T}}^*, \hat{\lambda}_0^{\mathcal{T}}]$. I have introduced a new symbol $\hat{\mathbf{x}}_0^{\mathcal{T}}$ to denote the time-reversed version of $\mathbf{x}_0^{\mathcal{T}}$: the order in which the states are visited is reversed, and the signs of all the momenta are flipped. Similarly, $\hat{\lambda}_0^{\mathcal{T}}$ indicates the time-reverse of the control protocol $\lambda_0^{\mathcal{T}}$. If any magnetic fields are present, they are automatically included among the parameters λ, and their signs are reversed in the reverse protocol. An individual state whose momenta have been reversed and a single set of control parameters with magnetic fields reversed are denoted by $\mathbf{x}^*$ and λ^*, respectively.

If $\mathbf{y}_0$ is chosen from an equilibrium distribution over environment states, it turns out that the statistical irreversibility of a given system trajectory $\mathbf{x}_0^{\mathcal{T}}$ is given by the change in the entropy of the environment ΔS_e over the course of the trajectory:

$$\frac{p[\hat{\mathbf{x}}_0^{\mathcal{T}}|\mathbf{x}_{\mathcal{T}}^*, \hat{\lambda}_0^{\mathcal{T}}]}{p[\mathbf{x}_0^{\mathcal{T}}|\mathbf{x}_0, \lambda_0^{\mathcal{T}}]} = e^{-\Delta S_e/k_B}. \tag{2.24}$$

As I explain below, Equation 2.24 can be taken as a basic statement of the requirement of consistency between a time-asymmetric coarse-grained dynamics and the time-reversal symmetry of the fundamental dynamics. For this reason, it is often referred to in the literature as the “microscopic reversibility relation” [17]. It can also be viewed as a generalization of the requirement of detailed balance, and is sometimes called “local detailed balance” (cf. [91]). An analogous relation holds for quantum systems, thanks to the time-reversal symmetry of the Schrödinger equation [40].

2.2.2 Microcanonical Derivation

I will now examine the connection between Equation (2.24) and time-reversal symmetry in the Hamiltonian framework I have just described. I will deviate from standard derivations of this result, which initialize the environment in a canonical ensemble [45, 53], because this connection is clearer when the initial energy of the whole setup is fixed. In the limit of infinite environment size, my microcanonical approach gives the same answer as the older literature, confirming that the usual "equivalence of ensembles" in the thermodynamic limit remains valid in this context.

The whole setup including the environment and the system begins at time $t = 0$ with total energy E. If the system begins in state $\mathbf{x}_0$, then the energy available to the environment is $E - H_{\text{sys}}(\mathbf{x}_0, \lambda_0)$. I will denote the phase space volume occupied by environment states $\mathbf{y}$ with this energy as $\Omega(E - H_{\text{sys}}(\mathbf{x}_0, \lambda_0))$. To compute the probability $p[\mathbf{x}_0^{\mathcal{T}}|\mathbf{x}_0, \lambda_0^{\mathcal{T}}]$ of observing a given system trajectory $\mathbf{x}_0^{\mathcal{T}}$, I count how many of these allowed environment states give rise to $\mathbf{x}_0^{\mathcal{T}}$ under the equations of motion (2.22), when combined with the system initial condition $\mathbf{x}_0$ and the control protocol $\lambda_0^{\mathcal{T}}$. I denote the phase space volume occupied by these states as $\Omega[\mathbf{x}_0^{\mathcal{T}}|\mathbf{x}_0, \lambda_0^{\mathcal{T}}]$. Now I choose the initial environment state from a uniform distribution over the energetically allowed states contained in $\Omega(E - H_{\text{sys}}(\mathbf{x}_0, \lambda_0))$. As illustrated in Figure 2-4, the probability of choosing one of the states that produces the desired trajectory is:

$$p[\mathbf{x}_0^{\mathcal{T}}|\mathbf{x}_0, \lambda_0^{\mathcal{T}}] = \frac{\Omega[\mathbf{x}_0^{\mathcal{T}}|\mathbf{x}_0, \lambda_0^{\mathcal{T}}]}{\Omega(E - H_{\text{sys}}(\mathbf{x}_0, \lambda_0))}. \tag{2.25}$$

The energy of the environment changes over the course of the trajectory, as heat flows in and out of the system. At the end of the trajectory, it is equal to $E + W - H_{\text{sys}}(\mathbf{x}_{\mathcal{T}}, \lambda_{\mathcal{T}})$, where the work W done by manipulation of λ is

$$W = \int_0^{\mathcal{T}} \frac{\partial H_{\text{sys}}(\mathbf{x}_t, \lambda_t)}{\partial \lambda} \dot{\lambda}_t dt. \tag{2.26}$$

Because the control parameter has a direct effect only on the system, and not on the environment, W is fully determined by the protocol $\lambda_0^{\mathcal{T}}$ and the system trajectory $\mathbf{x}_0^{\mathcal{T}}$,

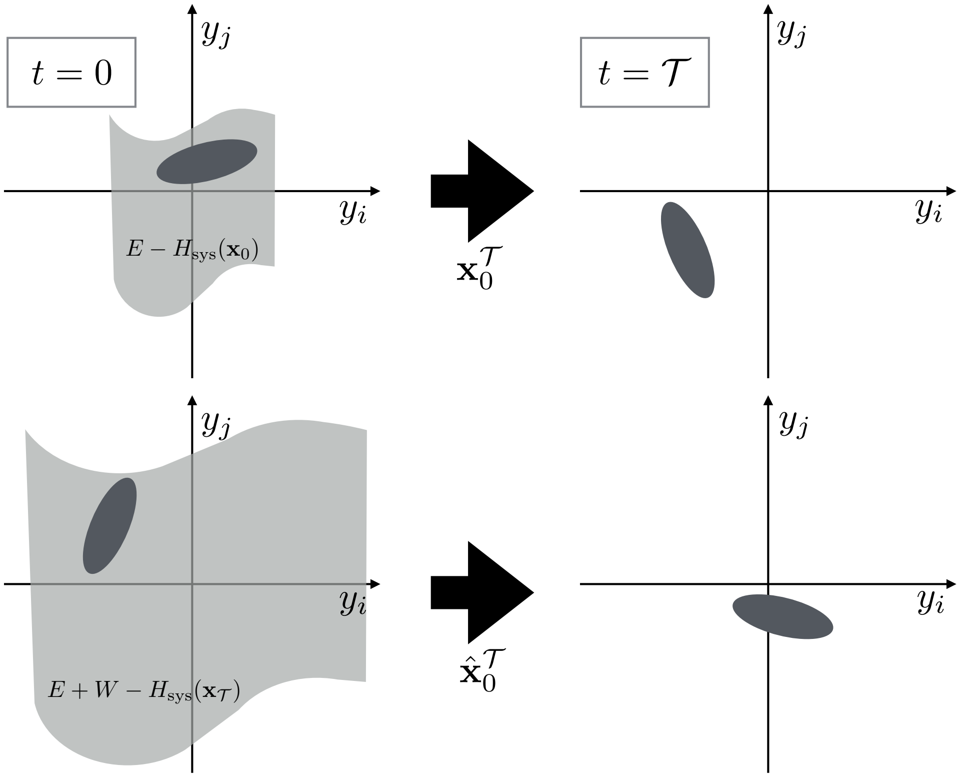
$t = 0$
y_j
y_i
$E - H_{\text{sys}}(\mathbf{x}_0)$
$\mathbf{x}_0^{\mathcal{T}}$
$t = \mathcal{T}$
y_j
y_i
y_j
y_i
$E + W - H_{\text{sys}}(\mathbf{x}_{\mathcal{T}})$
$\hat{\mathbf{x}}_0^{\mathcal{T}}$
y_j
y_i

Figure 2-4: Top: Given an initial system state $\mathbf{x}_0$, at time $t = 0$, the trajectory $\mathbf{x}_0^{\mathcal{T}}$ is determined by the choice of initial environment state $\mathbf{y}_0$. If $\mathbf{y}_0$ is chosen from a uniform distribution over a constant energy surface, the probability of choosing a state that generates a given $\mathbf{x}_0^{\mathcal{T}}$ is the ratio of the phase space volume occupied by these states (dark shaded region) to the total phase space volume available (light shaded region). Phase space volume is conserved by the Hamiltonian dynamics, and so the dark shaded region at time $t = 0$ evolves into a new region at time $t = \mathcal{T}$ with the same volume but a new energy. Bottom: Reversing the momenta of the dark shaded region from the $t = \mathcal{T}$ snapshot generates the initial conditions for the reverse trajectory $\hat{\mathbf{x}}_0^{\mathcal{T}}$. The probability of observing this trajectory is again given by the ratio of this region's volume to the volume of the entire set of environment states that share the same energy. The fact that both dark shaded regions occupy the same volume is a geometric statement of time-reversal symmetry, expressed symbolically in Equation (2.28).

regardless of what happens in the environment.

If I reverse the momenta, choose the environment conditions from this new energy surface, and run the protocol λ_t in reverse, then the probability of observing $\hat{\mathbf{x}}_0^{\mathcal{T}}$ is given by:

$$p[\hat{\mathbf{x}}_0^{\mathcal{T}}|\mathbf{x}_{\mathcal{T}}^*, \hat{\lambda}_0^{\mathcal{T}}] = \frac{\Omega[\hat{\mathbf{x}}_0^{\mathcal{T}}|\mathbf{x}_{\mathcal{T}}^*, \hat{\lambda}_0^{\mathcal{T}}]}{\Omega(E + W - H_{\text{sys}}(\mathbf{x}_{\mathcal{T}}, \lambda_{\mathcal{T}}))}. \tag{2.27}$$

Figure (2-4) shows how the numerators of Equations (2.25) and (2.27) are related by time-reversal symmetry:

$$\Omega[\hat{\mathbf{x}}_0^{\mathcal{T}}|\mathbf{x}_{\mathcal{T}}^*, \hat{\lambda}_0^{\mathcal{T}}] = \Omega[\mathbf{x}_0^{\mathcal{T}}|\mathbf{x}_0, \lambda_0^{\mathcal{T}}], \tag{2.28}$$

as long as $\Omega[\mathbf{x}_0^{\mathcal{T}}|\mathbf{x}_0, \lambda_0^{\mathcal{T}}]$ is evaluated using a measure that is invariant under the system equations of motion (which is true of the usual Liouville measure $dp\,dq$). When combined with Boltzmann's formula $S = k_B \ln \Omega$, this way of expressing the symmetry immediately leads to the desired result:

$$\frac{p[\hat{\mathbf{x}}_0^{\mathcal{T}}|\mathbf{x}_{\mathcal{T}}^*, \hat{\lambda}_0^{\mathcal{T}}]}{p[\mathbf{x}_0^{\mathcal{T}}|\mathbf{x}_0, \lambda_0^{\mathcal{T}}]} = \frac{\Omega(E - H_{\text{sys}}(\mathbf{x}_0, \lambda_0))}{\Omega(E + W - H_{\text{sys}}(\mathbf{x}_{\mathcal{T}}, \lambda_{\mathcal{T}}))} = e^{-\Delta S_e/k_B}. \tag{2.29}$$

2.2.3 Environment Entropy

The microcanonical definition of temperature $\frac{1}{T} = \beta = \frac{\partial S_e}{\partial E}$ implies that

$$\Delta S_e[\mathbf{x}_0^{\mathcal{T}}] = \beta \Delta Q[\mathbf{x}_0^{\mathcal{T}}], \tag{2.30}$$

where the heat $\Delta Q[\mathbf{x}_0^{\mathcal{T}}] = \Delta H_{\text{tot}} - \Delta H_{\text{sys}}$ is the change in the energy associated with the environment degrees of freedom $\mathbf{y}$ when the system executes the trajectory $\mathbf{x}_0^{\mathcal{T}}$.

As illustrated in Figure 2-3, more complex environments can be modeled with several ideal thermal and chemical reservoirs (indexed by α) at temperatures $T^{(\alpha)}$, and chemical potentials $\mu_j^{(\alpha)} = -T^{(\alpha)} \frac{\partial S_e^{(\alpha)}}{\partial n_j^{(\alpha)}}$, where $n_j^{(\alpha)}$ is the number of particles of type j in reservoir α. My derivation can be generalized to handle this broader class of environments if each of the reservoirs is separately initialized in a uniform distribution over states of fixed energy and particle number, and is coupled to the system at the beginning of the forward trajectory. The derivation from the canonical ensemble in [53] includes the possibility of particle exchange and multiple reservoirs from the beginning.

Now the system can be driven out of equilibrium without any time-variation in λ, carrying fluxes of energy and matter from one reservoir to another. The entropy change in the environment becomes (cf. [81]):

$$\Delta S_e(\mathbf{x}_0^{\mathcal{T}}) = \sum_\alpha \beta^{(\alpha)} \left(\Delta Q^{(\alpha)}[\mathbf{x}_0^{\mathcal{T}}] - \sum_j \mu_j^{(\alpha)} \Delta n_j^{(\alpha)}[\mathbf{x}_0^{\mathcal{T}}] \right). \tag{2.31}$$

Throughout this chapter, I will make an analogy between systems driven by external forces and those driven by thermal/chemical gradients, defining a generalized work $\mathcal{W}$ and an arbitrary reference temperature T that restore the form of the First Law for isothermal systems:

$$T\Delta S_e = \mathcal{W} - \Delta H_{\text{sys}} \tag{2.32}$$

where

$$\begin{aligned} \mathcal{W} &\equiv W + T\Delta S_e - \sum_\alpha \Delta Q^{(\alpha)} && (2.33) \\ &= W + T\sum_\alpha \left[\Delta Q^{(\alpha)}(\beta^{(\alpha)} - \beta) - \beta^{(\alpha)} \sum_j \mu_j^{(\alpha)} \Delta n_j^{(\alpha)} \right]. && (2.34) \end{aligned}$$

This allows the results I obtain in terms of work statistics to be readily generalized to cases of chemical or thermal driving.

2.2.4 Path Ensemble Averages

The microscopic reversibility relation (2.24) places a strict constraint on the averages of macroscopic observables, which can be expressed by integrating out the unobserved degrees of freedom. This results in an equality (2.40) between two averages of an arbitrary functional $\mathcal{O}[\mathbf{x}_0^{\mathcal{T}}]$ over ensembles of trajectories $\mathbf{x}_0^{\mathcal{T}}$. As first pointed out by Gavin Crooks, most of the central results of nonequilibrium statistical mechanics – including the Onsager relations, the fluctuation-dissipation theorem, and the Jarzynski Equality, in addition to McLennan's result from the previous section – can be obtained from this relation [18].

Equation (2.24) relates conditional trajectory probabilities to each other, where the initial condition is specified a priori. To obtain a path ensemble average, we also need to specify the distribution from which the initial condition is to be drawn. For now, I will simply use $p_{\text{rev}}(\mathbf{x}^*_{\mathcal{T}})$ to denote the initial conditions for the reverse trajectories, and $p_{\text{fwd}}(\mathbf{x}_0)$ for the forward trajectories. With some trivial rearranging of Equation (2.24), I find

$$p_{\text{rev}}(\mathbf{x}^*_{\mathcal{T}})p[\hat{\mathbf{x}}_0^{\mathcal{T}}|\mathbf{x}^*_{\mathcal{T}},\hat{\lambda}_0^{\mathcal{T}}] = e^{-\Delta S_e/k_B}\frac{p_{\text{rev}}(\mathbf{x}^*_{\mathcal{T}})}{p_{\text{fwd}}(\mathbf{x}_0)}p_{\text{fwd}}(\mathbf{x}_0)p[\mathbf{x}_0^{\mathcal{T}}|\mathbf{x}_0,\lambda_0^{\mathcal{T}}]. \tag{2.35}$$

Both sides now contain normalized probability distributions over the entire space of system trajectories $\mathbf{x}_0^{\mathcal{T}}$. Multiplying by the trajectory functional $\mathcal{O}[\mathbf{x}_0^{\mathcal{T}}]$, I integrate over trajectories to find:

$$\left\langle \mathcal{O}[\mathbf{x}_0^{\mathcal{T}}]\right\rangle_{\text{rev},\mathcal{T}} = \left\langle \mathcal{O}[\mathbf{x}_0^{\mathcal{T}}]e^{-\frac{\Delta S_e[\mathbf{x}_0^{\mathcal{T}}]}{k_B}+\ln\frac{p_{\text{rev}}(\mathbf{x}^*_{\mathcal{T}})}{p_{\text{fwd}}(\mathbf{x}_0)}}\right\rangle_{\text{fwd},\mathcal{T}} \tag{2.36}$$

where

$$\left\langle \mathcal{O}[\mathbf{x}_0^{\mathcal{T}}]\right\rangle_{\text{fwd},\mathcal{T}} \equiv \int \mathcal{D}[\mathbf{x}_0^{\mathcal{T}}]\mathcal{O}[\mathbf{x}_0^{\mathcal{T}}]p_{\text{fwd}}(\mathbf{x}_0)p[\mathbf{x}_0^{\mathcal{T}}|\mathbf{x}_0,\lambda_0^{\mathcal{T}}] \tag{2.37}$$

is the average of $\mathcal{O}[\mathbf{x}_0^{\mathcal{T}}]$ in the forward trajectory ensemble with initial conditions chosen from $p_{\text{fwd}}(\mathbf{x}_0)$, and

$$\left\langle \mathcal{O}[\mathbf{x}_0^{\mathcal{T}}]\right\rangle_{\text{rev},\mathcal{T}} \equiv \int \mathcal{D}[\mathbf{x}_0^{\mathcal{T}}]\mathcal{O}[\mathbf{x}_0^{\mathcal{T}}]p_{\text{rev}}(\mathbf{x}^*_{\mathcal{T}})p[\hat{\mathbf{x}}_0^{\mathcal{T}}|\mathbf{x}^*_{\mathcal{T}},\hat{\lambda}_0^{\mathcal{T}}] \tag{2.38}$$

is the average over the reverse trajectory ensemble with initial conditions chosen from $p_{\text{rev}}(\mathbf{x}^*_{\mathcal{T}})$.

A particularly important case has both the forward and reverse processes initialized in the Boltzmann distribution:

$$\ln \frac{p_{\mathrm{rev}}(\mathbf{x}^*_{\mathcal{T}})}{p_{\mathrm{fwd}}(\mathbf{x}_0)} = -\beta[H_{\mathrm{sys}}(\mathbf{x}^*_{\mathcal{T}}, \lambda_{\mathcal{T}}) - H_{\mathrm{sys}}(\mathbf{x}_0, \lambda_0) - F(\lambda_{\mathcal{T}}) + F(\lambda_0)] \tag{2.39}$$

where $F(\lambda) = -k_B T \ln \int d\mathbf{x} \exp[-\beta H_{\mathrm{sys}}(\mathbf{x}, \lambda)]$ is the free energy. I will use Equation (2.32) to write this in terms of the thermodynamic work $\mathcal{W}$ defined by Equation (2.34):

$$\left\langle \mathcal{O}[\mathbf{x}_0^{\mathcal{T}}] \right\rangle_{\mathrm{rev},\mathcal{T}} = \left\langle \mathcal{O}[\mathbf{x}_0^{\mathcal{T}}] e^{-\beta \mathcal{W}} \right\rangle_{\mathrm{fwd},\mathcal{T}} e^{\beta \Delta F}. \tag{2.40}$$

This is the fundamental expression I will work with for the rest of this chapter.

Now different choices of $\mathcal{O}[\mathbf{x}_0^{\mathcal{T}}]$ will produce different theorems. Choosing $\mathcal{O}[\mathbf{x}_0^{\mathcal{T}}] = 1$ in Equation (2.40) generates the Jarzynski equality [44]:

$$1 = \left\langle e^{-\beta \mathcal{W}} \right\rangle_{\mathrm{fwd},\mathcal{T}} e^{\beta \Delta F}. \tag{2.41}$$

Using this result, I can write Equation (2.40) in an explicitly normalized form, which will be important when I take the $\mathcal{T} \to \infty$ limit to study the steady state:

$$\left\langle \mathcal{O}[\mathbf{x}_0^{\mathcal{T}}] \right\rangle_{\mathrm{rev},\mathcal{T}} = \frac{\left\langle \mathcal{O}[\mathbf{x}_0^{\mathcal{T}}] e^{-\beta \mathcal{W}} \right\rangle_{\mathrm{fwd},\mathcal{T}}}{\left\langle e^{-\beta \mathcal{W}} \right\rangle_{\mathrm{fwd},\mathcal{T}}}. \tag{2.42}$$

2.3 Stochastic Models

Theoretical calculations based on Equations (2.24) and (2.31) are usually performed using some form of coarse-grained stochastic dynamics, rather than the Hamiltonian framework used above. In this section, I will briefly introduce two standard frameworks for stochastic modeling, illustrated in Figure 2-5, which I will make use of throughout the rest of this thesis.

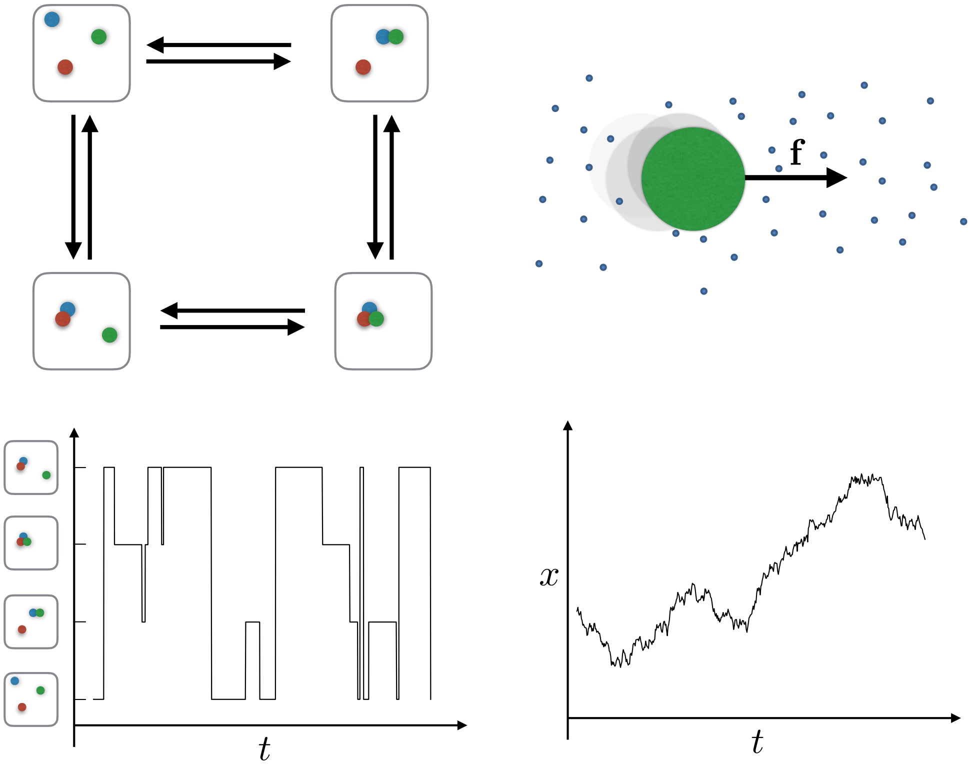

Figure 2-5: Color. Top left: A set of three particles can be found in one of four distinct states, depending on which particles are bound together. Arrows represent allowed transitions between states. Bottom left: If the transitions between discrete states are Markovian, the dynamics are described by a Markov jump process. Shown here is a trajectory for the three-particle system using equal rates for all transitions. Top right: A small particle is suspended in a solvent, and subject to a force field **f** along with a random force due to bombardment by solvent molecules. Bottom right: The random forces exerted by solvent molecules cause the particle to execute a noisy trajectory described by a Langevin equation, with a net drift in the direction of the force.

2.3.1 Markov Jump Process

Markov jump processes are a kind of stochastic dynamics exemplified by chemical reactions, as illustrated on the left side of Figure 2-5. The system evolves in a sequence of instantaneous "jumps" among discrete states $i, j, k, \ldots$, and the probability $w_{ji}dt$ that a system in state i executes a jump to state j in a given infinitesimal time window dt is independent of the system's history. Since the time required for a pair of atoms to enter or leave a bound state is typically very short compared to other relevant timescales (set by diffusion, for example), this can provide a very good model for the discrete changes in number of each kind of molecule over time in a chemical reaction, while abstracting from the quantum-mechanical nature of the transition.

The same mathematical framework can be used for any system that exhibits identifiable discrete states at some level of coarse-graining. The only requirement is a separation of time scales: the relaxation dynamics within each discrete state must be much faster than the transitions between states, so that the probability of starting a jump from a given internal configuration within the state is history-independent.

The transition rates are bound by clear thermodynamic constraints whenever the internal configuration probabilities are given by the Boltzmann distribution, and the environment consists of a set of equilibrated thermal/chemical reservoirs. A variation on the derivation of Equation (2.40) from microscopic reversibility (2.24) can be employed to show that

$$\frac{w_{ij}}{w_{ji}} = e^{-\beta(\mathcal{W}_{ji} - \Delta F_{ji})} \tag{2.43}$$

where $\mathcal{W}_{ji}$ is the generalized work done on the way from i to j, as defined in Equation (2.34), and ΔF_{ji} is the difference between the free energies of the two states. Special care must be taken when the transition from i to j can be accomplished in more than one way, as in the example of Chapter 5. Then $\mathcal{W}_{ji}$ can take on multiple possible values, depending on the pathway. In such cases, one must assign a separate rate for w_{ji}^{ρ} for each pathway ρ before imposing (2.43) [39].

2.3.2 Langevin Dynamics

The other stochastic modeling framework I will employ is inspired by Brownian motion, depicted on the right side of Figure 2-5. The equations of motion for the position $\mathbf{q}$ and momentum $\mathbf{p}$ of a Brownian particle of mass m subject to a force field $\mathbf{f}(\mathbf{q})$ are

$$\dot{\mathbf{p}} = \mathbf{f} - \frac{b}{m}\mathbf{p} + \mathbf{f}_t^{(r)} \tag{2.44}$$

$$\dot{\mathbf{q}} = \frac{1}{m}\mathbf{p} \tag{2.45}$$

where $\mathbf{f}_t^{(r)}$ is a random force with mean $\langle \mathbf{f}_t^{(r)} \rangle = 0$ due to thermal collisions with solvent molecules, and b is the particle's drag coefficient. The Langevin equation arises in the limit of infinitely fast decay of correlations in the random force, which is a good approximation for micron-scale Brownian particles being kicked by much faster moving water molecules [29]. In this limit, the random force becomes proportional to a vector of Gaussian white noise $\boldsymbol{\xi}_t$ defined by its mean and autocorrelation function:

$$\mathbf{f}_t^{(r)} = k\boldsymbol{\xi}_t \tag{2.46}$$

$$\langle \boldsymbol{\xi}_t \rangle = 0 \tag{2.47}$$

$$\langle \xi_t^i \xi_{t'}^j \rangle = \delta(t - t')\delta_{ij} \tag{2.48}$$

where ξ^i and ξ^j are elements of the vector $\boldsymbol{\xi}$, $\delta(t)$ is the Dirac delta function, δ_{ij} is the Kronecker delta and k is a scalar constant of proportionality.

This information is sufficient to compute the left-hand side of the microscopic reversibility relation (2.24) in terms of the constant k, and the right-hand side is determined by the existing definitions of work and energy. In Appendix A I use these expressions to confirm that there exists a value of k consistent with microscopic reversibility for arbitrary choices of $\mathbf{f}$. This value is $k = \sqrt{2k_BTb}$, so that

$$\mathbf{f}_t^{(r)} = \sqrt{2k_BTb}\boldsymbol{\xi}_t \tag{2.49}$$

which is a form of the famous Einstein relation [21].

At constant $\mathbf{f}$, Equation (2.44) causes the momentum $\mathbf{p}$ to lose memory of its initial condition on the time scale m/b. At times significantly longer than this, $\mathbf{p}(t)$ is sampled from its steady-state distribution with mean $\langle \mathbf{p} \rangle = m\mathbf{f}/b$ and variance mk_BT regardless of its starting state $\mathbf{p}(0)$. Since the mass m of a particle scales with its volume and the drag b with linear size, this quantity is smaller for smaller particles in the same solvent. For micron-scale particles in water, this timescale is extremely short – a few hundred nanoseconds. In such cases, it can be an excellent approximation to regard this relaxation process as instantaneous, so that the distribution over $\mathbf{p}$ is always in the steady state corresponding to the current force value, even if the force is changing in time. In this $m \to 0$ limit, the variance of the velocity $\dot{\mathbf{q}} = \mathbf{p}/m$ diverges, and the steady-state fluctuations of $\dot{\mathbf{q}}$ about its mean are themselves described by a vector of Gaussian white noise:

$$\dot{\mathbf{q}} = \frac{1}{b}\mathbf{f}(\mathbf{q}) + \sqrt{\frac{2k_BT}{b}}\boldsymbol{\xi}_t. \tag{2.50}$$

This is known as the "overdamped" Langevin equation, which will be the basis of the simulation I describe in Chapter 3.

Just as the Markov jump process is used to model many systems that are quite different from chemical reactions, so too the overdamped Langevin equation (2.50) is used as a generic phenomenological model for all kinds of continuous stochastic processes, as long as they possess the requisite separation of time scales between the "noise" and the deterministic part of the dynamics. The requirements of microscopic reversibility will vary according to the physical interpretation of the equation, and the proportionality constant k governing the strength of the random force is no longer necessarily related to the drag coefficient and temperature.

2.4 Coarse-Grained Steady-State Distribution from Forward Statistics

With this background in place, I can now turn to my original results [67]. I first obtain a generalization of the McLennan distribution (2.14) governing macroscopic fluctuations in all systems obeying microscopic reversibility (2.24), arbitrarily far from thermal equilibrium. In Section 2.5, I will obtain McLennan's result as a special case of my general expression, and investigate the conditions under which this approximation is valid.

2.4.1 General Expression

The symmetry of path ensemble averages expressed in Equation (2.42) implicitly contains a macroscopic fluctuation theory that generalizes Equation (2.13) to steady states arbitrarily far from equilibrium – whether driven by periodic variation in λ as in the piston example or by chemical or thermal gradients. Recall that the "steady state" of a periodically driven system is defined by making observations at integer multiples of the drive period, as described in the explanation of Equation (2.12) in my initial example. This means that λ has a fixed value for all the time points of interest, and I will suppress the explicit dependence on λ for the remaining derivations.

To obtain the macroscopic fluctuation theory for these driven steady states, I first partition the system phase space based on some observable properties. In the example of Section (2.1), the property is the position of the piston, which sets the volume of the cylinder. This property can be used to carve up the microscopic phase space (including the positions and velocities of the piston and all the particles) into discrete regions, such that a point $\mathbf{x}$ falls in region $\mathbf{X}$ if the volume of the cylinder is within some margin δV of a specified volume $V_{\mathbf{X}}$.

Now I use this partition of phase space to define a trajectory observable $\mathcal{O}[\mathbf{x}_0^{\mathcal{T}}] = \chi(\mathbf{x}_{\mathcal{T}} \in \mathbf{X})$, where $\chi(A)$ equals 1 if A is true, and 0 if A is false. Since $\mathbf{x}_{\mathcal{T}}$ determines the *initial* condition of the reverse trajectory, and is sampled from the Boltzmann distribution, the left hand side of the path average relation (2.42) becomes

$$\begin{aligned}\langle \chi(\mathbf{x}_{\mathcal{T}} \in \mathbf{X}) \rangle_{\mathrm{rev},\mathcal{T}} &= \int d\mathbf{x}\, \chi(\mathbf{x} \in \mathbf{X}) e^{-\beta(H_{\mathrm{sys}}(\mathbf{x}^*) - F)} \\ &= p_{\mathrm{eq}}(\mathbf{X}), \qquad (2.51)\end{aligned}$$

where $p_{\mathrm{eq}}(\mathbf{X})$ is the probability of finding $\mathbf{x} \in \mathbf{X}$ in thermal equilibrium. If magnetic fields are present, it is important to note that $H_{\mathrm{sys}}(\mathbf{x}^*)$ here is really shorthand for $H_{\mathrm{sys}}(\mathbf{x}^*, \lambda^*_{\mathcal{T}})$, which is equal to $H_{\mathrm{sys}}(\mathbf{x}, \lambda_{\mathcal{T}})$ thanks to the reversal of magnetic field direction implied in λ^*.

The other side of the path average relation (2.42) can be expressed in terms of the probability $p_{\mathrm{fwd},\mathcal{T}}(\mathbf{X}) = \langle \chi(\mathbf{x}_{\mathcal{T}} \in \mathbf{X}) \rangle_{\mathrm{fwd},\mathcal{T}}$ of finding the system in $\mathbf{X}$ at time $t = \mathcal{T}$:

$$\begin{aligned}\left\langle \chi(\mathbf{x}_{\mathcal{T}} \in \mathbf{X}) e^{-\beta \mathcal{W}} \right\rangle_{\mathrm{fwd},\mathcal{T}} &= p_{\mathrm{fwd},\mathcal{T}}(\mathbf{X}) \frac{\left\langle \chi(\mathbf{x}_{\mathcal{T}} \in \mathbf{X}) e^{-\beta \mathcal{W}} \right\rangle_{\mathrm{fwd},\mathcal{T}}}{\langle \chi(\mathbf{x}_{\mathcal{T}} \in \mathbf{X}) \rangle_{\mathrm{fwd},\mathcal{T}}} \qquad (2.52) \\ &= p_{\mathrm{fwd},\mathcal{T}}(\mathbf{X}) \left\langle e^{-\beta \mathcal{W}} \right\rangle_{\mathrm{fwd},\mathcal{T},\mathbf{X}} . \qquad (2.53)\end{aligned}$$

I have streamlined the notation by introducing a restricted trajectory ensemble average in the second line, which only includes trajectories that end in $\mathbf{X}$ at time $\mathcal{T}$.

Inserting Equations (2.51) and (2.52) into (2.42) gives a general expression for the finite-time evolution of $p_{\mathrm{fwd},\mathcal{T}}(\mathbf{X})$:

$$p_{\mathrm{fwd},\mathcal{T}}(\mathbf{X}) = p_{\mathrm{eq}}(\mathbf{X}) \frac{\left\langle e^{-\beta \mathcal{W}} \right\rangle_{\mathrm{fwd},\mathcal{T}}}{\left\langle e^{-\beta \mathcal{W}} \right\rangle_{\mathrm{fwd},\mathcal{T},\mathbf{X}}} \qquad (2.54)$$

For an ergodic system, which loses memory of its initial conditions in finite time, the desired steady-state probability can be found by simply taking the long-time limit

$$p_{\mathrm{ss}}(\mathbf{X}) = \lim_{\mathcal{T} \to \infty} p_{\mathrm{fwd},\mathcal{T}}(\mathbf{X}). \qquad (2.55)$$

But evaluating this limit requires some care, because $\mathcal{W}$ diverges.

2.4.2 Cumulant Expansion

To take the $\mathcal{T} \to \infty$ limit of Equation (2.54), I will make use of the fact that $\ln\langle e^{-\beta\mathcal{W}}\rangle$ is the cumulant generating function for the distribution over $\mathcal{W}$:

$$\ln \left\langle e^{-\beta\mathcal{W}}\right\rangle_{\text{fwd},\mathcal{T}} = \sum_{m=1}^{\infty} \frac{(-\beta)^m}{m!} \left\langle\mathcal{W}^m\right\rangle^c_{\text{fwd},\mathcal{T}} \tag{2.56}$$

Explicit formulas for the cumulants $\langle\mathcal{W}^m\rangle^c_{\text{fwd},\mathcal{T}}$ are obtained from the coefficients of a power series expansion in β of $\ln\left\langle e^{-\beta\mathcal{W}}\right\rangle_{\text{fwd},\mathcal{T}}$. A helpful review of the properties of cumulants $\langle\mathcal{W}^m\rangle^c_{\text{fwd},\mathcal{T}}$ in the context of a related derivation can be found in [54]. The first cumulant $\langle\mathcal{W}\rangle^c_{\text{fwd},\mathcal{T}} = \langle\mathcal{W}\rangle_{\text{fwd},\mathcal{T}}$ is the mean of the distribution, the second $\langle\mathcal{W}^2\rangle^c_{\text{fwd},\mathcal{T}}$ is the variance, and the higher-order terms provide progressively more information about the shape of the distribution. A Gaussian distribution is fully described by the first two cumulants, with all the higher-order terms vanishing.

Thanks to the explicit normalization of Equation (2.42), the expressions for the steady-state distribution will involve differences between cumulants $\left\langle\mathcal{W}^k\right\rangle^c_{\text{fwd},\mathcal{T},\mathbf{X}}$ of the restricted trajectory ensemble, and the cumulants $\left\langle\mathcal{W}^k\right\rangle^c_{\text{fwd},\mathcal{T}}$ for the full ensemble.

As illustrated in Figure 2-6, these cumulant differences converge to a finite limit as $\mathcal{T} \to \infty$ for ergodic systems [54]. I can therefore define:

$$\Delta\langle\mathcal{W}^k\rangle^c(\mathbf{X}) = \lim_{\mathcal{T}\to\infty} \left[\left\langle\mathcal{W}^k\right\rangle^c_{\text{fwd},\mathcal{T},\mathbf{X}} - \left\langle\mathcal{W}^k\right\rangle^c_{\text{fwd},\mathcal{T}}\right] \tag{2.57}$$

The key property of cumulants for the present analysis is that shifting the mean of the distribution while leaving its shape unchanged has no effect on the cumulants of order 2 and higher [54]. It will therefore be convenient group all these higher cumulants together, and refer to the resulting quantity as the "excess fluctuations" Φ_{ex}:

$$\Phi_{\text{ex}}(\mathbf{X}) = \sum_{k=2}^{\infty} \frac{(-\beta)^k}{k!} \Delta\langle\mathcal{W}^k\rangle^c(\mathbf{X}). \tag{2.58}$$

Note that this term does not simplify in the thermodynamic limit, even though the

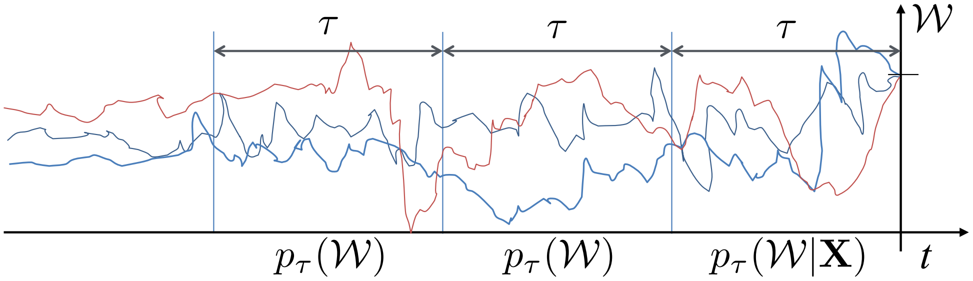

Figure 2-6: Color. Ergodicity means that a long trajectory can be divided into a sequence of uncorrelated segments, so that the average over a single trajectory becomes equivalent to the average over an ensemble of independent systems. The stipulation that trajectories end in $\mathbf{X}$ only affects the final segment, so the cumulant differences only depend on the work statistics in this finite time window.

Central Limit Theorem guarantees that the work distribution will look progressively more Gaussian as the system size increases, and only the first term of this sum is nonzero for a Gaussian distribution. All the cumulants $\langle \mathcal{W}^k \rangle^c_{\text{fwd},\mathcal{T}}$ of an extensive quantity like $\mathcal{W}$ become proportional to the system size, and so their relative sizes converge to finite limits as $N \to \infty$ (cf. [49, p. 46]). This apparent paradox results from the sensitive dependence of Φ_{ex} on rare fluctuations in the far tail of the work distribution, where the Central Limit Theorem does not apply (cf. [46]).

I will also give a special symbol to the first cumulant difference, and call it the "excess work," because it is the mean additional work done during the fluctuation to state $\mathbf{X}$, beyond the work already being done in the steady state:

$$\mathcal{W}_{\text{ex}}(\mathbf{X}) = \Delta\langle \mathcal{W} \rangle(\mathbf{X}). \tag{2.59}$$

This notation is potentially confusing, because several alternative definitions of "excess work" already exist in the literature. Equation (2.59) is most closely related to the notion of "excess heat" proposed by Oono and Paniconi in the context of phenomenological thermodynamics [79], applied to stochastic processes by Komatsu and Nakagawa [55]. This definition involves subtracting off the steady-state rate of heat production from a transient relaxation trajectory to obtain a finite heat associated with the transition. Basu, Maes and Netočný applied an analogous procedure to define an excess work associated with the relaxation to a new steady state [3]. The definition of Equation (2.59) is almost identical to that of Basu *et al.*, except that the work is evaluated along the trajectories that *generate* the fluctuation.

In terms of these quantities, the $\mathcal{T} \to \infty$ limit of Equation (2.54) is [67]

$$p_{\text{ss}}(\mathbf{X}) = p_{\text{eq}}(\mathbf{X}) e^{\beta \mathcal{W}_{\text{ex}}(\mathbf{X}) - \Phi_{\text{ex}}(\mathbf{X})}. \tag{2.60}$$

The rest of this chapter and the example in the next one will be devoted to unpacking the implications of this expression.

2.4.3 Discussion

The quantity Φ_{ex} in the exponent of (2.60) – involving all the cumulants of the work distribution even in the large system limit – is not a something we are used to dealing with from other areas of physics. It usually much more challenging to measure or even estimate than the steady-state probability distribution itself. Equation (2.60) is thus most useful when Φ_{ex} is independent of $\mathbf{X}$, and the expression reduces to the McLennan form (2.14) up to a normalization constant. The advantage of knowing the exact expression (2.60) is that it provides a basis for estimating the size of correction terms to the McLennan approximation, so that its full range of validity can be carefully established.

Starting from Equation (2.40), one can obtain a whole family of exact expressions for the steady-state distribution by making different choices for $\mathcal{O}[\mathbf{x}_0^{\mathcal{T}}]$. Most expressions that have been studied so far from this path ensemble average approach are special cases of the general form:

$$\mathcal{O}[\mathbf{x}_0^{\mathcal{T}}] = \delta(\mathbf{x}_0 - \mathbf{x}) e^{\alpha \beta \mathcal{W}[\mathbf{x}_0^{\mathcal{T}}]} \tag{2.61}$$

where $\mathbf{x}$ is the microstate whose probability is being computed and α is a number between 0 and 1 (cf. [54, eq. 4.19], which is not quite the same, because $\mathbf{x}_{\mathcal{T}}$ is also restricted). Macrostate probabilities can be found by replacing the δ-function with an indicator function χ as done above. Note that this form restricts the initial system state at $t = 0$, as opposed to the final-state restriction of Equations (2.51) and (2.52) imposed at $t = \mathcal{T}$. The final-state restriction in the forward ensemble can be transformed into an initial-state restriction in the reverse ensemble by a trivial relabeling of the time axis, and corresponds to the choice $\alpha = 1$ in Equation (2.61).

The best choice of α depends on what the expression will be used for. The simplest choice, and the one first examined historically, is $\alpha = 0$ [17, 18]. A symmetrized form has also been studied, using $\alpha = 1/2$ [53]. This form provides a convenient cancellation of second-order terms in a series expansion in small driving force. But neither of these choices is appropriate for the expansion about linearized coarse-grained dynamics that I will construct in the next section.

The first reason for this has to do with coarse-graining. The effective dynamics of $\mathbf{X}$ will depend in general on the distribution over internal configurations $\mathbf{x} \in \mathbf{X}$ at each point in time. $\mathbf{x}_0$ is always sampled from the Boltzmann distribution in the forward ensemble, but by time $\mathcal{T}$ the distribution within a given $\mathbf{X}$ has relaxed to a nonequilibrium steady state that could be very different. The opposite is true for the reverse ensemble: $\mathbf{x}_{\mathcal{T}}$ is Boltzmann-distributed, and $\mathbf{x}_0$ is sampled from the new steady state. Phenomenological equations that accurately capture the fluctuations of $\mathbf{X}$ in the steady state are sufficient to obtain the required work statistics only if the relevant parts of the $\mathbf{X}$ trajectory have the internal configurations sampled from the steady state. Since the trajectory functional defined in Equation (2.61) imposes a restriction on $\mathbf{x}_0$, I need all the dependence on $\mathcal{W}$ to be in the reverse average to guarantee that $\mathbf{x}$ is properly distributed within $\mathbf{X}$ during the transient that determines the "excess" quantities. This happens only when $\alpha = 1$.

The second limitation of (2.61) with $\alpha < 1$ becomes relevant when the steady-state distribution under the forward driving protocol $\lambda_0^{\mathcal{T}}$ is different from the distribution under the reverse protocol $\hat{\lambda}_0^{\mathcal{T}}$, as in the example of Chapter 3. When $\alpha < 1$, the resulting expression for the steady-state probability of $\mathbf{x}$ under the forward protocol includes work averages under the reverse protocol. To compute the probability of a small fluctuation, it may be necessary to consider trajectories that are very rare in the reverse protocol steady state. There is then no reason to expect that a set of approximate equations of motion linearized around the forward steady state should be adequate to determine the relevant work statistics.

Equation (2.60), based on the choice $\alpha = 1$, avoids both these problems. This equation relates the probabilities of typical fluctuations in the observables $\mathbf{X}$ to typical fluctuations of $\mathcal{W}$ in the steady state. The relevant work statistics can be plausibly estimated using an approximate coarse-grained dynamics valid near the steady-state mean.

Note that all the expressions in this family involve work statistics under the driven dynamics, and they can only provide substantive predictions in regimes such as the one discussed in Section 2.5, where the effect of the driving on the work fluctuations can be estimated without knowing the steady state distribution. An alternative approach makes use only of equilibrium statistics, so that the nonequilibrium behavior can be predicted in principle to arbitrary accuracy based on measurements of equilibrium correlations [65, 16]. This approach makes use of a new dynamical variable called the "traffic" or "dynamical activity," however, and the predictions demand prior knowledge of how this new variable depends on the strength of the driving force. Progress in this direction demands building up stronger intuition for how the dynamical activity behaves in systems of interest.

2.5 Extended Linear Response

I will now use Equation (2.60) to determine the range of validity of the McLennan distribution (2.14), originally derived within linear response theory. As I noted above, the form of (2.60) makes this question equivalent to the problem of determining when Φ_{ex} is independent of $\mathbf{X}$.

The most obvious way to make Φ_{ex} constant is to make it vanish. This happens in the weak driving limit, since all the terms in Φ_{ex} are of higher order than $\mathcal{W}_{\mathrm{ex}}$ in the strength of the driving force $\mathcal{F}$. This driving force could be the amplitude of a periodic variation in a control parameter, the shear rate in a flow-driven system, or the chemical potential difference or temperature difference in boundary-driven system. The $\mathcal{F} \to 0$ limit gives rise to traditional linear response theory, as exemplified by the Einstein Relation [21], the Green-Kubo relation [65, 53, 54] and similar results connecting near-equilibrium behavior to equilibrium fluctuations [66].

At larger values of $\mathcal{F}$, Φ_{ex} does not vanish, but it can be treated as part of the normalization constant as long as it is independent of $\mathbf{X}$. To see when the $\mathbf{X}$-dependence comes in, I must make some additional assumptions about the statistics of $\mathcal{W}$.

I will do this by postulating phenomenological equations for the fluctuation trajectories in the nonequilibrium steady state of a set of observables $\mathbf{X}$, which are taken to be instantaneous averages over a finite but macroscopic system volume. I will initially specialize to the case of a system driven by an externally imposed flow field that is constant in time, such as the sheared colloid discussed in Chapter 3. As shown in Appendix A, the instantaneous work rate in this class of systems is entirely determined by the system's current microstate. Thus I can let one of the coarse-grained variables in the vector $\mathbf{X}$ control the exact work rate:

$$\dot{\mathcal{W}} = V\mathcal{F} \cdot (X_1 + J_{\text{ss}}) \tag{2.62}$$

where J is the current conjugate to the thermodynamic force $\mathcal{F}$ supplied by the flow, and the first element of $\mathbf{X}$ is the deviation of the current from its $\mathcal{F}$-dependent mean steady state value $X_1 = J - J_{\text{ss}}(\mathcal{F})$.

I will obtain a set of conditions on the phenomenological dynamics of this $\mathbf{X}$ that guarantee that $\Phi_{\text{ex}}(\mathbf{X})$ is independent of $\mathbf{X}$. By perturbing about this case, I will determine the factors that control the impact of Φ_{ex} on the steady state distribution.

The bulk of the section will follow my original presentation of this material in [67]. At the end of the section, I will show how to generalize the results to systems driven by thermal or chemical gradients, where the work rate is determined by the time-derivative $\dot{\mathbf{X}}$ and not by $\mathbf{X}$ directly.

2.5.1 Phenomenological Equations and Fluctuation Trajectories

I first consider the case where the dynamics of $\mathbf{X}$ are well-described by a linear diffusion process, whose individual trajectories follow a form of the overdamped Langevin equation introduced in Section 2.3.2:

$$\dot{\mathbf{X}} = -A(\mathcal{F})\mathbf{X} + B(\mathcal{F})\boldsymbol{\xi}_t. \tag{2.63}$$

I have defined $\mathbf{X}$ such that $\mathbf{X} = 0$ is the most probable value in the steady state, and A and B are constant matrices independent of $\mathbf{X}$. A is positive definite, and B is diagonal with all positive entries. To allow for extensions beyond the traditional linear response regime, I will allow A to depend on $\mathcal{F}$. As I explain in Section 2.5.3, this implies that B also must depend on $\mathcal{F}$ to maintain thermodynamic consistency. I will require the elements of B to scale as $1/\sqrt{V}$ when the volume changes (while A remains constant), so that the size of fluctuations in the intensive variables $\mathbf{X}$ satisfies the Central Limit Theorem.

The solution for $t \geq 0$ with initial condition $\mathbf{X}_0$ is:

$$\mathbf{X}_t = e^{-At}\mathbf{X}_0 - \int_0^t dt'\, e^{-A(t-t')}B\boldsymbol{\xi}_{t'}. \tag{2.64}$$

To evaluate the conditional averages in Φ_{ex}, I need to obtain the ensemble of trajectories that *end* in $\mathbf{X}_{\mathcal{T}}$. This cannot be extracted easily from Equation (2.63) directly, because the noise realization $\boldsymbol{\xi}_t$ must be sampled from a modified distribution conditioned on this ending state. As shown in Appendix B, these trajectories can be found by solving a different Langevin equation with $\boldsymbol{\xi}_t$ sampled from the original δ-correlated distribution:

$$\dot{\mathbf{X}} = \tilde{A}\mathbf{X} + B\boldsymbol{\xi}_t, \tag{2.65}$$

where $\tilde{A} = A^T$ if $AA^T = A^TA$ and B is proportional to the identity matrix (in general it is given by a more complicated form derived in Appendix B). Replacing A by $\tilde{A}$ ensures that any circulating currents in the steady state of the original dynamics preserve their direction when the sign on the $A\mathbf{X}$ term is flipped.

The solution to Equation (2.65) can be found by integrating backwards from $\mathcal{T}$ to time $t < \mathcal{T}$:

$$\mathbf{X}_t = e^{\tilde{A}(t-\mathcal{T})}\mathbf{X}_{\mathcal{T}} - \int_t^{\mathcal{T}} dt'\, e^{\tilde{A}(t-t')} B \boldsymbol{\xi}_{t'}. \tag{2.66}$$

For notational simplicity, I will drop the tilde of $\tilde{A}$ from now on, since the original A will not be needed in the subsequent derivations.

2.5.2 Work Statistics for Linear Dynamics

The work done over a given trajectory $\mathbf{X}_0^{\mathcal{T}}$ is given by

$$\begin{aligned}
\mathcal{W} &= V\mathcal{F}\int_0^{\mathcal{T}} dt\, J_t && (2.67)\\
&= V\mathcal{F}\int_0^{\mathcal{T}} dt\, \hat{X}_1 \cdot \left[e^{A(t-\mathcal{T})}\mathbf{X}_{\mathcal{T}} - \int_t^{\mathcal{T}} dt'\, e^{A(t-t')} B\boldsymbol{\xi}_{t'} + J_{\mathrm{ss}}\right] && (2.68)\\
&= \mathcal{W}'(\mathbf{X}_{\mathcal{T}}) + \mathcal{W}_0 && (2.69)
\end{aligned}$$

where $\hat{X}_1$ is the unit vector in the X_1 direction, and $\mathcal{W}'$ is the part of the work that contains the dependence on the final condition $\mathbf{X}_{\mathcal{T}}$.

The $\mathbf{X}_{\mathcal{T}}$-dependent term $\mathcal{W}'$ has a finite $\mathcal{T} \to \infty$ limit and is independent of the noise realization $\boldsymbol{\xi}_0^{\mathcal{T}}$:

$$\lim_{\mathcal{T}\to\infty} \mathcal{W}'(\mathbf{X}) = V\mathcal{F}\hat{X}_1 \cdot A^{-1}\mathbf{X}. \tag{2.70}$$

Since this quantity is deterministic and contains all the $\mathbf{X}_{\mathcal{T}}$ dependence, we can immediately conclude that it is equal to $\mathcal{W}_{\mathrm{ex}}$ up to an additive constant, and that the $\mathbf{X}$-dependent part of Φ_{ex} vanishes. The remaining constants can be easily computed using Equations (2.59) and (2.58), yielding:

$$\mathcal{W}_{\text{ex}}(\mathbf{X}) = V\mathcal{F}\hat{X}_1 \cdot A^{-1}\mathbf{X} \tag{2.71}$$

$$\Phi_{\text{ex}} = \frac{1}{2}(\beta V \mathcal{F})^2 \langle (\hat{X}_1 \cdot A^{-1}\mathbf{X})^2 \rangle_{\text{ss}}. \tag{2.72}$$

But my central conclusion is in fact independent of the exact values of these constants, which are important only for normalizing the distribution. As long as a set of observables containing $\dot{\mathcal{W}}$ can be found that obeys the linear equation (2.63), this brief argument establishes that the steady-state fluctuations are given by the McLennan form

$$p_{\text{ss}}(\mathbf{X}) \propto p_{\text{eq}}(\mathbf{X}) e^{\beta \mathcal{W}_{\text{ex}}(\mathbf{X})}, \tag{2.73}$$

with the nonequilibrium correction fully determined by the mean work done on the way to the fluctuation.

This derivation of Equation (2.73) does not rely on the Gaussianity or whiteness of the noise term $\boldsymbol{\xi}_t$. As long as the noise term in Equation (2.63) is independent of $\mathbf{X}$, the linearity of the equation guarantees that the excess work $\mathcal{W}_{\text{ex}}(\mathbf{X})$ is the sum of a deterministic term that carries the $\mathbf{X}$-dependence and a stochastic term independent of $\mathbf{X}$. Colored or non-Gaussian noise will change the expression for Φ_{ex}, but will not introduce any $\mathbf{X}$-dependence into that quantity. In particular, Equation (2.73) remains the correct distribution for underdamped fluctuation dynamics with exponential noise, which have been successfully employed to model shear stress fluctuations in molecular dynamics simulations of simple fluids [6]. The underdamped equation for X can be converted into a pair of overdamped equations by treating $\dot{X}$ as an independent observable.

The perturbative calculations of Section 2.5.4, however, become more challenging when the noise is non-Gaussian or self-correlated. Before applying Equation (2.73) to such systems, one should verify that the expansion around linearity remains well-behaved.

2.5.3 Linearity and Nonlinearity

Since $\mathcal{W}_{\text{ex}}$ is a linear function of $\mathbf{X}$, adding it to the exponent in Equation (2.73) only changes the mean and not the covariance matrix of the Gaussian equilibrium distribution $p_{\text{eq}}(\mathbf{X})$. When $\mathbf{X}$ is one-dimensional, this implies

$$\frac{B(\mathcal{F})^2}{A(\mathcal{F})} = \frac{B(0)^2}{A(0)}. \tag{2.74}$$

This relationship will place an important constraint on the behavior of the nonlinear correction term in the discussion surrounding Equation (2.87) below.

But $\mathcal{W}_{\text{ex}}$ as given in Equation (2.71) is not necessarily a linear function of $\mathcal{F}$, because $A(\mathcal{F})$ can also depend on this parameter. To see how this fact extends traditional linear response theory, we can compute the typical current in a macroscopic system with a single observable $X = J - J_{\text{ss}}$, and compare this with the prediction of the Green-Kubo formula. The typical current is found by maximizing the probability (2.73), which yields:

$$J_{\text{ss}}(\mathcal{F}) = \frac{\beta V \mathcal{F}}{A(\mathcal{F})} \langle J^2 \rangle_{\text{eq}}, \tag{2.75}$$

where I have used the fact that the equilibrium distribution $p_{\text{eq}}(X)$ that follows from Equation (2.63) is Gaussian. The Green-Kubo formula of linear response theory predicts (cf. [25]):

$$J_{\text{ss}}(\mathcal{F}) = \beta V \mathcal{F} \int_0^\infty dt \langle J_t J_0 \rangle_{\text{eq}}. \tag{2.76}$$

Under the linear overdamped Langevin dynamics of Equation (2.63), Equation (2.76) becomes

$$J_{\text{ss}}^{(0)}(\mathcal{F}) = \frac{\beta V \mathcal{F}}{A(0)} \langle J^2 \rangle_{\text{eq}}. \tag{2.77}$$

Equations (2.77) and (2.75) start to differ from each other when the damping rate A begins to depart from its equilibrium value $A(0)$. The relative size of the difference is given by

$$\frac{J_{\text{ss}}^{(0)} - J_{\text{ss}}}{J_{\text{ss}}} = \frac{A(\mathcal{F})}{A(0)} - 1. \tag{2.78}$$

Equation (2.73) thus provides an extension of linear response forms like (2.76) that remains valid even when J_{ss} ceases to be a linear function of $\mathcal{F}$.

2.5.4 Nonlinear Correction

I now introduce a nonlinear term into the fluctuation dynamics (2.65). My goal is to identify the physical property of the system that controls how well the real steady-state distribution is approximated by Equation (2.73) when that expression is no longer exact. For this calculation, I again focus on the 1-D case, and consider an ensemble of fluctuation trajectories for $X = J - J_{\text{ss}}(\mathcal{F})$ described by

$$\dot{X} = A(\mathcal{F})X + \frac{\epsilon(\mathcal{F})}{2}X^2 + B(\mathcal{F})\xi_t \tag{2.79}$$

where ξ_t is again a Gaussian white noise term with mean 0 and autocorrelation function $\langle \xi_0 \xi_t \rangle = \delta_t$. The coefficient ϵ has dimensions of 1/[time][current]. Note that in one dimension, the ensemble of fluctuation trajectories is always simply a mirror image of the ensemble of relaxation trajectories, since there are no circulating steady-state current whose properties have to be preserved under the transformation.

The solution can be written as a power series in ϵ:

$$X(t) = X_t^{(0)} + \epsilon X_t^{(1)} + \epsilon^2 X_t^{(2)} + \dots. \tag{2.80}$$

Plugging this in to the equation of motion and collecting terms in powers of ϵ gives

$$\dot{X}^{(0)} = AX^{(0)} + B\xi_t \tag{2.81}$$

$$\dot{X}^{(1)} = AX^{(1)} + \frac{1}{2}(X^{(0)})^2 \tag{2.82}$$

In Appendix C, I use these two equations and the expression for the work (2.67) given above to show that:

$$\mathcal{W}_{\text{ex}}(X) = \frac{V\mathcal{F}}{A}X + \epsilon\frac{V\mathcal{F}}{4A^2}X^2 + \mathcal{N} + O(\epsilon^2) \tag{2.83}$$

$$\Phi_{\text{ex}}(X) = \epsilon\frac{\beta^2V^2\mathcal{F}^2B^2}{2A^4}X + \mathcal{N}' + O(\epsilon^2). \tag{2.84}$$

where $\mathcal{N}$, $\mathcal{N}'$ are constants independent of X.

Since Φ_{ex} depends on X, the general Equation (2.60) for the steady state distribution no longer reduces exactly to the McLennan form (2.73). The variational principle based on the McLennan form should still provide a good approximation of the typical steady-state behavior as long as $\frac{d}{dX}\Phi_{\text{ex}}(X) \ll \frac{d}{dX}\beta\mathcal{W}_{\text{ex}}(X)$ near $X = 0$. By comparing Equations (2.83) and (2.84) we see that this is true whenever

$$\tilde{\epsilon} \equiv \epsilon\frac{\beta V\mathcal{F}B^2}{2A^3} \ll 1. \tag{2.85}$$

This expression defines a dimensionless quantity $\tilde{\epsilon}$ that controls the accuracy of the McLennan approximation.

To assess the physical significance of $\tilde{\epsilon}$, we can compare it to the quadratic $O(\epsilon)$ term in the expression for the excess work (2.83). Since the variance in the unperturbed steady-state distribution is $\sigma_X^2 = B^2/2A$, we can write

$$\tilde{\epsilon} = 4\beta[\mathcal{W}_{\text{ex}}(\sigma_X) - \mathcal{W}_{\text{ex}}^{(0)}(\sigma_X)]. \tag{2.86}$$

where $\mathcal{W}_{\text{ex}}^{(0)}$ is the value computed under the linear dynamics alone, with $\epsilon = 0$. This quantity is thus equal to four times the extra mean work difference due to the nonlinear term during a typical fluctuation, in units of k_BT.

2.5.5 Degree of Nonequilibrium

Since the McLennan distribution (2.14) is sufficient to produce straightforward generalizations of “near-equilibrium” results like the Green-Kubo relation, the parameter $\tilde{\epsilon}$ of Equation (2.86) serves as a good measure of the “degree of nonequilibrium.” When

$\tilde{\epsilon} \ll 1$, fluctuation probabilities are still directly determined by entropy and energy exchange, and in this sense the system remains close to thermal equilibrium, even if the probabilities are far from the Boltzmann distribution. As $\tilde{\epsilon}$ becomes large, more subtle features of the system dynamics come into play. The symmetry expressed in Equation (2.40) still holds, but it becomes impossible to translate this into any concrete prediction about observable behavior without additional constraints on the rare fluctuations that control the higher cumulants. The $\tilde{\epsilon} \gg 1$ regime is truly far from equilibrium, in the sense that generalizations of equilibrium thermodynamics can no longer provide predictions or intuition about the system's evolution.

Combined with the assumptions about the dependence of A and B on $\mathcal{F}$ and V, Equation (2.85) shows how the degree of nonequilibrium $\tilde{\epsilon}$ depends on the strength of the driving force and on the system size. For small values of $\mathcal{F}$, these three parameters remain close to their equilibrium values $A(0), B(0), \epsilon(0)$, and so $\tilde{\epsilon} \propto \mathcal{F}$. This is the regime treated by linear response theory. To see what can happen at larger values, we can consider a system where $A = A(0)(1+k\mathcal{F})$, $B^2 = B(0)^2(1+k\mathcal{F})$, $\epsilon = \epsilon(0)(1+k\mathcal{F})$ for some constant k. These choices satisfy the constraint (2.13) on the relationship between A and B for $\epsilon(0) = 0$. They represent a force $\mathcal{F}$ that accelerates the relaxation to steady state while driving the system out of equilibrium, as the shear flow will do in Chapter 3. These stipulations imply that $\tilde{\epsilon}$ increases monotonically to its limit

$$\lim_{\mathcal{F}\to\infty} \tilde{\epsilon} = \epsilon(0)\frac{\beta V B^2(0)}{2kA(0)^3}. \tag{2.87}$$

If this quantity is small, then the system will remain "near equilibrium" for arbitrarily large values of the driving force $\mathcal{F}$.

The fact that $B^2 \propto 1/V$ is the only V-dependent parameter in Equation (2.85) immediately implies that $\tilde{\epsilon}$ is independent of system size. This is an important feature, which guarantees that $\tilde{\epsilon}$ remains an informative quantity even in the thermodynamic limit $V \to \infty$. Other plausible candidates for measuring the degree of nonequilibrium do not have this property. For example, the excess work done (in units of $k_B T$) during a typical fluctuation $\beta\mathcal{W}_{\rm ex}(\sigma_X)$ seems like a reasonable measure of how well the externally supplied work couples to the distribution over states. But Equation (2.83) reveals that this quantity grows as $\sqrt{V}$, because $\sigma_X \propto 1/\sqrt{V}$. To obtain a V-independent quantity in terms of excess work in Equation (2.86), I had to subtract off the part due to the linear terms, and isolate the nonlinear contribution.

2.6 Application to Other Models

For concreteness, I focused the analysis of Section 2.5 on a specific class of systems driven by externally imposed flow fields. This restriction came into play when I imposed a thermodynamic interpretation on the stochastic process described by Equation (2.63), stipulating that

$$\dot{\mathcal{W}} = V\mathcal{F} \cdot (X_1 + J_{\mathrm{ss}}). \tag{2.88}$$

But versions of the linear Langevin equation (2.63) and the nonlinear perturbation (2.79) can be used as a phenomenological description of the near-steady-state dynamics under any kind of driving force. In this section, I show how to apply the mathematical results of Section 2.5 to systems driven by gradients of temperature or chemical potential.

2.6.1 Transport of Energy and Particles

Consider a system in contact with several reservoirs that are not in equilibrium with each other, as described in Section 2.2.3 above. Energy and matter can flow through the system from one reservoir to another. I will label system macrostates with a variable $\mathbf{X}$ that contains the variation in system energy per unit volume $X_1 = [H_{\mathrm{sys}}(\mathbf{x}) - \langle H_{\mathrm{sys}}\rangle_{\mathrm{ss}}]/V$ as well as in the concentrations $X_i = [n_i - \langle n_i\rangle_{\mathrm{ss}}]/V$ of each kind of particle. To compute the work rate due to the imbalances of temperature and chemical potential, I will need to keep track of the flux from each reservoir separately. The linear Langevin equation (2.65) for computing the ensemble of fluctuation trajectories thus becomes:

$$\dot{\mathbf{X}} = \sum_\alpha \dot{\mathbf{X}}^{(\alpha)} = \sum_\alpha [\mathcal{A}^{(\alpha)}\mathbf{X} + \mathcal{B}^{(\alpha)}\boldsymbol{\xi}_t^{(\alpha)} + \dot{\mathbf{X}}_{\text{ss}}^{(\alpha)}] \tag{2.89}$$

with $A = \sum_\alpha \mathcal{A}^{(\alpha)}$ and $B^2 = \sum_\alpha (\mathcal{B}^{(\alpha)})^2$ so that the net change in $\mathbf{X}$ is still described by (2.65). The noise realizations $\boldsymbol{\xi}_t^{(\alpha)}$ are sampled independently for each reservoir. If the reservoirs are not in equilibrium with each other, then some of the $\dot{\mathbf{X}}^{(\alpha)}$'s will remain nonzero in the steady state $\mathbf{X} = 0$. This steady-state flux is contained in the constant terms $\dot{\mathbf{X}}_{\text{ss}}^{(\alpha)}$, which must satisfy $\sum_\alpha \dot{\mathbf{X}}_{\text{ss}}^{(\alpha)} = 0$.

The work statistics can be computed using the definition (2.34) of $\mathcal{W}$:

$$\dot{\mathcal{W}}(\mathbf{X}) = -k_B T V \sum_\alpha \left[(\beta^{(\alpha)} - \beta)\dot{X}_1^{(\alpha)} - \sum_{i>1} \beta^{(\alpha)}\mu_i^{(\alpha)}\dot{X}_i^{(\alpha)} \right] \tag{2.90}$$

$$= -k_B T V \sum_{\alpha,i} (\beta^{(\alpha)} - \beta) \Bigg[(\mathcal{A}_{1i}^{(\alpha)} X_i + \mathcal{B}_{1i}^{(\alpha)} \xi_t^i)$$

$$- \sum_{j>1} \beta^{(\alpha)}\mu_j^{(\alpha)} (\mathcal{A}_{ji}^{(\alpha)} X_i + \mathcal{B}_{ji}^{(\alpha)} \xi_t^i) \Bigg] + \dot{\mathcal{W}}_0 \tag{2.91}$$

$$= V(\mathbf{F}\cdot\mathbf{X} + \mathbf{B}\cdot\boldsymbol{\xi}_t) + \dot{\mathcal{W}}_0 \tag{2.92}$$

where the vectors $\mathbf{F}$ and $\mathbf{B}$ are defined by this equation, and $\dot{\mathcal{W}}_0$ is a constant that contains the contribution of the $\dot{\mathbf{X}}_{\text{ss}}^{(\alpha)}$. $\mathbf{F}$ and $\mathbf{B}$ are both linear in the temperature and chemical potential differences when these differences are small, but can become nonlinear for larger differences, since $\mathcal{A}^{(\alpha)}$ and $\mathcal{B}^{(\alpha)}$ can be functions of all the reservoir parameters $\beta^{(\alpha)}, \mu_j^{(\alpha)}$.

Since $\langle \boldsymbol{\xi}_t \rangle = 0$, this expression for $\dot{\mathcal{W}}$ is formally identical with that obtained in Section 2.5.2 for the purposes of computing the mean work, except that $\hat{X}_1$ is replaced by $\mathbf{F}$. Φ_{ex} remains independent of $\mathbf{X}$, because all the dependence on the final condition $\mathbf{X}_\mathcal{T}$ is still contained in a term independent of $\boldsymbol{\xi}_t$. Thus I obtain results very similar to Equations (2.71) and (2.72)

$$\mathcal{W}_{\text{ex}}(\mathbf{X}) = V\mathbf{F} \cdot A^{-1}\mathbf{X} \tag{2.93}$$

$$\Phi_{\text{ex}} = \frac{1}{2}(\beta V \mathcal{F})^2 \langle (\mathbf{F} \cdot A^{-1}\mathbf{X})^2 \rangle_{\text{ss}}, \tag{2.94}$$

and the fluctuations of $\mathbf{X}$ continue to be described by the McLennan form (2.73).

To study perturbations away from this linear case, I again specialize to the one-dimensional equation (2.79), now divided into contributions from two chemical reservoirs with chemical potentials $\mu^{(1)} \geq \mu^{(2)}$:

$$\dot{X} = \mathcal{A}^{(1)}X + \epsilon^{(1)}X^2 + \mathcal{B}^{(1)}\boldsymbol{\xi}_t^{(1)} + \dot{X}_{\text{ss}}^{(1)} + \mathcal{A}^{(2)}X + \epsilon^{(2)}X^2 + \mathcal{B}^{(2)}\boldsymbol{\xi}_t^{(2)} + \dot{X}_{\text{ss}}^{(2)}. \tag{2.95}$$

with $\epsilon = \epsilon^{(1)} + \epsilon^{(2)}$ for consistency with (2.79), and $\dot{X}_{\text{ss}}^{(2)} = -\dot{X}_{\text{ss}}^{(1)}$. The work rate is

$$\begin{aligned}\dot{\mathcal{W}} =& V\Bigg([\mu^{(1)}\mathcal{A}^{(1)} + \mu^{(2)}\mathcal{A}^{(2)}]X + [\mu^{(1)}\epsilon^{(1)} + \mu^{(2)}\epsilon^{(2)}]X^2 \\ &+ \sqrt{(\mu^{(1)}\mathcal{B}^{(1)})^2 + (\mu^{(2)}\mathcal{B}^{(2)})^2}\boldsymbol{\xi}_t + [\mu^{(1)} - \mu^{(2)}]\dot{X}_{\text{ss}}^{(1)} \Bigg).\end{aligned} \tag{2.96}$$

Using the calculations in Appendix C, I find:

$$\mathcal{W}_{\text{ex}}(X) = V\bar{\mu}X + V\frac{\epsilon}{4A}\left(\bar{\mu} + 2\bar{\mu}'\right)X^2 + \mathcal{N} + O(\epsilon^2) \tag{2.97}$$

where

$$\bar{\mu} \equiv \frac{\mu^{(1)}\mathcal{A}^{(1)} + \mu^{(2)}\mathcal{A}^{(2)}}{A} \tag{2.98}$$

$$\bar{\mu}' \equiv \frac{\mu^{(1)}\epsilon^{(1)} + \mu^{(2)}\epsilon^{(2)}}{\epsilon} \tag{2.99}$$

$$\bar{\mu}'' \equiv \frac{\sqrt{(\mu^{(1)}\mathcal{B}^{(1)})^2 + (\mu^{(2)}\mathcal{B}^{(2)})^2}}{B} \tag{2.100}$$

are average chemical potentials weighted by the linear rates, the nonlinear corrections, and the noise amplitudes, respectively.

The fluctuation term can be written as:

$$\Phi_{\text{ex}}(\mathbf{X}) = \tilde{\epsilon}\beta\mathcal{W}_{\text{ex}}(X) + \mathcal{N}' + O(\epsilon^2) \tag{2.101}$$

where

$$\tilde{\epsilon} = \beta\frac{\epsilon V}{A}\left(\bar{\mu} + 2\bar{\mu}' - \bar{\mu}'' - 3\frac{\bar{\mu}'\bar{\mu}''}{\bar{\mu}}\right)\frac{B^2}{2A} \tag{2.102}$$

is the dimensionless measure of the strength of the nonlinearity. This cannot be exactly identified with the $O(\epsilon)$ term from $\mathcal{W}_{\text{ex}}$ as in Equation (2.86), because of the $\bar{\mu}''$ terms that are absent from $\mathcal{W}_{\text{ex}}$. Instead, the expansion parameter and the extra excess work due to the nonlinearity are related by

$$\tilde{\epsilon} = 4\beta[\mathcal{W}_{\text{ex}}(\sigma_X) - \mathcal{W}_{\text{ex}}^{(0)}(\sigma_X)]\left(1 - \bar{\mu}''\frac{\bar{\mu} + 3\bar{\mu}'}{\bar{\mu}(\bar{\mu} + 2\bar{\mu}')}\right). \tag{2.103}$$

But if the contributions of the two reservoirs to each of the terms of Equation (2.95) are in similar proportions, so that $\bar{\mu} \approx \bar{\mu}' \approx \bar{\mu}''$, this simplifies to

$$\tilde{\epsilon} = -\frac{4}{3}\beta[\mathcal{W}_{\text{ex}}(\sigma_X) - \mathcal{W}_{\text{ex}}^{(0)}(\sigma_X)]. \tag{2.104}$$

Thus as long as the three different ways of averaging $\mu^{(1)}$ and $\mu^{(2)}$ give approximately the same answer, the magnitude of the expansion parameter can still be estimated based on the contribution of nonlinearities to the excess work.

2.6.2 Chemical Reactions

A similar analysis can be applied to a well-mixed solution of reacting chemicals, where the nonequilibrium driving force is provided by a chemostat that maintains some of the chemical concentrations at fixed values. The equation for the work rate (2.90) becomes slightly more complicated, because particles are removed from the chemical reservoir by being converted into different kinds of particles (cf. [39, 81] for a complete presentation of the thermodynamics of such systems). In the example I will present in Chapter 5, I make some simplifying approximations that allow Equation (2.90) to apply as written.

The more fundamental problem in applying the results of Section 2.5 to chemical reactions is the failure of the Langevin equation (2.63) to capture the exact work statistics required to evaluate Φ_{ex}.

At the molecular level, chemical reaction dynamics involve transitions among discrete states, which can be described by a Markov jump process as discussed in Section 2.3.1 above. The Langevin equation is only a valid approximation in the limit of large system size. It turns out that the key assumption is that many individual chemical reactions occur in the time required for the concentration to change appreciably [39]. Then on the time scale of the concentration dynamics, the net effect of all these rapid jumps looks like Gaussian white noise.

Although the Langevin equation accurately describes the *typical* fluctuation dynamics in a macroscopic chemical system, assessing the $\mathbf{X}$-dependence of Φ_{ex} requires considering the extremely *rare* fluctuations that influence the higher cumulants. As shown in [39], the Langevin equation successfully provides the exact entropy statistics only in the detailed-balance equilibrium state where the forward and reverse jump rates for each reaction are equal. When these rates become sufficiently unequal, then the typical trajectories successfully modeled by the Langevin equation only contain jumps in the more likely direction, with the stochasticity resulting entirely from the timing of the jumps. The probabilities of the rare trajectories that contain reverse jumps are no longer related in any obvious way to the statistics of these small fluctuations.

This disconnect between observable small fluctuations and reverse trajectory probabilities constitutes an important aspect of the "edge of thermodynamics." When the Langevin approximation successfully provides the full work statistics, then Φ_{ex} can be exactly computed from measurements of typical fluctuations that have plausible physical relevance. When the approximation fails, Φ_{ex} contains new information about rare trajectories that become astronomically improbable in the limit of large system size, and may never occur in any actual realizations of the process. This seems to destroy the utility of thermodynamic expressions like (2.60). But such expressions can still remain relevant when the real probabilities are *bounded* by the prediction of the Langevin equation. Such a bound has recently been discovered for the asymptotic statistics of the total entropy change in steady states of generic Markov jump processes [30], but it is not yet clear if something similar can be done for the conditional averages of (2.60).

Chapter 3

Shear Thinning in Brownian Colloids

Strongly driven colloidal suspensions are commonly used as examples of far-from-equilibrium steady state systems where the relationship between currents and applied fields can violate the predictions of linear response theory (cf. [27, 35, 56, 94, 98]). In this chapter, I describe how to measure the key quantities $\mathcal{W}_{\mathrm{ex}}$ and Φ_{ex} of Chapter 2 in a numerical simulation of a sheared colloid. I find that the form of the steady state distribution given in equation (2.73) for vanishing nonlinearities generates a qualitatively correct prediction of the observed decrease in viscosity with increasing shear rate, while the $O(\tilde{\epsilon})$ correction from the first term in Φ_{ex} is sufficient to maintain quantitative agreement with the actual distribution well into the thinning regime. This system is thus poised on the edge of the expanded near-equilibrium regime defined in Chapter 2. The simple approximation based on $\mathcal{W}_{\mathrm{ex}}$ can be expressed in terms of a few easily accessible parameters, and provides solid physical intuition for the basic phenomenon even when the quantitative predictions begin to fail.

3.1 Setup

Consider a suspension of small identical spheres in a liquid bath. The particles are small enough that Brownian motion can equilibrate their spatial configuration rapidly compared to the timescale of the experiment, producing a steady state independent of initial conditions. Electrostatic repulsion keeps the spheres far enough apart that

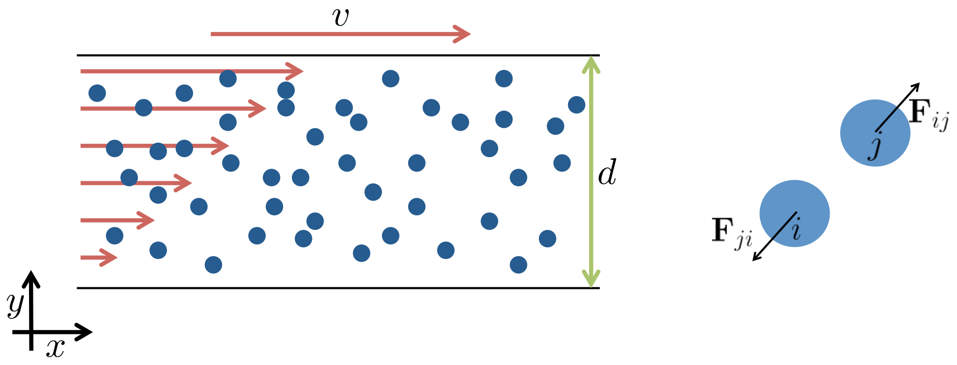

Figure 3-1: Color. Shear cell with periodic boundary conditions along flow direction. The reflecting walls on the top and bottom are separated by a distance d. The top wall moves at constant speed v in the x direction, while the bottom wall is fixed, causing a linear gradient $\dot{\gamma} = v/d$ in the solvent flow velocity along the y direction. The suspended colloidal particles repel each other with equal and opposite forces $\mathbf{F}_{ij} = -\mathbf{F}_{ji}$.

the disturbance each particle creates in the flow field has no effect on the trajectories of the other particles, while ions in the solvent screen the charges and exponentially suppress the interaction at large separations.

3.1.1 Flow-Induced Steady State

As illustrated in Figure 3-1, a nonequilibrium steady state can be created by moving one wall of the chamber containing the suspension at a constant velocity v while keeping the opposite wall fixed, thus setting up a steady shear flow in the gap of width d between the walls. The strength of the shear flow can be quantified in a form independent of the system dimensions as the "shear rate" $\dot{\gamma} = v/d$. A constant shear rate can be maintained by using periodic boundary conditions in the flow direction (which can be approximated in an experiment by using a cylindrical geometry). I will define coordinates such that the moving wall travels in the $+x$ direction and the y axis points from the stationary wall to the moving wall.

Three important dimensionless parameters for the dynamics of a sheared colloid are the Reynolds number $\mathrm{Re} = \rho\dot{\gamma}a^2/\eta_0$, the Peclet number $\mathrm{Pe} = \dot{\gamma}a^2 b/k_B T$, and the volume fraction $\phi = (4/3)\pi a^3 \rho_N$. Here ρ is the mass density of the fluid (assumed to be comparable to the density of the particles), a is the radius of a particle, η_0 is the viscosity of the suspending fluid, b is the drag coefficient of a particle ($= 6\pi\eta_0 a$ for a sphere with no-slip boundary conditions), and ρ_N is the number density of suspended particles.

In the Re $\ll 1$ limit, the instantaneous velocity of the particles can be regarded as fully determined by their spatial configuration (up to the rapidly equilibrating contribution from Brownian motion), so the set of particle positions is sufficient to define the full microstate, and overdamped Langevin equation (2.50) becomes applicable. Re can be kept in this regime while sweeping Pe up to any desired maximum value $\mathrm{Pe}_{\mathrm{max}}$ by choosing a viscosity such that $\eta_0 \gg \sqrt{\rho \mathrm{Pe}_{\mathrm{max}} k_B T/a}$.

Pe measures the importance of motion by convection in the shear flow relative to diffusive motion. It thus provides a dimensionless measure of the strength of the driving force, so that Pe $\ll 1$ is the linear-response regime, and Pe $\gg 1$ constitutes the “far from equilibrium” regime where shear thinning occurs.

ϕ governs the frequency with which particles interact with each other. If it increases beyond a certain critical point ϕ_c, the system will undergo a phase transition to a crystalline state. I will focus on the regime $\phi < \phi_c$, which is easiest to simulate accurately, and which is where the extended linear response prediction (2.73) is most likely to hold.

3.1.2 Equations of Motion

To describe this system mathematically, I will use the model employed in [94, 98] for the investigation of departures from near-equilibrium linear-response behavior in nonequilibrium steady states. This model is expressed as a set of overdamped Langevin equations:

$$\dot{x}_i = y_i \dot{\gamma} + \frac{1}{b} \sum_j \hat{\mathbf{x}} \cdot \mathbf{F}_{ji} + \sqrt{\frac{2k_B T}{b}} \xi_{x,i}(t) \tag{3.1}$$

$$\dot{y}_i = \frac{1}{b} \sum_j \hat{\mathbf{y}} \cdot \mathbf{F}_{ji} + \sqrt{\frac{2k_B T}{b}} \xi_{y,i}(t) \tag{3.2}$$

where $\mathbf{F}_{ji}$ is the conservative force exerted on particle i by particle j. The new feature of these equations as compared to Equation (2.50) from Chapter 2 is the addition of the flow term $y_i\dot{\gamma}$ to the equation for $\dot{x}_i$. This incorporates the effect of the imposed shear flow, and will alter the formula for the externally supplied work as described in Section 3.2 below.

I choose the force $\mathbf{F}_{ji}$ to be a screened Coulomb repulsion, with potential energy $U(r) = k_B T e^{-r/\lambda} z l_B / r$ as a function of the distance r separating a pair of particles. λ is the screening length, l_B is the Bjerrum length, and z is the number of elementary charges on each particle.

Equations (3.1-3.2) can be numerically simulated with the dilute limit of the Brownian Dynamics of Ermak and McCammon [24] or of the Stokesian Dynamics of Brady and Bossis [12], which generate the following discretized equations of motion in the regime I am considering:

$$x_i(t+\Delta t) = x_i(t) + y_i(t)\dot{\gamma}\Delta t + \frac{1}{b}\sum_j \hat{\mathbf{x}} \cdot \mathbf{F}_{ji}\Delta t + \Delta x_i^r \tag{3.3}$$

$$y_i(t+\Delta t) = y_i(t) + \frac{1}{b}\sum_j \hat{\mathbf{y}} \cdot \mathbf{F}_{ji}\Delta t + \Delta y_i^r \tag{3.4}$$

Equations (3.3) and (3.4) are simply iterated by the computer with a small enough time step that the results are insensitive to variations in time-step size.

As mentioned at the beginning of this section, I am considering the case where the particle size is much smaller than λ or zl_B, so that hydrodynamic interactions (particle-particle interactions mediated by disturbances in the solvent flow) have a negligible impact on the particle trajectories. This is what allows me to use the “dilute limit” of the Stokesian or Brownian Dynamics, where the mobility and resistance tensors are diagonal and independent of particle positions. Another consequence of this limit is that the actual particle radius a does not appear in the equations of motion; I therefore use the screening length λ as the microscopic length scale for computing the Peclet number and measuring distance from equilibrium.

3.1.3 Shear Stress

The macroscopic viscosity of the whole suspension at equilibrium will be larger than η_0, because both the disturbance of the flow field produced by individual particles and the mutual repulsion between pairs of particles make the suspension harder to shear than the bare fluid. As the suspension is sheared, however, the contribution of the particle repulsion to the viscosity decreases, and the suspension shear thins (cf. [11]). The particles cause the shear stress to vary with position in the suspension, so I define an overall shear stress for the system by averaging the local shear stress at the moving wall of the system over the whole wall area. This will be convenient for computing the work done by the moving wall later on, and gives a macroscopic parameter that can be directly observed in experiment via a measurement of the force applied to the wall. As shown in Appendix D, for a suspension of particles in a Newtonian solvent in the limit of zero Re with no hydrodynamic interactions, the instantaneous mean shear stress $\sigma_{xy}^{\mathrm{wall}}$ exerted by the fluid on the moving wall is:

$$\sigma_{xy}^{\mathrm{wall}} = \sigma_{xy}^{I} + \sigma_{xy}^{0}. \tag{3.5}$$

with

$$\sigma_{xy}^{I} = \frac{1}{2V} \sum_{i \neq j} \hat{\mathbf{x}} \cdot \mathbf{F}_{ij} \Delta y_{ij}. \tag{3.6}$$

Here V is the system volume, $\hat{\mathbf{x}}$ is the unit vector in the $+x$ direction, and $\Delta y_{ij} = y_j - y_i$. The right-hand side can be unambiguously determined from the system microstate, which I am taking to be the list of positions of all the particles. I can therefore choose $X = \sigma_{xy}^{I}$ as a coarse-grained observable within the framework of Chapter 2. The remaining term σ_{xy}^{0} is independent of the particle positions, so the work done by the moving wall will depend on the particle configuration through σ_{xy}^{I} alone.

When the shear rate is small compared to the diffusive relaxation rate, the overall viscosity of the suspension can be computed from the equilibrium fluctuations in σ_{xy}

using linear response theory [25]. As the shear rate continues to increase, the viscosity begins to deviate from this value as the suspension shear thins. In the following sections, I will use the expression for the steady-state distribution in Equation (2.60) from Chapter 2 to determine the most probable value of σ^I_{xy} and hence the contribution $\eta^I = -\sigma^I_{xy}/\dot{\gamma}$ to the viscosity in both the linear response regime and the shear thinning regime.

3.2 Probabilities and Work Statistics

Physically, the rate at which the moving wall does work on the fluid is given by the force $-A\sigma^{\text{wall}}_{xy}$ it exerts against the fluid (where A is the surface area of the wall) times the speed of the wall $\dot{\gamma}d$. Using equation (3.5), I thus obtain:

$$\dot{\mathcal{W}} = -V\dot{\gamma}\sigma^I_{xy} + \dot{\mathcal{W}}_0. \tag{3.7}$$

where $\dot{\mathcal{W}}_0$ is the part of the work that does not depend on the configuration of the particles. In Appendix A, I verify that this expression for the work rate satisfies microscopic reversibility (2.24) under the equations of motion (3.1-3.2). Since $\dot{\mathcal{W}}_0$ only affects the normalization, but not the shape of the distribution, I will set it to zero for the purpose of the calculations in this section.

Using equation (3.7), I can now compute the work done along any stochastic trajectory $\sigma^I_{xy}(t)$ by integrating the trajectory with respect to time. The distribution of $\mathcal{W}$ for trajectories ending at a given σ^I_{xy} value can then be estimated using a variant of an established method for obtaining "pre-history" ensembles in experiments on noisy electrical circuits [64]. Specifically, I let the system relax to the steady state at some value of $\dot{\gamma}$ and run there for a long time, while continuously recording the fluctuations in σ^I_{xy} (which can be determined directly from the fluctuations in the force applied to the moving plate in an experiment). As shown in Figure (3-2), I then choose some time interval $\mathcal{T}$ (much longer than the relaxation time to the steady state) and compute both the work $\mathcal{W}$ and the final value $\sigma^I_{xy,\mathcal{T}}$ of σ^I_{xy} for every segment of length

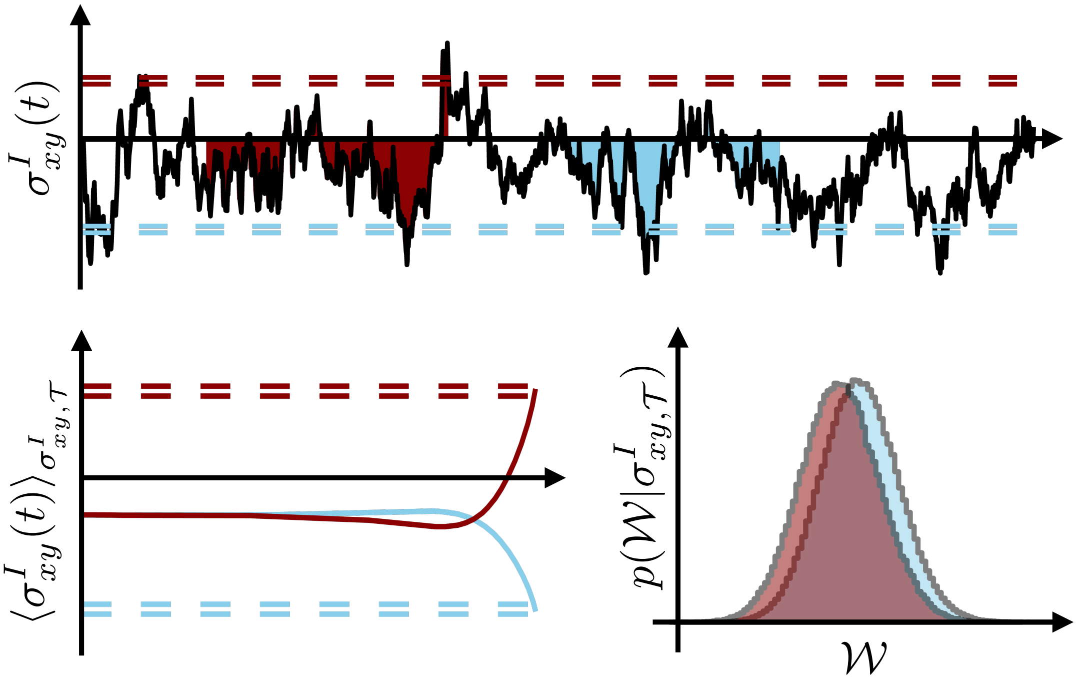

Figure 3-2: Color. Top: Portion of the raw $\sigma^{I}_{xy}(t)$ timeseries at high shear rate Pe = 19. I obtained the conditional work distribution by collecting trajectory segments based on their ending state. Two ending-state bins are indicated by dotted lines, and an example of a trajectory segment that ends in each bin is shaded in the corresponding color. The shaded areas are proportional to the interaction-dependent part of the work, according to equation (3.7). Bottom left: Average of all trajectory segments that end in each of the two bins from the top panel. Bottom right: Work distributions for trajectories ending in each of the two bins, shaded in the corresponding colors.

$\mathcal{T}$ in the whole trajectory. Finally, I bin the work values by the corresponding value of $\sigma^I_{xy,\mathcal{T}}$ to obtain the distribution of work for each bin, from which I can estimate the cumulant differences $\Delta\langle\mathcal{W}^n\rangle^c(\sigma^I_{xy})$ up to an additive constant independent of σ^I_{xy}.

3.2.1 Simulation Results

I simulated a sheared colloidal monolayer of $N = 100$ particles using the equations of motion (3.3) and (3.4). The colloid was confined to a square box of side length 20, with reflecting boundary conditions on the moving wall and the opposite wall, and periodic boundary conditions on the other sides. The other parameters were chosen as $k_BT = b = \lambda = zl_B = 1$. I ran this simulation for 20 different values of Pe, from 0 to 19, generating trajectories with lengths up to $t = 72,000$ in the given units, with time step size 0.001. The simulations were initialized with uniform random distributions of particle positions, and the initial transients were removed from the timeseries before analysis.

The first panel of Figure 3-3 shows how $F \equiv -k_BT \ln p_{\text{eq}}(\sigma^I_{xy})$ and the other two terms in the general expression for the steady-state distribution (2.60) depend on σ^I_{xy}, with the nonequilibrium terms evaluated at three different values of the shear rate. F is parabolic near $\sigma^I_{xy} = 0$, but requires a fourth-order polynomial to fit the far tails. $\mathcal{W}_{\text{ex}}$ is linear in σ^I_{xy} at low shear rates, starts curving slightly by Pe $=$ 10, and becomes noticeably quadratic by Pe $=$ 19, indicating that the $O(\epsilon)$ term in the expansion around linearized dynamics (2.83) has become important. Φ_{ex} is independent of σ^I_{xy} at low shear rates, but starts becoming σ^I_{xy}-dependent at about the same shear rate as $\mathcal{W}_{\text{ex}}(\sigma^I_{xy})$ begins to deviate from linearity, as predicted by Equation (2.84).

The second panel of Figure 3-3 shows the location of the peak of the steady-state distribution σ^{I*}_{xy} as a function of Pe. This is the value of the shear stress in observed in the thermodynamic limit $V \to \infty$ when the fluctuations become negligible. I plot the prediction based on the work statistics using Equation (2.60), and compare this with the distribution directly sampled from the simulation. I have also plotted the prediction of the McLennan approximation (2.73) that relies only on the mean excess

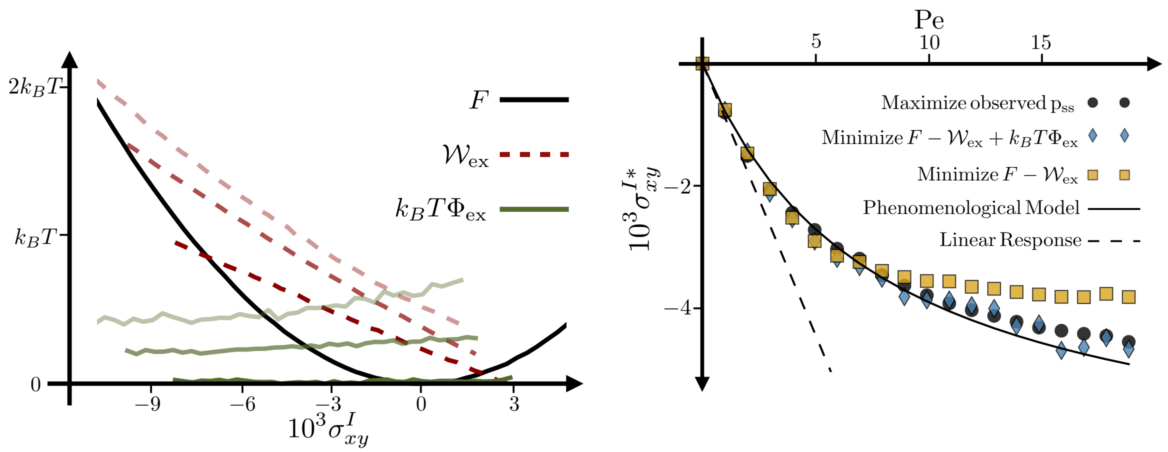

Figure 3-3: Color. Left: Using the method illustrated in Figure 3-2, I compute $\mathcal{W}_{\mathrm{ex}}(\sigma_{xy}^I)$ and $\Phi_{\mathrm{ex}}(\sigma_{xy}^I) \approx (\beta/2)\Delta\langle\mathcal{W}^2\rangle^c(\sigma_{xy}^I)$ for a range of values of σ_{xy}^I in my numerical simulation, and plot this data for Pe values 4, 10, and 19, increasing from bottom to top. Also plotted is the equilibrium free energy F extracted from the Pe $=$ 0 simulation run. Right: Empirical location σ_{xy}^{I*} of peak of probability distribution, compared with four thermodynamic predictions. The diamonds maximize the general expression for the steady-state distribution from Equation (2.60), including the Φ_{ex} term. The squares maximize the McLennan distribution contained in Equation (2.73), ignoring the contribution of Φ_{ex}. The dotted line is the Green-Kubo linear-response prediction of Equation (2.76), obtained from the σ_{xy}^I systems at equilibrium with Pe $=$ 0. Finally, the solid line is the prediction of the phenomenological model described in Section 3.2.2.

work. The McLennan distribution correctly captures the qualitative shear thinning behavior, where $\sigma_{xy}^{I*}I$ departs from the linear-response dependence on $\dot\gamma$ and eventually saturates, causing the contribution of the particle interactions to the viscosity $-\sigma_{xy}^I/\dot\gamma$ to fall off as $1/\dot\gamma$. This approximation predicts that the saturation occurs sooner than it actually does, but the correction based on my estimate of Φ_{ex} appears to entirely compensate for the discrepancy.

The straight dotted line in this panel is the linear-response prediction for the mean shear stress, computed from the equilibrium fluctuations using the Green-Kubo formula given in Equation (2.76). It correctly predicts the initial slope of the linear part of the curve, but starts noticeably departing from the true σ_{xy}^{I*} as soon as the dimensionless shear rate parameter Pe becomes greater than 1. The solid black line also uses the equilibrium fluctuations to compute the initial response near Pe $=$ 0, but then allows the relaxation rate to increase with Pe to account for shear-induced stirring as described below.

3.2.2 Physical Intuition

The fact that the McLennan distribution (2.73) correctly describes shear thinning suggests the use of a linear phenomenological model, like the one from Section 2.5, to develop some physical intuition for this behavior. The Gaussianity of $p_{\text{eq}}(\sigma^I_{xy})$ and the linearity of $\mathcal{W}_{\text{ex}}(\sigma^I_{xy})$ are both consistent with the linear dynamics of Equation (2.63). I will write these dynamics in terms of the timescale $\tau(\dot{\gamma})$ for relaxation to the steady state, the noise amplitude B and a term C that controls the location of the steady-state mean:

$$\dot{\sigma}^I_{xy} = -\frac{1}{\tau}\sigma^I_{xy} + C + B\xi(t). \tag{3.8}$$

Using this model in conjunction with the expression for the work rate (3.7), I find the mean excess work done on the way to a given σ^I_{xy} value in terms of the relaxation time τ:

$$\mathcal{W}_{\text{ex}}(\sigma^I_{xy}) = -V\dot{\gamma}\tau\sigma^I_{xy}. \tag{3.9}$$

The equilibrium distribution for this model can be written in terms of the equilibrium values τ_0 and B_0 of the model parameters τ and B:

$$\begin{aligned} p_{\text{eq}}(\sigma^I_{xy}) &= \frac{1}{\sqrt{B_0^2\tau_0}} e^{-\frac{(\sigma^I_{xy})^2}{B_0^2\tau_0}} \qquad (3.10)\\ &\equiv \frac{1}{\sqrt{B_0^2\tau_0}} e^{-\beta F(\sigma^I_{xy})}. \qquad (3.11)\end{aligned}$$

where I have introduced the free energy $F = k_B T(\sigma^I_{xy})^2/(B_0^2\tau_0)$ to simplify the notation and highlight the connection to the variational principle of Section 2.1.3. Since Φ_{ex} is independent of σ^I_{xy} under this linear model, the steady-state distribution is given by the McLennan form (2.73):

$$
\begin{aligned}
p_{\mathrm{ss}}(\sigma^I_{xy}) &\propto p_{\mathrm{eq}}(\sigma^I_{xy}) e^{\beta \mathcal{W}_{\mathrm{ex}}(\sigma^I_{xy})} && (3.12) \\
&\propto e^{-\beta(F - \mathcal{W}_{\mathrm{ex}})} && (3.13)
\end{aligned}
$$

The value of σ^I_{xy} observed in the steady state of a macroscopic system is found by maximizing p_{ss}, which is equivalent to minimizing $F - \mathcal{W}_{\mathrm{ex}}$. This minimum occurs when

$$
0 = \frac{\partial}{\partial \sigma^I_{xy}}(F - \mathcal{W}_{\mathrm{ex}}) = \frac{2k_B T}{B_0^2 \tau_0} \sigma^{I*}_{xy} + V\dot{\gamma}\tau. \tag{3.14}
$$

Solving for σ^{I*}_{xy} yields

$$
\begin{aligned}
\sigma^{I*}_{xy} &= -\frac{1}{2}\beta V B_0^2 \tau_0 \dot{\gamma}\tau && (3.15) \\
&= -\beta\dot{\gamma}\tau V \langle (\sigma^I_{xy})^2 \rangle_{\mathrm{eq}} && (3.16)
\end{aligned}
$$

where I have replaced the equilibrium parameters B_0^2 and τ_0 with the directly measurable equilibrium variance $\langle (\sigma^I_{xy})^2 \rangle_{\mathrm{eq}} = B_0^2 \tau_0 / 2$.

The contribution of interactions between particles to the viscosity is thus given by

$$
\begin{aligned}
\eta^I &\equiv -\frac{\sigma^{I*}_{xy}}{\dot{\gamma}} && (3.17) \\
&= \beta\tau V \langle (\sigma^I_{xy})^2 \rangle_{\mathrm{eq}}. && (3.18)
\end{aligned}
$$

The viscosity continues to be related to the relaxation time of stress fluctuations, as in classical linear response theory. But now the relaxation time τ is allowed to depend on the driving force $\dot{\gamma}$. Near equilibrium, relaxation to the steady state is primarily driven by diffusion, and $\tau = \tau_0$ is determined by the diffusion coefficient of the particles along with their number density and interaction potential. But as the shear rate increases, the imposed flow field stirs the particles, and randomizes their configuration faster than diffusion alone. As $\dot{\gamma} \to \infty$, diffusion becomes irrelevant, and $\tau \sim 1/\dot{\gamma}$ is the time required for two neighboring particles to be pushed past each other by the shear flow. This is a natural mechanism to generate the expression for τ as a function of driving force discussed in Section 2.5.5, which can keep the system in the near-equilibrium regime for arbitrarily large force values:

$$\tau = \frac{\tau_0}{1 + k\dot{\gamma}\tau_0}. \tag{3.19}$$

The solid line in the second panel of Figure 3-3 is the σ_{xy}^{I*} prediction of Equation (3.16), with the relaxation time τ given by Equation (3.19). τ_0 is determined from equilibrium fluctuations, and $k = 1.78$ was determined by fitting to the σ_{xy}^{I*} vs. Pe data. This model does a good job reproducing the line shape, with a single free parameter. Since it abstracts from all the details of the particle shape and the interaction potential, it also provides more general intuition for why shear thinning is such a generic phenomenon in suspensions of hard particles. Whenever the shear flow helps to accelerate the relaxation of the particle configuration to its steady-state distribution, τ should follow Equation (3.19) for small and large values of $\dot{\gamma}$, and the viscosity will be smaller at higher shear rates.

Chapter 4

Regulating Disassembly in Clathrin-Mediated Endocytosis

The life of eukaryotic cells depends on a constant traffic in lipid membranes among topologically separate structures. Vesicles are pinched off from the plasma membrane that surrounds the cell and fused with organelles in the interior for processing of their contents. Other vesicles carry newly synthesized membrane proteins from the endoplasmic reticulum to the plasma membrane, or deliver secretory proteins from the Golgi apparatus to the cellular exterior.

From a physical perspective, this constant and organized flux of membrane is quite a spectacular phenomenon – one that remains poorly understood at a fundamental level. Thanks to advances in membrane thermodynamics, we do have some idea of the conditions under which membrane buds spontaneously form and pinch off [90], and molecular biologists have identified the many proteins that help to regulate these processes in cells [87]. But there is still much to learn about how these collections of proteins cooperate to synthesize vesicles in a robust and tunable way.

The most experimentally accessible example of such a process is clathrin-mediated endocytosis (CME). This is a primary pathway for vesicle creation at the plasma membrane in mammalian cells, in which curved protein lattices self-assemble on the membrane in order to bend it [52]. The regular structure of these lattices makes them easy to identify and study with electron microscopy [83], and their location at the outer cell membrane renders them accessible to high-resolution *in vivo* imaging techniques [74]. As shown in Figure 4-1, the lattice is composed of three-legged subunits known as clathrin triskelia, which are themselves composed of six proteins: three tightly bound "heavy chains" that make up the three legs, each associated with a smaller "light chain" [52]. Purified clathrin can self-assemble into its distinctive spherical lattice structure *in vitro* [51], facilitating the study of the physics and chemistry of the process in simplified systems with known components [8].

In order to carry out their functional role, these structures have to combine two seemingly incompatible properties. The binding energy of the subunits must be sufficiently strong to overcome the resistance of the membrane elasticity. But the structure also has to rapidly disassemble when the vesicle has been successfully created. Mutant cells that fail to remove the protein coats from new vesicles exhibit numerous pathologies. The cell uses stored chemical energy to resolve this impasse, via chaperone proteins that actively destabilize the coat.

CME thus provides a promising platform in which to test the utility of results like those of Chapter 2 for understanding the collective behavior of real biological subsystems. It will soon be possible to make the relevant measurements of relaxation rates and steady-state distributions in an artificially controlled environment, where all components are known and provided in controlled amounts.

The properties of these nonequilibrium steady states, however, form just one element of the CME phenomenon. Even if we understood all the characteristics of the disassembly transition, the question of the trigger mechanism would still remain. Which parameter changes in order to send the system over the threshold? And how is this parameter change coupled to the completion of the vesicle?

In this chapter, I present my contribution to an experiment designed to answer these questions. This experiment employed a purpose-built protein to indirectly monitor the concentrations of putative trigger molecules. Inferring the dynamics of these molecules from the observed behavior of the sensor protein required integrating a large quantity of background knowledge, control experiments, and reasonable but unverified assumptions. As these inference tasks become ever more complex, it is in-

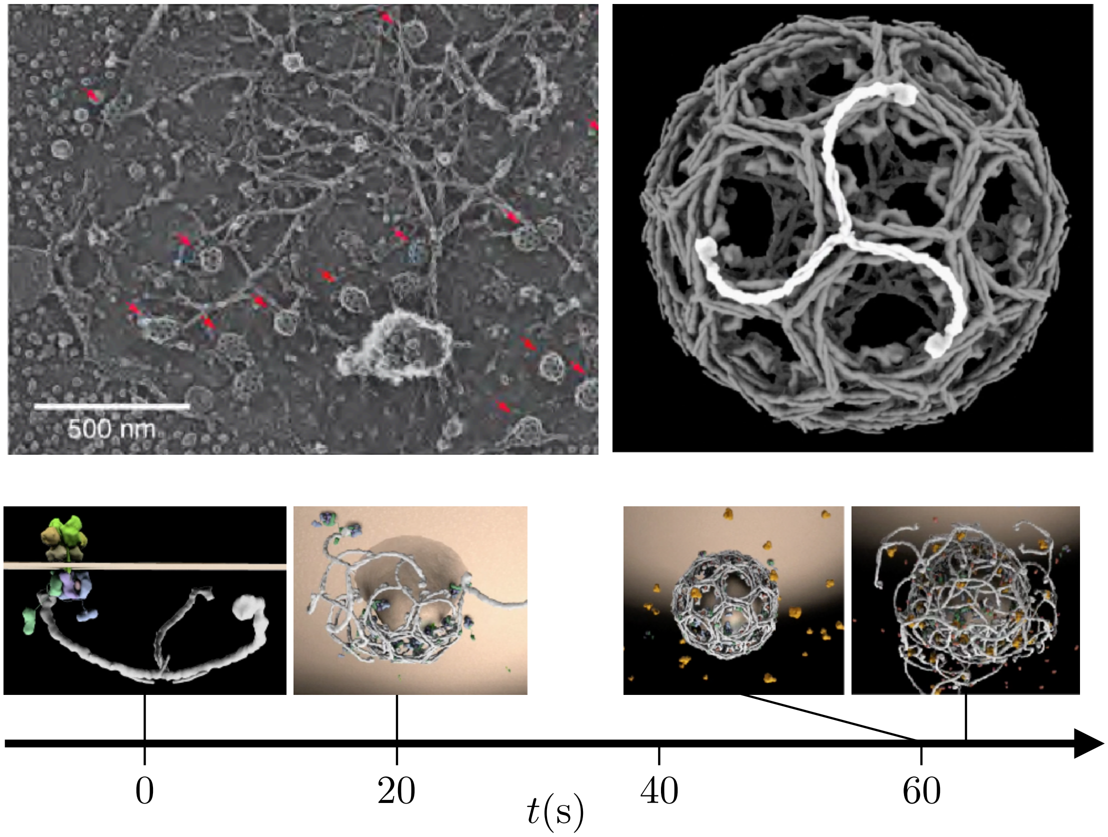

Figure 4-1: Color. Top left: Electron-microscope image of clathrin coats on the inner surface of the plasma membrane, reprinted from [42]. Top right: Structure of a complete clathrin coat, reprinted from [50]. Bottom: Dynamics of clathrin-mediated endocytosis. Clathrin triskelia bind to the plasma membrane and self-assemble into a lattice. Once the lattice is complete, the enclosed patch of membrane is severed from the bulk to create a topologically distinct closed surface of membrane, called a vesicle. Finally the clathrin coat disassembles, setting the vesicle free for further processing. The entire process takes about a minute. 3D renderings taken from the animation accompanying [50].

creasingly important to formalize the process in order to ensure that all assumptions are made explicit and to quantify the reliability of the result. My contribution to this experimental work on the clathrin system was precisely this quantitative formalization. After summarize the goals and methods of the experiment, I will explain how I translated the intuitive judgments of the biologists into a Bayesian framework, drawing on tools developed for systems biology research, and present the results of my analysis.

4.1 Biological Background

Before introducing the experiment, I need to set up the problem in more detail. In this section, I provide some key background information on the proteins involved in the disassembly process, and on the information-bearing phospholipids in the plasma membrane that are hypothesized to provide the triggering mechanism. Both aspects of clathrin disassembly are of significant biophysical interest in their own right, apart from their role in this particular process. The disassembly mechanism is generic, and operates on many other kinds of structures; and the special phospholipids I will be discussing are involved in almost every signaling pathway of eukaryotic cells [2].

4.1.1 Auxilin and Hsp70

As mentioned above, disassembly of the clathrin lattice after vesicle completion is powered by ATP hydrolysis. The enzyme that couples these two reactions together is a member of the Hsp70 family called Hsc70. Hsp70 enzymes break apart many kinds of protein structures, including dangerous aggregates that begin to form when proteins misfold at high temperatures – hence the name “heat shock protein.” Using an ATP-driven cycle through different conformations, these proteins can bind to their target with very high apparent affinity, displacing the other proteins that were bound to it in the structure, and then spontaneously dissociate from the target at a later point in the cycle [92, 93, 78].

At the concentrations actually maintained in the cytosol, the rate of Hsp70-

mediated disassembly is much lower than typical assembly rates, and structures are only minimally perturbed. For disassembly to occur, the local concentration of Hsp70 near the target must be raised by special proteins that bind Hsp70 on one end and the target on the other [92].

There are two proteins that accomplish this task for clathrin disassembly. For historical reasons, one of them is called auxilin (Aux) [97] and the other cyclin G-associated kinase (GAK) [48]. Once it became clear that GAK plays a similar functional role to auxilin, GAK received the more suggestive name auxilin 2, with the original auxilin now relabeled as auxilin 1. For the purpose of the present discussion, I will simply refer to both of them as auxilin.

Both these proteins contain three domains that are essential to their function in the disassembly process: a domain that binds to a pocket formed by three clathrin triskelia in an assembled coat, another that binds to the plasma membrane, and a third that binds to Hsc70. The name auxilin (from the Latin *auxilium* which means "help") comes from the fact that large quantities of auxilin actually *help* clathrin to assemble into coats *in vitro*, since they simultaneously bind to multiple subunits of the lattice and thus increase its stability.

It has been observed in previous experiments with fluorescently labeled auxilin that these molecules are absent from the clathrin structure as it nucleates and grows on the plasma membrane, and they suddenly arrive after the structure is complete, around the time the neck of the nascent vesicle pinches off [69]. This explains how the structure can avoid disassembling until the task of vesicle creation has been accomplished. But it opens a new question: how do these molecules "know" that they should arrive at this time?

4.1.2 Phosphoinositides and Fission Sensing

For the past decade, the Kirchhausen Lab has been proposing a plausible hypothesis to answer this question, based on the membrane-binding ability of auxilin [69, 32]. At physiological concentrations, the membrane-binding activity is essential to bring auxilin to the clathrin structure; the clathrin-binding interaction is too weak on its

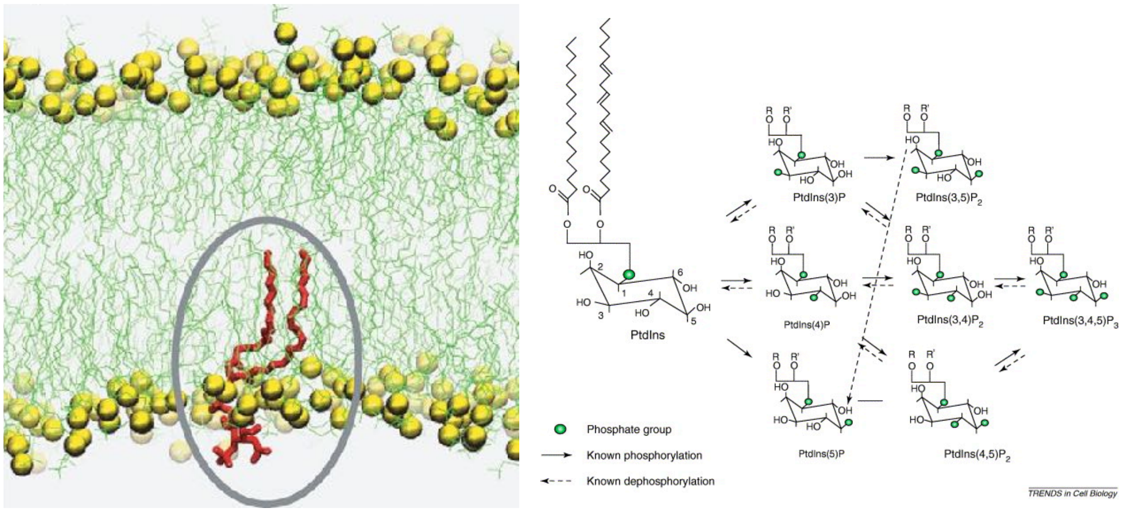

Figure 4-2: Color. Left: Frame from a simulation reported in [84] of the phospholipid bilayer that surrounds the cell, with a molecule of PI(4,5)P$_2$ highlighted in red. Right: PIP molecular structure, with network of conversion reactions catalzyed by known enzymes. Taken from [14].

own. When a truncated version of auxilin is added to cells, with the membrane-binding domain removed, the sudden recruitment after vesicle completion no longer occurs [69].

They hypothesize that the presence of the clathrin coat increases the local concentration of certain enzymes, which can specifically bind to the coat proteins. These enzymes catalyze the transfer of phosphate groups from ATP to specific sites on some of the phospholipid molecules that make up the membrane. As long as the nascent vesicle is still attached to the plasma membrane, the modified phospholipids rapidly diffuse out of the bud to the bulk of the membrane. Since the surface area of the bud is negligible compared to the area of the whole membrane (smaller by a factor of about 10,000), the concentration of modified phospholipids in the bulk membrane remains unaffected, and the enzymes act too slowly (compared to the diffusion rate) to pull the local concentration away from this value. A sudden change occurs when the vesicle pinches off from the plasma membrane. Phospholipids can no longer escape from the vesicle, and even a small number of slow enzymes can significantly change the membrane composition.

The standard phospholipid used by eukaryotic cells for this sort of membrane-based signaling is phosphatidylinositol (PI). Although PI makes up only one percent of the total lipid content of the plasma membrane [61], its unique structure makes it extremely important for information storage and transfer [2]. As shown in Figure 4-2, the region of PI exposed to the interior of the cell contains a ring of six carbon atoms, with hydroxyl groups hanging from five of the vertices, numbered 1-5 (the sixth connects the ring to the rest of the molecule). The cell can in principle transfer phosphates from ATP to any subset of these five sites, generating $5!\sum_{n=1}^{5}\frac{1}{(5-n)!n!} = 31$ possible distinct molecular species (in addition to the original PI), known as phosphoinositides (PIP). Only three of the five sites (3, 4 and 5) are used in known biochemical processes, so only seven of these possible types are actually found in cell membranes [2]. The standard notation for distinguishing these types denotes the phosphorylated sites (the sites with phosphates attached to them) in parentheses: PI(3)P, for example, is phosphorylated only on site 3.

The structure of the membrane-binding domains of auxilin suggests that they should interact with PIP's. The amino acid sequences of these domains are homologous to a known enzyme that can catalyze the removal of a phosphate group from at least three of the seven PIP's [37, 60]. A change in the active site of the enzyme abolishes the catalytic activity of the auxilin domains, but the 3D structure of the domain as determined by X-ray crystallography is still very similar to that of the enzyme, and should preserve affinity for some PIP species [32]. Biochemical assays qualitatively confirm this conjecture, but the affinity is extremely weak and difficult to quantify reliably [69, 32].

In light of this information, we can make the question about the disassembly trigger more specific: is the sudden recruitment of auxilin after vesicle fission caused by a change in PIP concentrations? The most satisfying answer to this question would come from suddenly blocking or enhancing the activity of the enzyme responsible for the concentration change, while recording the response of the auxilin dynamics. But before such an experiment can be performed, one must first establish that there *is* a concentration change. Ideally, one would also quantify the affinity of auxilin for PI and all seven PIP's, to determine whether the change in membrane composition is sufficient to drive its observed recruitment.

My main contribution to this project was to extract information about PIP concentrations in clathrin-coated membrane regions from an experiment designed and carried out by Dr. Kangmin He at the Kirchhausen lab. In Section 4.2 I describe the experiment and the goals of my quantitative analysis. After explaining some key elements of the initial data processing in Section 4.3, I present my inference procedure in Sections 4.4-4.5. Section 4.6 reports the results of this analysis, with comments on their significance and limitations.

4.2 Experimental Design

Detecting how many phosphates are attached to the carbon ring of a PIP and which positions they occupy seems a daunting task, especially if this must be done in real time inside living cells. What makes this measurement possible is a deep principle that undergirds much of post-genomic molecular biology. The biologically relevant features of a molecule are by definition the features that the cell can robustly recognize. And since the cell can only "recognize" something through the way it interacts with another molecule, we can design a sensor for any biologically relevant feature by harnessing this interaction.

In this case, the fact that PIP's are so important for information processing implies that many proteins in the cell can discriminate among them with high fidelity. For at least four of the seven PIP's, a naturally occurring protein has been successfully identified that binds only to that form and ignores the others. With routine techniques, one can take the DNA sequence that codes for any one of these proteins, append the sequence that codes for a green or red fluorescent protein with an appropriate linker, and insert a ring of DNA containing this sequence into a cell. The cell's membranes will now fluoresce with an intensity proportional to the local concentration of the given PIP. This technique has been used to show that different PIP's are typically localized to different kinds of membranes (plasma membrane, Golgi complex, endosomes, etc.), apparently helping the cell distinguish these structures from one another

[70].

In these experiments, the plasma membrane always appears with a relatively uniform fluorescence intensity. The sudden local change in concentration predicted to occur at the moment of vesicle fission is not observed. This does not rule out the existence of such changes, however, because the small size of the clathrin-coated vesicles demands an extremely sensitive measurement. The vesicles are typically about 50 nm in diameter, while the wavelength of the light used to visualize the PIP sensors is at least 500 nm. This means that the signal of interest gets smeared out by diffraction over an area 100 times larger than the size of the source. To detect a given concentration change in a coated vesicle, the measurement has to be 100 times more sensitive than would be required to measure the same concentration difference in the bulk of the plasma membrane.

4.2.1 Auxilin-based Sensors

According to our hypothesis, however, auxilin is capable of robustly reporting this local concentration change. Cells expressing fluorescent versions of these proteins display clearly visible flashes of localized fluorescence intensity [69]. So Kangmin designed a new sensor based on fluorescent auxilin by replacing the membrane-binding domain with one of the established PIP sensors (and removing the domain that binds to Hsc70 to avoid accidentally activating the uncoating process too early). The clathrin-binding ability of the auxilin molecule would combine cooperatively with the PIP-binding capacity of the original sensor, thereby amplifying the signal.

When he expressed these sensors in live cells, bright dots appeared scattered over the whole plasma membrane, as shown in Figure 4-3. These spots appeared at the same locations as fluorescent clathrin, which was observed simultaneously using a different color. Remarkably, the spots showed qualitatively different behavior over time depending on which PIP sensor was used, as shown in Figure 4-3. The PI(4,5)P$_2$ sensor, for example, gradually grew brighter as the clathrin coat assembled, and then rapidly disappeared after the clathrin fluorescence reached its maximum value. The PI(3)P sensor, on the other hand, behaved almost identically to the original auxilin, briefly flashing at the moment of vesicle fission.

As documented in [38], a large set of control experiments were performed to test the specificity of these sensors and verify that they are reporting on the intended PIP's. One particularly important experiment addressed a referee's concern that these sensors are based on the same protein whose behavior is to be explained. If some unknown factors independent of PIP signaling are actually producing the auxilin recruitment, then those same factors could be responsible for the supposed sensor readout. This concern is ameliorated somewhat by the fact that the sensor for PI(4,5)P_2 behaves in a way that is dramatically different from the original auxilin. But to perform a more decisive test, Kangmin built a new sensor using the clathrin-binding domain of an unrelated protein called epsin, which does not show any sudden spikes of recruitment. The results gave fresh confidence that everything was working as expected: the epsin-based sensor for each PIP showed the same qualitative behavior as the corresponding auxilin-based sensor.

4.2.2 Quantification

The images in Figure 4-3, when accompanied by the control experiments, convincingly establish that local PIP concentrations change suddenly at the moment of vesicle fission. In principle, no additional quantitative analysis is needed to answer this basic biological question. But quantification is desirable for at least three reasons. First, it allows us to attempt a more detailed inference, extracting the speed, size and duration of the change. This information will provide a basis for judging whether measured differences in affinities of auxilin for the seven PIP's are sufficient to account for their recruitment dynamics.

To identify the enzymes responsible for the concentration change, it would also be helpful to have an estimate of how many copies of the enzyme we should expect to find in each coat. When coupled with known rates for the relevant kinds of enzymes, this also tells us whether the diffusion-based mechanism for generating the sudden concentration change is still plausible.

Finally, the affinities of the original PIP sensor proteins for their binding partners

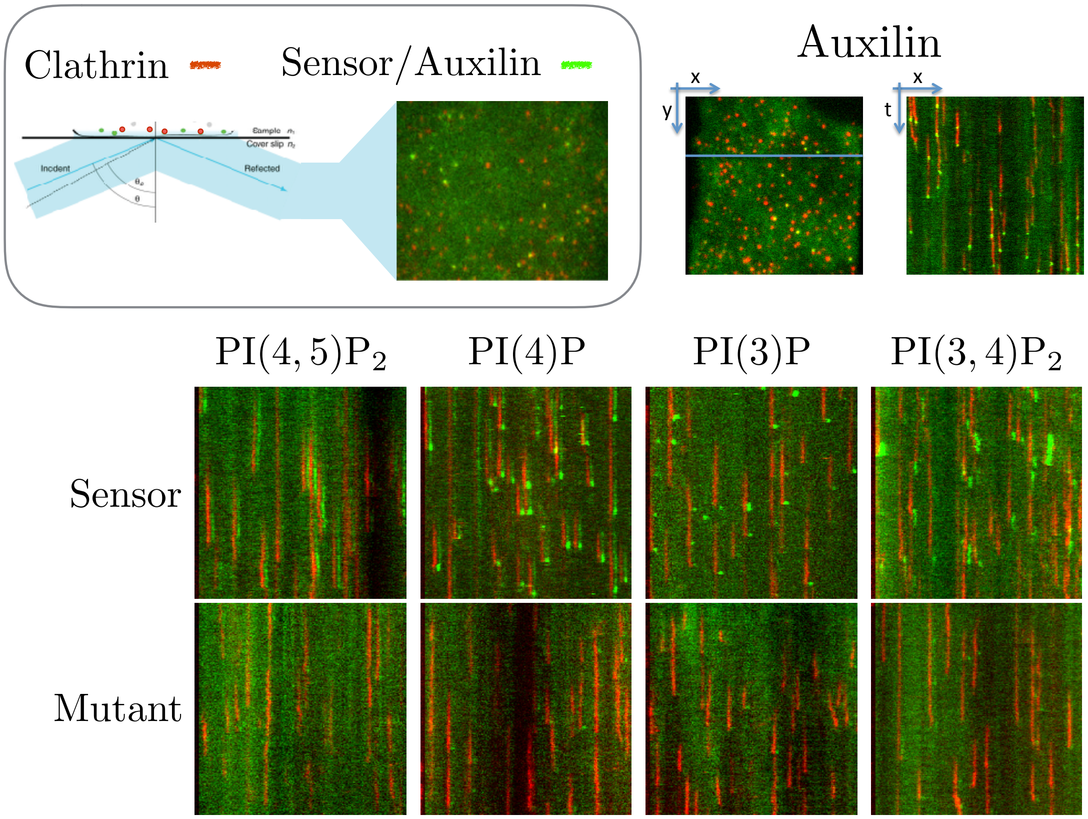

Figure 4-3: Color. Top left: Cells were gene-edited to attach red fluorescent proteins to all clathrin triskelia. Auxilin or auxilin-based sensors were attached to green fluorescent proteins. Total Internal Reflection Fluorescence (TIRF) microscopy was employed to collect sequential exposures of the fluorescent emission in each color channel from the ~500 nm-thick region near the surface of the cell illuminated by the evanescent field of the excitation laser. Top right: Data from an experiment with green fluorescent auxilin. A convenient way to collapse a movie to a single image is to draw a line through the imaged region and display the fluorescence intensities along this line for all frames in a movie. In the resulting image, called a kymograph, the vertical axis is time, increasing from top to bottom. Bottom: Kymographs from microscope movies of auxilin-based sensors for each of the four tested PIP's. Kymographs from a control experiment is also included, with identical conditions except that the PIP-binding domain of the sensor was mutated. The red channel has been slightly shifted along the x axis relative to the green channel, for easier visualization of both.

(without the added auxilin fragment) have been measured *in vitro*. We can use our estimate of PIP concentration changes to determine whether these published values are consistent with the absence of detectable signal in the original sensor experiments, and with the behavior of the auxilin-based sensors. It is generically expected that *in vitro* affinity measurements should differ from affinities because of crowding, enzyme activity, etc., and this experiment gives us an opportunity to quantify that difference in a concrete case.

The first step in this analysis, illustrated in Figure 4-4, is to convert the microscope movies into a set of trajectories that give the fluorescence intensities of clathrin and the sensor as a function of time for each clathrin assembly event. Then the trajectories need to be aligned and placed on the same time axis, so that the mean and standard error can be obtained for each time point. This step raises a number of important difficulties, which will be addressed in Section 4.3. The resulting traces provide the starting point for inferring PIP concentrations and enzyme copy numbers using the model of sensor kinetics presented in Section 4.4.

The initial quantification of the microscope movies was performed with a MATLAB script developed in the Kirchhausen lab for this purpose several years ago [1]. Because the clathrin-coated regions are so much smaller than the wavelength of the light used for imaging, each coat shows up on the microscope as a bright spot with an approximately Gaussian profile, whose width is set by the wavelength. The total fluorescence intensity emitted by this spot reflects the number of fluorescent molecules contained in it. Identifying and tracking such Gaussian dots over the course of the movie is a routine computational task, and the core of the script is based on standard algorithms. The first panel of Figure 4-5 shows the distributions of lifetimes of automatically detected objects under different experimental conditions. The short-lifetime regime contains other kinds of events in addition to the endocytosis process I am trying to study, including “abortive” clathrin coats that dissolve before they produce a new vesicle, and clathrin coats or clusters floating in the cytosol that temporarily enter the field of view. In the original paper for which the MATLAB script was developed, considerable attention was dedicated to screening out these events in

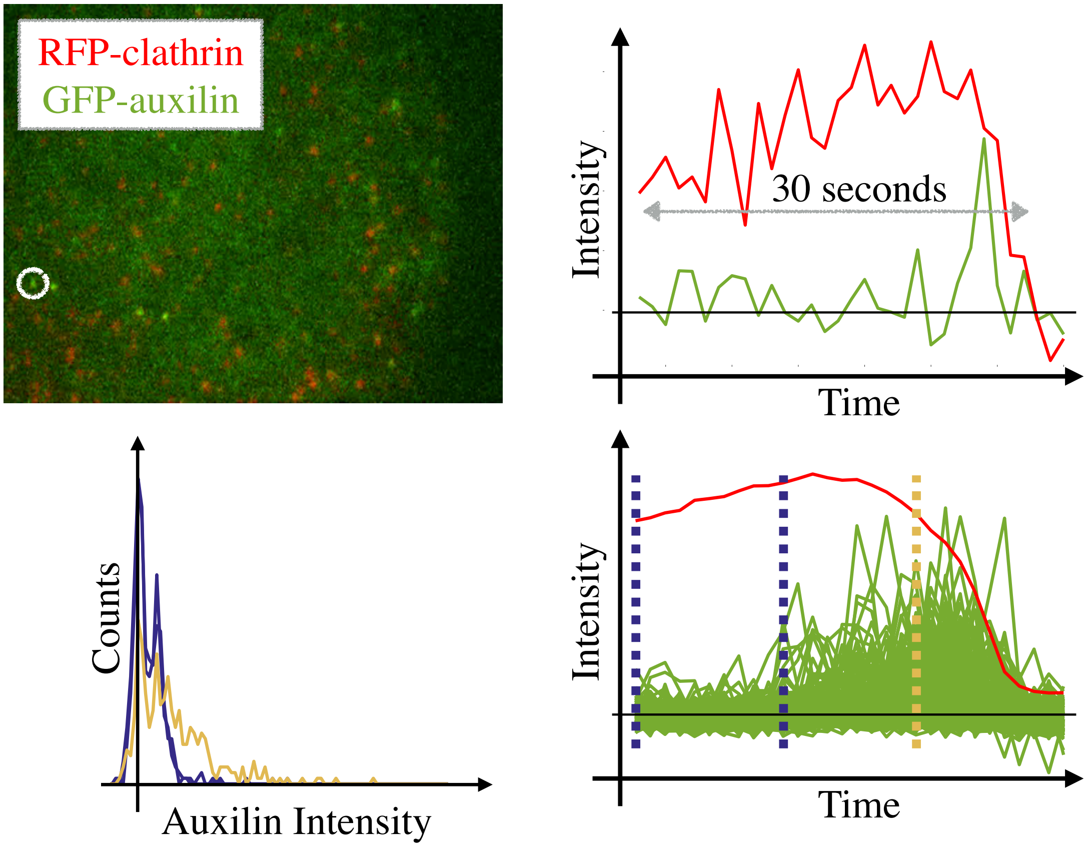

Figure 4-4: Color. Top left: Diffraction-limited spots of fluorescent clathrin are automatically detected in each frame of the microscope movies, and the total fluorescence intensities emanating from the clathrin and from the sensor in each spot are separately recorded. Top right: The spots are linked together from frame to frame using standard algorithms, generating a trajectory of the fluorescence intensity variations over the lifetime of each spot. Bottom right: Collecting all the trajectories from a given set of experimental conditions generates a statistical ensemble. Bottom left: Histograms of fluorescence intensity from the ensemble of sensor trajectories in the bottom right panel, taken at the three time points indicated by vertical dotted lines. The first two time points show the same distribution, indicating that the ensemble has reached a steady state. The mean number of sensor molecules in each object is less than one, and a peak is clearly visible at the average fluorescence intensity of a single molecule. After vesicle completion, when the clathrin fluorescence starts to decreases, the distribution changes.

order to obtain fully reproducible lifetime distributions. For separate reasons, described below, I only consider events that have lifetimes within the typical range for the standard assembly/disassembly cycle of 60-80 seconds, which automatically excludes these kinds of objects without any additional statistical tests.

4.3 Data Interpretation and Averaging

In Section 4.4 I will write down differential equations for the evolution of two quantities $N_C(t)$ and $N_S(t)$, which are to be compared with the fluorescence intensity data for clathrin and a sensor protein, respectively, averaged over some subset of detected events under a given set of experimental conditions. There are several good reasons for skepticism about the meaningfulness of this comparison, some due to my modeling assumptions and some to the perennial challenge of biological heterogeneity. In this section I summarize these challenges and describe how I addressed them.

4.3.1 Deterministic Modeling of Stochastic Processes

The first problems come from my use of deterministic chemical kinetics. The mean number of sensor molecules in a given coated structure is usually quite low in the cells chosen for analysis, so as to avoid interference with the natural dynamics. In this low number regime, the difference between the predictions of deterministic chemical kinetics and of the full Master Equation can be significant. Furthermore, the only way to incorporate vesicle fission into these deterministic equations is to suddenly change some of the clathrin and phosphoinositide kinetic parameters at a given time $t = t_{\text{fiss}}$. But fission is really a stochastic event, and averaging many traces with different t_{fiss} values smears out the features of the line shapes.

The first issue is alleviated by the fact that the Master Equation and deterministic chemical dynamics agree even at low numbers when the deterministic ODE's are linear. The only nonlinearities in the equation for the sensor dynamics come from saturation of binding sites, when the bound sensors take up a significant fraction of the total number of binding sites in the coat or on the membrane. But if the binding

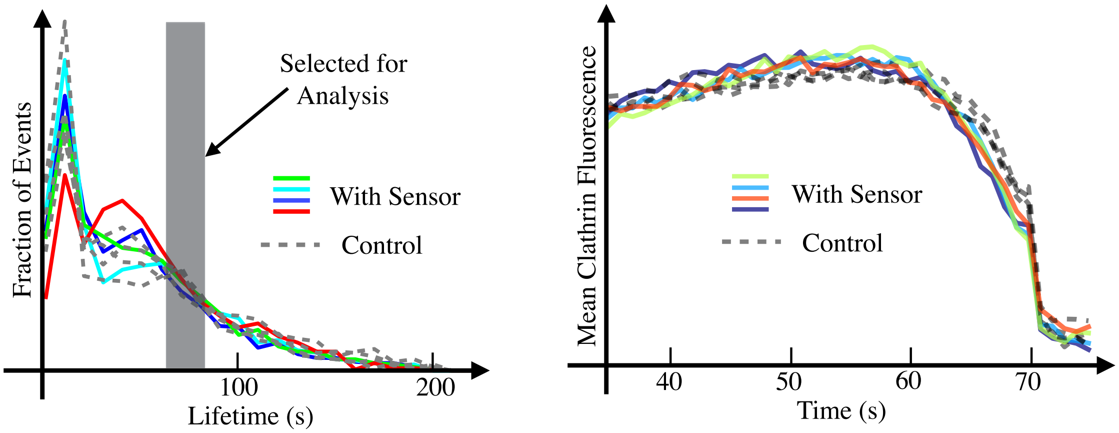

Figure 4-5: Color. Left: Distribution of lifetimes of automatically detected clathrin spots. Solid traces are from cells expressing the full auxilin-based sensor proteins, and dotted lines are from control cells that only express the PIP-binding part. To reduce the variability in clathrin coat sizes for the model fitting and analysis, I only used events within the indicated narrow window of lifetimes. Right: Clathrin fluorescence vs. time, averaged over all the events in the lifetime window from the left panel, for each of the sensor-expressing cells (solid lines) and the control cells (dotted lines).

sites were saturated, this would interfere with the native auxilin dynamics, which is what the low expression levels were supposed to prevent, and thereby perturb the clathrin uncoating process.

In Figure 4-5, I compare the average clathrin dynamics with and without sensor binding, and show that the presence of the sensor has a minimal effect. In the first panel, I plot the distribution of lifetimes of all the automatically detected clathrin spots from the microscope movies, with solid lines for the cells expressing the four sensors, and dotted lines for the cells expressing just the PIP-binding part. As discussed above, the PIP-binding part of the sensor does not specifically associate with the clathrin spots when expressed alone, and thus provides a good control that should not perturb the dynamics. At short lifetimes, the distributions differ from one another, possibly due to differences in the ratio of “abortive” clathrin coats to productive endocytosis events in the different experimental conditions. But for the longer lifetimes that contain only productive events, the distributions all agree up to experimental uncertainty.

As for the second problem, since I am primarily interested in what happens after fission, the cleanest way to address the stochasticity of t_{fiss} would be to use it as the basis for aligning the trajectories prior to averaging. In the absence of a direct measurement of t_{fiss}, the next best thing is to align based on the end of the event, which happens when the clathrin signal falls below the threshold for a significant event detection. This will clearly distort the beginning of the trajectory because of the range of event lifetimes, so I truncated the average data and kept only the final 40 frames of each trajectory for analysis.

I experimented with different alignment techniques: selecting trajectories with similar shapes and aligning based on least-squares minimization using the whole trajectory, using the peak of the post-scission sensor bursts, or using the point where the derivative of the trajectory becomes most negative. None of these methods were clearly superior to simply aligning the ends of the traces, but they complicate the analysis and makes the results less robustly reproducible.

The second panel of Figure 4-5 shows the averaged clathrin fluorescence signals after alignment based on the end of the event. I have again separately plotted the data from the experiments with each of the sensors and each of the control proteins, and have normalized the fluorescence intensity by dividing each trace by its mean value. The clathrin dynamics are very similar across all conditions.

4.3.2 Biological Heterogeneity

Additional problems arise when we consider that the model parameters are not necessarily the same for all cells or even for all coats within a single cell.

The canonical example of such heterogeneity in biology is variability in protein expression levels, which affects the model parameter N_{S0}. This natural heterogeneity becomes still more significant with proteins produced by transient transfection with a foreign plasmid, since the number of copies of the plasmid cannot be precisely controlled.

The linearity of the sensor dynamics discussed above alleviates this problem. In the limit where the differential equations for the sensor kinetics are approximately linear, it is easy to check that changing N_{S0} merely scales the sensor trajectories

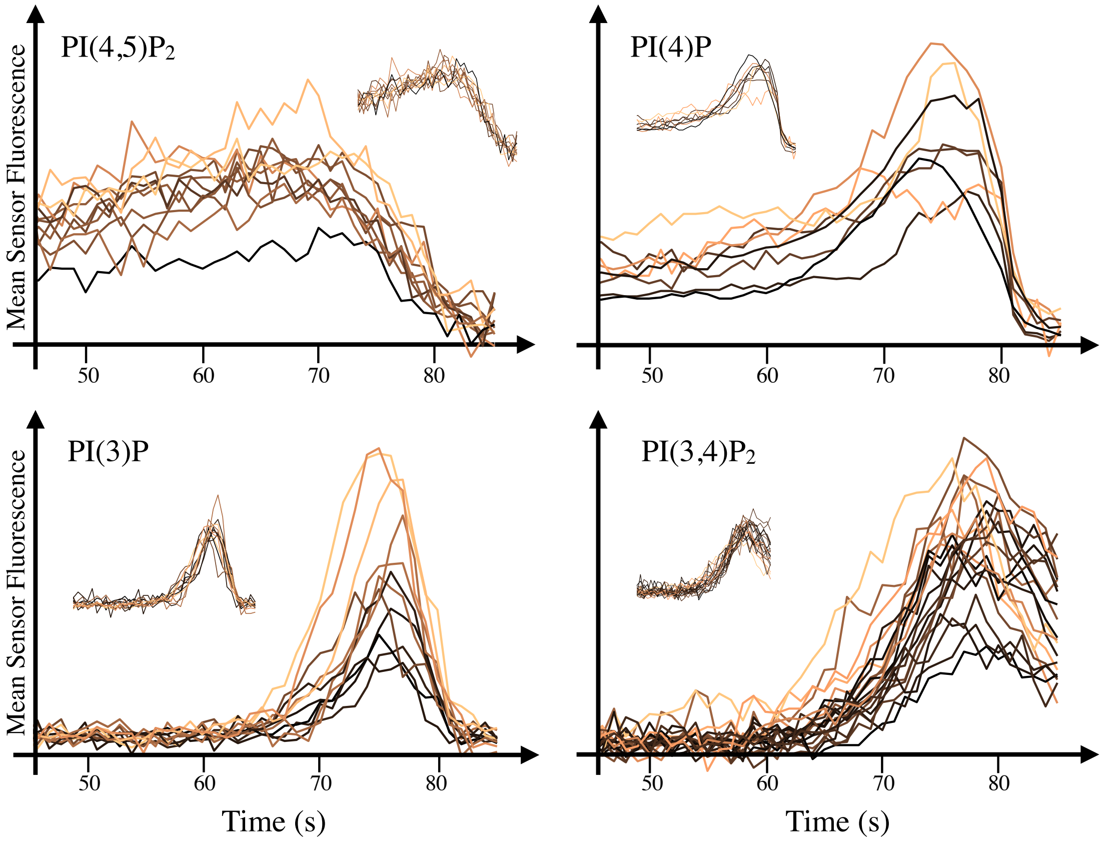

Figure 4-6: Color. Sensor fluorescence measurements were made on several cells for each sensor. Each trace in this plot represents the averaged fluorescence trajectory in a single cell, using the selection and alignment procedures described in Section 4.3. Color indicates relative expression level of sensor protein, with lighter colors corresponding to higher cytosolic sensor concentrations. Inset contains the same data after linearly rescaling the trajectories by dividing each one by its mean.

by a constant factor, and does not affect the shape. This means that the equations of Section 4.4 should still describe the sensor dynamics, with N_{S0} replaced by the average $\langle N_{S0} \rangle$ over all the cells in the dataset.

There is another kind of heterogeneity specific to this problem, which is the variability in the size of the completed vesicle. This variation affects the clathrin parameters that determine k_a and k_u, as well as the initial phosphoinositide numbers. Since the final vesicle size is tightly correlated with the lifetime of the coat formation event, I reduced the size range by only analyzing data from events with lifetimes between 60 and 80 seconds, as indicated in Figure 4-5.

The dynamics of each sensor were observed in multiple cells, allowing an empirical assessment of the overall variability from cell to cell. In Figure 4-6, I average the sensor fluorescence over all the 60-80 second events from each of the cells, after alignment based on the end of the trajectory as described above. Each cell contained between 7 and 67 events within this lifetime window. The traces are colored by the relative expression level of the sensor protein, as quantified by the background fluorescence in each cell, with lighter traces corresponding to higher sensor concentrations. The effect of the expression level variation is particularly visible in the PI(3)P sensor, which shows a wide range of peak heights correlated with the line color. But dividing each trace by its mean to normalize the intensities, as shown in the inset, causes all these traces to agree reasonably well.

4.4 Kinetic Model

The core of my procedure for inferring PIP dynamics from the sensor readings is a model of the sensor kinetics, which combines the conclusions of the control experiments and additional background knowledge into a single set of formulas.

The model describes the dynamics of three subpopulations of sensor molecules, as depicted in Figure 4-7: a sensor protein in the coat can be bound to clathrin, to a specific phosphoinositide binding partner in the cell membrane, or to both. I denote the mean number of sensor molecules in each state, averaged over many identically prepared coats, as N_{SC}, N_{SL} and N_{SLC}, respectively. The mean number of clathrin triskelia in the coat is $N_C = (N_{C0} + N_{SC} + N_{SLC})/3$, and the mean copy number of the relevant phosphoinositide in the coated region is $N_L = N_{L0} + N_{SL} + N_{SLC}$. N_{C0} and N_{L0} represent the number of free binding sites in the clathrin lattice and the membrane, and the factor of 3 comes from the fact that the clathrin lattice contains three binding sites per triskelion.

As discussed above in Section 4.3, I assume that the clathrin number N_C and the total number of sensor molecules $N_S = N_{SC} + N_{SL} + N_{SLC}$ are proportional to the suitably averaged fluorescence signals observed in the experiment. My goal is to infer the temporal dynamics of the unobservable quantity N_L from observations of

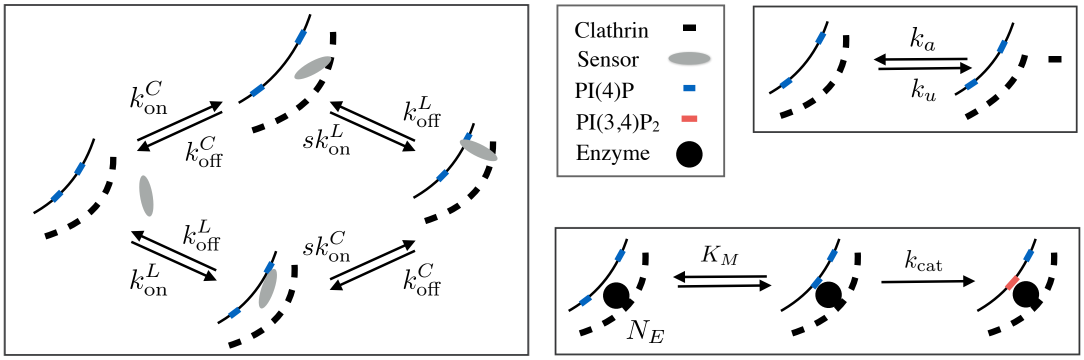

Figure 4-7: Color. Left: Model of sensor protein kinetics. Sensor can bind to its specific PIP substrate in the membrane and to the clathrin lattice. Binding to one speeds up binding to the other by a factor s. Top right: Model of clathrin kinetics. Association rate k_a and dissociation rate k_u are both functions of the current number of triskelia in the lattice N_C. Bottom right: Michaelis-Menten enzyme kinetics. The Michaelis constant K_M is the concentration of substrate molecules at which half of the enzymes are substrate-bound in the limit of low enzyme concentration.

N_C and N_S.

4.4.1 Sensor Kinetics

I model the kinetics of sensor binding as follows. Cytosolic sensor proteins, at concentration N_{S0} molecules per unit volume, can enter the coat by binding either to clathrin or to its phosphoinositide partner, with mean rates $k_{\text{on}}^C N_{C0}$ and $k_{\text{on}}^L N_{L0}$, respectively. Once one of the sensor domains is bound, its spatial confinement enhances the on-rate for the other domain, so that membrane-bound sensor binds clathrin at a higher rate $sk_{\text{on}}^C N_{C0}$ and clathrin-bound sensor likewise binds the membrane at rate $sk_{\text{on}}^L N_{L0}$. The clathrin-binding and membrane-binding domains can unbind independently of each other, with mean rates k_{off}^C and k_{off}^L respectively. The assumption of independent unbinding forces the speedup factor s to be the same for both membrane and clathrin on-rates. Finally, the clathrin-binding domain can fall off due to dissociation of the sensor-binding pocket when a clathrin triskelion leaves the coat, which happens at rate k_u.

The kinetics of truncated auxilin (with the membrane-binding domain removed) and of the phosphoinositide sensor proteins alone (without the clathrin-binding domain) are modeled the same way, but with the on-rates to membrane or clathrin set to zero.

Thus for a given sensor, the mean number of copies in each state in a single coat evolves according to the following set of coupled differential equations:

$$\begin{aligned}
\frac{dN_{SC}}{dt} &= k_{\text{on}}^{C} N_{C0} N_{S0} + k_{\text{off}}^{L} N_{SLC} - (k_{\text{off}}^{C} + s k_{\text{on}}^{L} N_{L0} + k_u) N_{SC} && (4.1)\\
\frac{dN_{SL}}{dt} &= k_{\text{on}}^{L} N_{L0} N_{S0} + (k_{\text{off}}^{C} + k_u) N_{SLC} - (k_{\text{off}}^{L} + s k_{\text{on}}^{C} N_{C0}) N_{SL} && (4.2)\\
\frac{dN_{SLC}}{dt} &= s k_{\text{on}}^{L} N_{L0} N_{SC} + s k_{\text{on}}^{C} N_{C0} N_{SL} - (k_{\text{off}}^{C} + k_{\text{off}}^{L} + k_u) N_{SLC} && (4.3)
\end{aligned}$$

The PI(3)P and PI(3,4)P$_2$ sensors each include two copies of the lipid-binding domain, and so the reaction network has to be expanded to include two new species: N_{SLL} and N_{SLLC}. The expanded set of equations can be obtained from the above three equations by adding an extra lipid-binding reaction at rate $s_L k_{\text{on}}^{L} N_{L0}$, and a corresponding dissociation reaction at rate k_{off}^{L}:

$$\begin{aligned}
\frac{dN_{SC}}{dt} &= k_{\text{on}}^{C} N_{C0} N_{S0} + k_{\text{off}}^{L} N_{SLC} - (k_{\text{off}}^{C} + s k_{\text{on}}^{L} N_{L0} + k_u) N_{SC} && (4.4)\\
\frac{dN_{SL}}{dt} &= k_{\text{on}}^{L} N_{L0} N_{S0} + (k_{\text{off}}^{C} + k_u) N_{SLC}\\
&\qquad - (k_{\text{off}}^{L} + s k_{\text{on}}^{C} N_{C0} + s_L k_{\text{on}}^{L} N_{L0}) N_{SL} + k_{\text{off}}^{L} N_{SLL} && (4.5)\\
\frac{dN_{SLL}}{dt} &= s_L k_{\text{on}}^{L} N_{L0} N_{SL} + (k_{\text{off}}^{C} + k_u) N_{SLLC} - (k_{\text{off}}^{L} + s k_{\text{on}}^{C} N_{C0}) N_{SLL} && (4.6)\\
\frac{dN_{SLC}}{dt} &= s k_{\text{on}}^{L} N_{L0} N_{SC} + s k_{\text{on}}^{C} N_{C0} N_{SL}\\
&\qquad - (k_{\text{off}}^{C} + k_{\text{off}}^{L} + k_u + s_L k_{\text{on}}^{L} N_{L0}) N_{SLC} + k_{\text{off}}^{L} N_{SLLC} && (4.7)\\
\frac{dN_{SLLC}}{dt} &= s_L k_{\text{on}}^{L} N_{L0} N_{SLC} + s k_{\text{on}}^{C} N_{C0} N_{SLL} - (k_{\text{off}}^{C} + k_{\text{off}}^{L} + k_u) N_{SLLC}. && (4.8)
\end{aligned}$$

4.4.2 Phosphoinositide Kinetics

If all the kinetic parameters were known, one could infer $N_L(t)$ directly from these equations and the observed signals $N_S(t)$ and $N_C(t)$. But as in most *in vivo* biosensor inference tasks, many parameters are not known. This is partially due to the difficulty of measuring weak interactions in standard biochemical assays. But the fundamental problem lies in the fact that the parameters of any simple model of *in vivo* kinetics are not fully determined by intrinsic properties of the molecules involved. In the crowded environment of the cytosol and cell membrane, the molecules of interest do not form a closed system, but in fact interact with thousands of other molecular species that are not explicitly modeled. The parameters of the model are thus "effective" parameters, which implicitly include the effect of all these extraneous interactions. The membrane affinities $k_{\text{on}}^L/k_{\text{off}}^L$, for example, have been measured *in vitro* by a variety of different techniques, but as we will see in Table 4.2, the reported values differ by up to an order of magnitude from the effective affinity in the cell.

In the absence of known parameter values, the above equations define a family of curves $N_L(t)$ compatible with a given data set $[N_S(t), N_C(t)]$. Extracting and analyzing this family is much easier if we first restrict the space of allowed $N_L(t)$ lineshapes to a subspace of finite dimension. A mathematically natural way would be to represent $N_L(t)$ as a Fourier series, and truncate the series at the highest physically allowed frequency. But the space of Fourier coefficients in this case would still have a very high dimension, and would include many shapes (e.g., rapid oscillations) that are *a priori* implausible. To obtain at least a rough sense of the phosphoinositide dynamics compatible with the data, I used physically motivated assumptions to restrict $N_L(t)$ to a space of only a few dimensions. This restriction represents an additional assumption, which rules out some lineshapes that would otherwise be allowed by the data.

Specifically, I assume that the PIP's are generated by enzymes acting on preexisting membrane components, and removed by additional enzymes, all at fixed concentration and with Michaelis-Menten kinetics. The turnover rate for each reaction is determined by the number of enzymes N_E in the coated vesicle, the turnover number k_{cat} and the Michaelis constant K_M, as illustrated in Figure 4-7. This results in the following differential equations for the mean copy numbers N_L, N_L' of each sensor-binding phosphoinositide species and its precursor, respectively, in the coated vesicle:

$$\frac{dN_L}{dt} = k_{\text{cat}} N_E \frac{N'_L}{K_M + N'_L} \tag{4.9}$$

$$\frac{dN'_L}{dt} = -k_{\text{cat}} N_E \frac{N'_L}{K_M + N'_L}. \tag{4.10}$$

My model of clathrin-associated phosphoinositide conversions includes four reactions, each modeled by a set of equations like (4.9-4.10):

$$\text{PI} \rightarrow \text{PI(3)P} \tag{4.11}$$

$$\text{PI}(4,5)\text{P}_2 \rightarrow \text{PI(4)P} \tag{4.12}$$

$$\text{PI(4)P} \rightarrow \text{PI}(3,4)\text{P}_2 \tag{4.13}$$

$$\text{PI(3)P} \rightarrow \text{PI}(3,5)\text{P}_2. \tag{4.14}$$

There are only four reactions because PI(4,5)P_2, PI(4)P and PI(3,4)P_2 form a single chain of transformations. Each reaction is characterized by two independent parameters, $k_{\text{cat}} N_E$ and K_M.

Before vesicle fission, all new lipid species generated rapidly diffuse out of the coat. Assuming a diffusion coefficient of 5.4 $\mu\text{m}^2/\text{s}$ as reported in [28] for cell membranes on sub-200nm length scales, I estimate that the mean time for a new molecule to leave a half-completed 70nm-diameter vesicle is about 1 ms, whereas the typical turnover time for a lipid-modifying enzyme is about 20 ms. Since there are typically only a few copies of an enzyme in the coat, and certainly far fewer than 20, finding a pit with even a single modified lipid still inside it should be a rare event. I include this effect in my parameterization by keeping all enzyme activity turned off ($k_{\text{cat}} = 0$) until the moment of vesicle fission.

I constrained the pre-fission phosphoinositide levels to be consistent with known measurements of plasma membrane composition [41, 61]. I set the PI(3)P and PI(3,4)P_2 concentrations to be essentially zero (fixed at 0.001 molecules per coated region), since they are not supposed to be present in the plasma membrane. The initial PI, PI(4)P and PI(4,5)P_2 concentrations were left as free parameters, but constrained as described in Section 4.5 to have the right order of magnitude.

4.4.3 Clathrin Kinetics

Although the clathrin trajectory $N_C(t)$ can be obtained directly from the data, without any modeling, I had to make some assumptions about the clathrin kinetics in order to fix the rate k_u that appears in the equations for the sensor dynamics. The data can tell us the net rate of change of the clathrin copy number, but not how that rate is partitioned between association and dissociation rates. It should be possible to obtain this decomposition from the fluctuations in individual trajectories or from photobleaching experiments, but some technical challenges still remain before these techniques can yield reliable results in this system.

The equation for the clathrin dynamics can be written without loss of generality as

$$\frac{dN_C}{dt} = k_a(N_C) - k_u(N_C)N_C \qquad (4.15)$$

where both k_a and k_u can depend on N_C, and $k_u(N_C)$ is the same quantity that appears in the sensor dynamics.

I now choose $k_u(N_C)$ by setting both $k_a(N_C)$ and $k_u(N_C)$ to be linearly decreasing functions of N_C. This should be a good approximation in the regime where the coat is nearly complete, and is also a reasonable approximation for any N_C after fission. The off-rate k_u should decrease as the coat grows, because clathrin is primarily added and removed from the edge of the coat, and the perimeter-to-area ratio is a monotonically decreasing function of area. When the coat is more than halfway complete, the on-rate k_a should also decrease, because the perimeter itself is a decreasing function of N_C.

With these assumptions, the rates can be written as:

$$k_u = k_u^{0/1}\left(1 - \frac{N_C}{M_u^{0/1}}\right) \tag{4.16}$$

$$k_a = k_a^{0/1}\left(1 - \frac{N_C}{M_a^{0/1}}\right) \tag{4.17}$$

where the parameters $k_u^0, M_u^0, k_a^0, M_a^0$ govern the kinetics before fission, and are changed to $k_u^1, M_u^1, k_a^1, M_a^1$ after fission. I chose these eight parameters to obtain the best fit to the averaged clathrin fluorescence data.

4.4.4 Summary of Modeling Assumptions

All the assumptions contained in the above equations can be reduced to the following four groups:

- Sensor kinetics: two domains, cooperative binding, independent unbinding; only difference among sensors is off-rate from lipid
- Clathrin kinetics: phenomenological trajectory with kink at fission, linearly decreasing rates
- Phosphoinositide kinetics: Michaelis-Menten conversion rates turn on at fission
- Correspondence rule: deterministic mass action kinetics describes evolution of mean copy numbers.

4.5 Data Fitting and Sensitivity Analysis

Under the assumptions of Section 4.4, the task of inferring the most likely phosphoinositide dynamics is reduced to an optimization problem. I can think of the model as a function $\mathbf{y}(\boldsymbol{\theta})$ that maps the 24-dimensional vector $\boldsymbol{\theta}$ containing the model parameters $\{k_{\text{on}}\}, \{k_{\text{off}}\}, \{s\}, \{N_E\}, \{K_M\}, \{N_{S0}\}, \{N_L(0)\}$ to a vector $\mathbf{y}$ containing the predicted mean fluorescence intensities at each time point for all four sensor proteins and their truncated control versions. The observed mean fluorescence intensities are represented by another vector $\mathbf{d}$ that lives in the same space as $\mathbf{y}$. By minimizing the norm of the vector $\mathbf{d} - \mathbf{y}(\boldsymbol{\theta})$, I can identify a vector $\boldsymbol{\theta}^*$ that provides the best fit to the data. The inferred phosphoinositide dynamics $N_L(t)$ are then found by integrating the relevant kinetic equations from Section 4.4 using the parameters $\{N_E\}, \{K_M\}, \{N_L(0)\}$ from $\boldsymbol{\theta}^*$.

But this result is not really meaningful without some quantification of uncertainty. This is especially important in large kinetic models like this one, which are known to generically contain "sloppy" directions in parameter space along which the norm of $\mathbf{d} - \mathbf{y}(\boldsymbol{\theta})$ is practically constant [34, 96]. Steady-state concentrations, for example, typically depend on ratios of creation and degradation rates, and the individual parameters can take on any value as long as this ratio agrees with the data.

Software packages are now available that automatically and efficiently quantify the uncertainty in a high-dimensional least-squares fit. But these algorithms must be used with care, making sure that the underlying assumptions actually apply to a given problem, and that the results are interpreted correctly. In this section, I review the reasoning behind one of these algorithms, written by Ryan Gutenkunst during his PhD at Cornell [77, 33] and describe how I applied it.

4.5.1 Bayesian Statistics

I want to obtain the probability distribution $p(\boldsymbol{\theta}|\mathbf{d})$ quantifying the likelihood that a given set of parameters $\boldsymbol{\theta}$ correctly describes the system, given a set of noisy observations $\mathbf{d}$. More precisely, I want to study $p(\ln \boldsymbol{\theta}|\mathbf{d})$, where the logarithm automatically enforces the requirement that all the parameters be positive.

The conditional distribution $p(\ln \boldsymbol{\theta}|\mathbf{d})$ can be broken up into two factors using Bayes' Rule: $p(\ln \boldsymbol{\theta}|\mathbf{d}) \propto p(\mathbf{d}|\ln \boldsymbol{\theta})p(\ln \boldsymbol{\theta})$. The first factor is the probability of observing the given data $\mathbf{d}$, given that $\boldsymbol{\theta}$ is the true vector of parameters. To compute this factor, I model the noise as a vector of independent Gaussian random variables $\boldsymbol{\xi}$ (one for each data point) with standard deviations $\boldsymbol{\sigma}_\xi$. It will be notationally convenient to define a diagonal covariance matrix Σ_ξ whose diagonal elements are the squares of the standard deviations. Then the probability of making the observations $\mathbf{d}$

in a system described by $\boldsymbol{\theta}$ is simply the probability of observing the noise realization $\boldsymbol{\xi} = \mathbf{d} - \mathbf{y}(\boldsymbol{\theta})$:

$$p(\mathbf{d}|\ln\boldsymbol{\theta}) \propto \exp\left[-\frac{1}{2}(\mathbf{d} - \mathbf{y}(\ln\boldsymbol{\theta}))^T \Sigma_\xi^{-1}(\mathbf{d} - \mathbf{y}(\ln\boldsymbol{\theta}))\right], \tag{4.18}$$

where I have written the exponent in matrix notation and treated $\mathbf{d}$ and $\mathbf{y}$ as column vectors. Superscript T indicates a matrix transpose.

This noise model requires some justification in the context of the current problem. The main source of variability in the fluorescence intensity trajectories is the intrinsic stochasticity of the binding/unbinding dynamics. As illustrated in Figure 4-4, the distribution of fluorescence intensities over all the trajectories at a given time point is not Gaussian, but closer to Poissonian, and this difference is especially important in the parts of the trajectory where the mean number of sensor molecules is of order 1. The reason the Gaussian noise model is justified is that each element of $\mathbf{d}$ is an average of many independent observations. As long as the number of observations is large enough, the central limit theorem guarantees that the distribution of this sample mean will be nearly Gaussian. The variance of this Gaussian is simply the variance of the distribution from which the individual samples are drawn, divided by the number of samples. I approximated the variance of the underlying distribution by the variance of the set of sampled values at each time point.

The second factor in Bayes' formula is the prior probability $p(\ln\mathbf{d})$, which gives the likelihood of $\boldsymbol{\theta}$ being the true parameter set based on prior observations. This term was employed in the fitting process to incorporate important information about the initial numbers of lipids and the relative sensor expression levels that is not contained in the data. I took $p(\ln\boldsymbol{\theta})$ to be a product of independent Gaussian distributions, one for each parameter, with means $\ln\boldsymbol{\theta}_0$ and standard deviations $\boldsymbol{\sigma}_0$. Again, it will be to define a diagonal matrix of variances Σ_0 whose elements are the squares of the elements of $\boldsymbol{\sigma}_0$.

The full expression for $p(\ln\boldsymbol{\theta}|\mathbf{d})$ can now be written as

$$p(\ln\boldsymbol{\theta}|\mathbf{d}) \propto \exp\left[-\frac{1}{2}C(\ln\boldsymbol{\theta},\mathbf{d})\right] \tag{4.19}$$

where the "cost function" is

$$C_{\mathbf{d}}(\ln\boldsymbol{\theta}) = (\mathbf{d}-\mathbf{y}(\ln\boldsymbol{\theta}))^T\Sigma_\xi^{-1}(\mathbf{d}-\mathbf{y}(\ln\boldsymbol{\theta})) + (\ln\boldsymbol{\theta}-\ln\boldsymbol{\theta}_0)^T\Sigma_0^{-1}(\ln\boldsymbol{\theta}-\ln\boldsymbol{\theta}_0) \tag{4.20}$$

Except for the extra term from the prior probabilities, this is the function that is minimized in a standard weighted least-squares optimization. Equation (4.19) says that the vector $\ln\bar{\boldsymbol{\theta}}$ that minimizes $C_{\mathbf{d}}$ is most likely to be the true parameter set. But depending on the width of the distribution $p(\ln\boldsymbol{\theta}|\mathbf{d})$, there may be many other values that are almost as likely.

4.5.2 Estimating Cost Function

Near its maximum at $\ln\bar{\boldsymbol{\theta}}$, $p(\ln\boldsymbol{\theta}|\mathbf{d})$ can be approximated using a second-order Taylor expansion of the cost function $C_{\mathbf{d}}(\ln\boldsymbol{\theta})$ from Equation (4.20):

$$\begin{aligned} C_{\mathbf{d}}(\ln\boldsymbol{\theta}) = C_{\mathbf{d}}(\ln\bar{\boldsymbol{\theta}}) &+ \nabla C_{\mathbf{d}}(\ln\bar{\boldsymbol{\theta}})^T(\ln\boldsymbol{\theta}-\ln\bar{\boldsymbol{\theta}}) \\ &+ \frac{1}{2}(\ln\boldsymbol{\theta}-\ln\bar{\boldsymbol{\theta}})^T H(\ln\bar{\boldsymbol{\theta}})(\ln\boldsymbol{\theta}-\ln\bar{\boldsymbol{\theta}}) + \ldots . \end{aligned} \tag{4.21}$$

The gradient term $\nabla C_{\mathbf{d}}(\ln\bar{\boldsymbol{\theta}})^T(\ln\boldsymbol{\theta}-\ln\bar{\boldsymbol{\theta}})$ vanishes because $\ln\bar{\boldsymbol{\theta}}$ was defined as the value that minimizes $C_{\mathbf{d}}$. H is the matrix of second derivatives of $C_{\mathbf{d}}$, whose elements are

$$H_{ij} \equiv \frac{\partial^2 C_{\mathbf{d}}}{\partial\ln\theta_i\partial\ln\theta_j} = -2(\mathbf{d}-\mathbf{y}(\ln\bar{\boldsymbol{\theta}}))^T\Sigma_\xi^{-1}\frac{\partial^2\mathbf{y}(\ln\bar{\boldsymbol{\theta}})}{\partial\ln\theta_i\partial\ln\theta_j} \tag{4.22}$$

$$-2\frac{\partial\mathbf{y}(\ln\bar{\boldsymbol{\theta}})^T}{\partial\ln\theta_i}\Sigma_\xi^{-1}\frac{\partial\mathbf{y}(\ln\bar{\boldsymbol{\theta}})}{\partial\ln\theta_j} \tag{4.23}$$

$$+2\left(\Sigma_0^{-1}\right)_{ij}. \tag{4.24}$$

The likelihood function (4.19) thus becomes a Gaussian

$$p(\ln \boldsymbol{\theta}|\mathbf{d}) \sim \exp\left[-\frac{1}{2}(\ln \boldsymbol{\theta} - \ln \bar{\boldsymbol{\theta}})^T \Sigma^{-1} (\ln \boldsymbol{\theta} - \ln \bar{\boldsymbol{\theta}})\right] \tag{4.25}$$

where $\Sigma = 2H^{-1}$ is the covariance matrix of the distribution.

The problem of estimating the cost function is thus reduced to the problem of estimating the Hessian of $C_{\mathbf{d}}(\ln \boldsymbol{\theta})$. This task is greatly simplified in the case where $\mathbf{d}$ lies in the space of possible model results, so that $\mathbf{y}(\ln \bar{\boldsymbol{\theta}}) = \mathbf{d}$. In this case, the first term in Equation (4.22) vanishes, and the Hessian can be obtained directly from the Jacobian $\frac{\partial \mathbf{y}(\ln \bar{\boldsymbol{\theta}})}{\partial \ln \theta_j}$. The Jacobian, in turn, can be robustly obtained by first analytically differentiating the dynamical equations of the model with respect to each θ_i and then using a standard ODE solver to obtain the sensitivity at each time point. Typically $\mathbf{d}$ does not lie in the space of possible model results, and most of the discrepancy is due to the fact that the model results are all smooth functions of time, but the noise makes $\mathbf{d}$ jagged. Replacing $\mathbf{d}$ by $\mathbf{y}(\ln \bar{\boldsymbol{\theta}})$ is then a natural way of smoothing the data, and it makes sense to evaluate the cost function after performing this smoothing.

4.5.3 Implementation Details

The data vector $\mathbf{d}$ includes the average sensor fluorescence intensities at forty time points for each of the four sensors, plotted in Figure 4-8, in addition to the intensities from control experiments using only the clathrin-binding part of the sensor or only the membrane-binding part. The measurement uncertainties that go into Σ_ξ were obtained by computing the standard error from the ensemble of fluorescence values at each time point. If d_i^α is the intensity corresponding to data element $\mathbf{d}_i$ as observed in a single endocytic event, then $(\Sigma_\xi)_{ii} = \sum_\alpha \frac{(d_i^\alpha - \langle d_i^\alpha \rangle)^2}{N^2}$.

The prior uncertainties contained in Σ_0 were all set to $10,000$ except for the elements corresponding to the cytosolic sensor concentrations $\{N_{S0}\}$ and some of the initial PIP concentrations $\{N_{L0}\}$, which were set to 1. This left most of the parameters effectively unconstrained: they were free to vary by a factor of $e^{100} \sim 10^{43}$ before the contribution to the cost function became significant. The concentrations $\{N_{S0}\}$ and

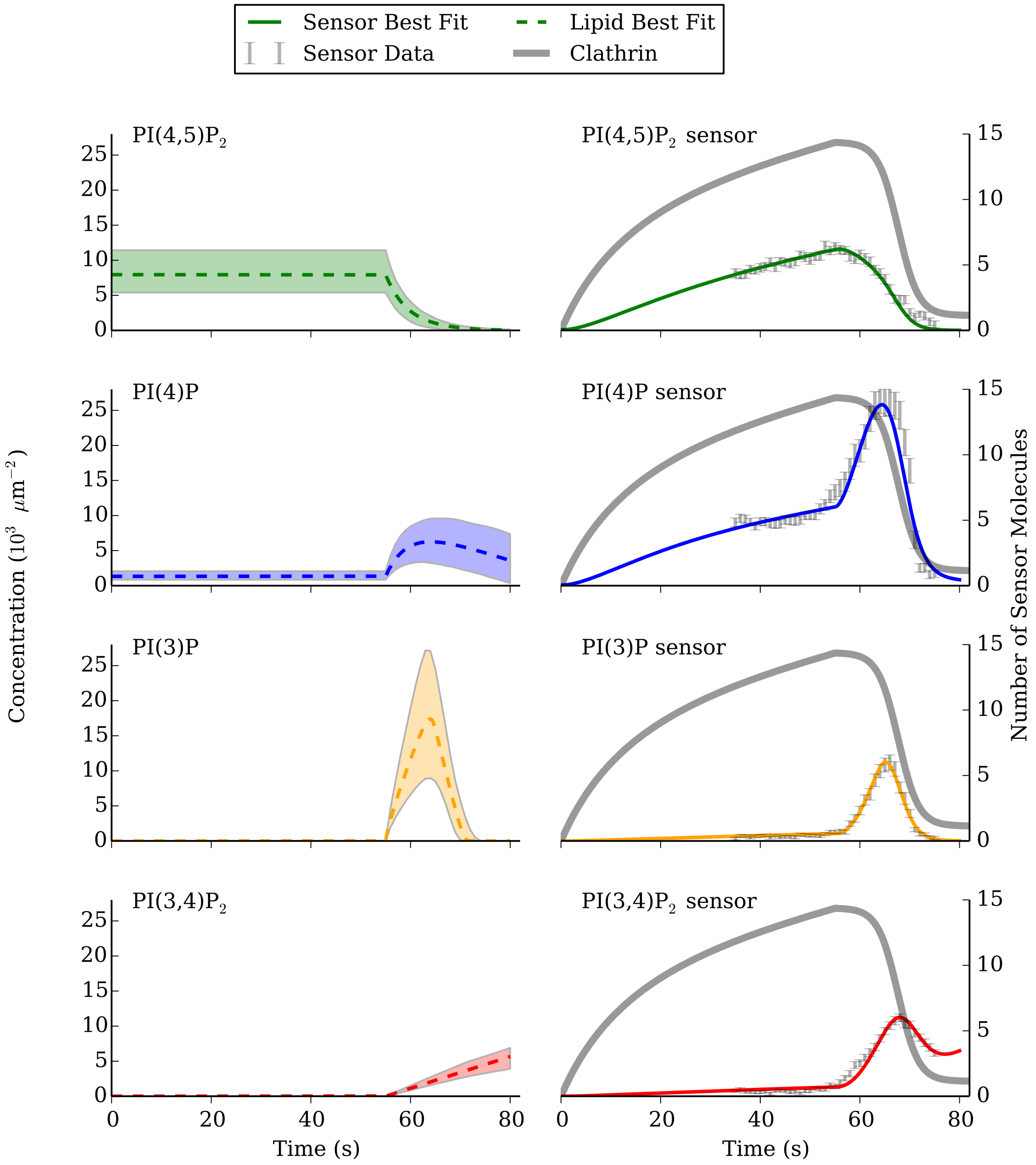

Figure 4-8: Color. Averaged sensor data and inferred PIP dynamics. The right-hand column shows comparisons between the experimental (traces with error bars) and simulated (traces with solid lines) recruitment of the auxilin1-based phosphoinositide sensors to the clathrin-coated structures using the best-fit model parameter. Error bars represent ± standard deviation of the mean σ_ξ. The corresponding phosphoinositide conversion dynamics obtained from the model are plotted as dashed lines in the left-hand plots. The colored band around the phosphoinositide traces represent the middle fifty percent of the likelihood distribution of possible phosphoinositide concentrations given the data, as described in Section 4.5.

$\{N_{L0}\}$ are not well constrained by the data, because any change in these parameters can be easily compensated by simply changing an on-rate or the number of enzymes. This kind of situation is one of the sources of "sloppiness" in large models, where the cost function is nearly flat along some directions in parameter space, and a good effective theory can be found by eliminating those directions [96]. But I am interested in the values of these on-rates and enzyme numbers, and have some prior knowledge that can remove the sloppiness. The relative values of $\{N_{S0}\}$ should all have the same order of magnitude, since the sensors were added using identical procedures, and the remaining degree of freedom simply sets the units of concentration. Likewise, the numbers of PI, PI(4)P and PI(4,5)P_2 molecules initially present in the coated region can be estimated from existing measurements and knowledge of the coat size [61, 2]. By making the cost function increase significantly when these parameters depart from their estimated values by a factor of e, I ensure that the best fit and the uncertainty ranges for all the free parameters are compatible with these biologically motivated constraints.

The final choice I had to make was how to convert from fluorescence intensity to absolute numbers of sensor molecules. It is possible, though challenging, to reliably perform this conversion at single-molecule precision, by performing a delicate calibration routine before each measurement. For the purposes of this experiment, such precise calibration would have been out of place, because the concentration of sensor molecules in the cells is itself variable from cell to cell. Instead, I took the rough conversion factor corresponding to the microscope settings and laser power used in these experiments, and applied it to all the data. The statistics of fluctuations in fluorescence intensity in Figure 4-4, for portions of the trajectory with few sensors present, show a definite peak near the nominal single-molecule fluorescence intensity. This confirms that the factor is close to the true value.

After obtaining the best-fit parameter set $\bar{\boldsymbol{\theta}}$ and the Hessian H_{ij}, I had to convert this information into error bars on the inferred PIP trajectories and on the other parameter values. For a Gaussian distribution one would use the standard deviation for this purpose, but since I have approximated $p(\boldsymbol{\theta}|\mathbf{d})$ by a log-normal distribution, this is no longer an appropriate choice. Instead, I divided the range of possible values for each parameter into four intervals containing equal probability. The boundary between the second and third intervals is the median, while the boundaries between the first and second and between the third and fourth intervals become the two ends of the confidence interval. The region between these boundaries, known as the interquartile range, contains the middle half of the probability in the distribution, and gives a meaningful measure of the width for an arbitrary distribution. In tables 4.1 and 4.2, I tabulate the mode and interquartile range for enzyme and sensor parameters, respectively, computing the interquartile boundaries for these lognormal distributions as $\bar{\theta}_i e^{\pm 0.675 \Sigma_{ii}}$.

This procedure integrates out the correlations in the original multivariate distribution. It gives the range of likely values for each parameter individually, under the assumption that the other parameters remain free to vary. But for evaluating the uncertainties in the PIP trajectories, the parameter correlations are crucial. Each PIP trajectory depends on at least four independent parameters, and some of the trajectories are linked with each other (since PI(4)P comes from PI(4,5)P_2, and PI(3,4)P_2 comes from PI(4)P). To preserve these correlations, I sampled 1,000 parameter sets from the full multivariate distribution $p(\boldsymbol{\theta}|\mathbf{d})$ and integrated the PIP kinetics for each one. This generated an ensemble of 1,000 trajectories, and allowed me to determine the boundaries of the shaded regions in Figure 4-8 by finding the interquartile range at each time point.

4.6 Conclusions

Figure 4-8 and Tables 4.1-4.2 display the results of this inference procedure, along with the uncertainties in the inferred PIP dynamics and parameter values. With this data in hand, we can return to the physical and biological questions that motivated this analysis in Section 4.2.2.

4.6.1 PIP Concentrations

Figure 4-8 contains the inferred PIP concentrations as a function of time, with error bars obtained as described in Section 4.5.

Most of the PIP trajectories look qualitatively different from the corresponding sensor signals, due to the influence of the clathrin dynamics. The PI(3,4)P_2 sensor provides the most extreme example: the PI(3,4)P_2 concentration increases monotonically after fission, but the sensor exhibits a pulse similar to the PI(3)P and PI(4)P sensor trajectories. It is easy to see why this happens: although the phosphoinositide concentration keeps increasing, the clathrin concentration is falling, which eventually reduces the overall affinity of the sensor for the vesicle. But this result now calls the PI(3)P and PI(4)P trajectories into question: could the same fits have been obtained with a monotonic increase in the concentrations of the corresponding lipids?

To answer this, we need to consider the ratio of the sensor concentrations before and after uncoating. In the high-cooperativity regime indicated by the control experiments, the maximum ratio with non-decreasing phosphoinositide concentration is the ratio of initial to final clathrin concentrations. While the PI(3,4)P_2 sensor trajectory has a smaller ratio, and is compatible with a monotonically increasing underlying signal, the other two sensors exhibit much larger ratios that can only be achieved if the corresponding PIP concentrations decrease during uncoating.

Comparing the final sections of the clathrin and sensor trajectories also provides information about the sensor dissociation rates. If these rates were much slower than the rate of clathrin loss, then the decreasing parts of the sensor trajectories would have the same shape as the clathrin signal, and only differ by an overall scale factor describing the fraction of clathrin triskelia that have a sensor protein bound. If anything, the sensor concentrations would decrease more slowly than the overall clathrin, because the clathrin-binding region of the sensor helps to stabilize the part of the coat where it is bound. The fact that the relative sensor signals for PI(4,5)P_2, PI(3)P and PI(4)P decrease more quickly than the clathrin signal implies that their k_{off}^L's must be considerably faster than $k_u = 0.5\ \mathrm{s}^{-1}$.

	N_E	K_M (molec./coat)	$v_{\max}$ (molec./s)
PI→PI(3)P	1.6 (0.9, 2.6)	18	70 (40, 120)
PI(4)P→PI(3,4)P_2	0.06 (0.05, 0.08)	< 0.01	3.5 (2.9, 4.2)
PI(4,5)P_2 →PI(4)P	25 (19, 32)	5000	45 (35, 60)
PI(3)P→PI(3,5)P_2	1.0 (0.6, 1.6)	10	45 (30, 75)

Table 4.1: Best-fit enzyme parameters and confidence intervals, computed as described in Section 4.5.3. Absolute enzyme numbers were obtained by setting $k_{\text{cat}} = 50 \text{ s}^{-1}$, which is near the top of the typical range of turnover rates for this kind of reaction.

4.6.2 Enzyme Numbers

Table 4.1 contains the best-fit parameters for the enzymes responsible for the PIP dynamics. There are two interesting things to note about these numbers. The first is that they are consistent with the diffusion-based trigger mechanism. The maximum turnover rate for a given reaction is:

$$v_{\max} = \frac{k_{\text{cat}} N_E}{1 + \frac{K_M}{N_{\max}}} \tag{4.26}$$

where $N_{\max}$ is the maximum number of substrate molecules encountered during the trajectory. Table 4.1 contains an upper bound on $v_{\max}$ computed by setting $N_{\max}$ equal to the initial number of PI molecules, since this is the maximum concentration reached by any of the lipid species. The largest values of these quantities within the confidence intervals are still very slow compared with the $\sim$ 1000molec./s rate required to keep up with diffusion and maintain an average concentration of one PIP molecule per coat inside the clathrin-coated region. I built this separation of timescales into my kinetic model by keeping enzymatic activity turned off until the moment of vesicle fission. If the rates required to fit the data had been significantly larger, I would have needed to relax this assumption and explicitly model the balance between local creation of new PIP's and diffusion in the membrane.

Second, the mean number per clathrin-coated vesicle for most of the enzymes are very small: all but the PI(4,5)P_2 phosphatase of order unity or smaller. This means that even a very weak association of an enzyme with clathrin is sufficient to generate the trigger mechanism, and it is not surprising that some of the essential enzymes have not yet been identified [38].

Such strong sensitivity to a small change in local enzyme concentration makes it easier to see how this trigger mechanism emerged over the course of evolutionary history. Enzymes that modify PIP's and proteins that detect them were already present in abundance before this mechanism arose [13]. Proteins generically tend to stick to each other, and the local concentration of many different protein species should be at least slightly affected by the presence of a dense, extended protein coat. It would be interesting to estimate the effect of the clathrin coat on concentrations of known PIP kinases and phosphatases based on their non-specific affinity alone, to see whether this would have been sufficient to generate the PIP-based trigger even before evolutionary fine-tuning had time to take place.

The separation of timescales and the low enzyme copy numbers could have been inferred even without the mathematical model. The key assumptions are that the initial numbers of PI and PI(4,5)P_2 are on the order of 100, and that the modification of an enzyme-bound PIP takes about 20 ms. The first assumption implies that a total catalytic rate of 25 molecules/s is sufficient to deplete each preexisting phosphoinositide pool in the coated vesicle before the PI(3)P and PI(4)P sensors reach their peak concentrations. And the second implies that this rate can be achieved with an average of half an enzyme per coat, if the enzyme is running at maximum capacity.

The model helps us to think more carefully through the issues that could complicate these estimates. First of all, the cooperative binding of the sensor to both the clathrin coat and the membrane could easily have been strong enough to make the sensor signal significantly lag behind the underlying PIP dynamics. If the dissociation rate were much slower than the association rate per free PIP, then the peak sensor concentration would not occur until the corresponding PIP's had been almost entirely cleared from the vesicle. So the PI and PI(4,5)P_2 concentrations could be depleted arbitrarily quickly and still generate the same timing for the PI(3)P and PI(4)P sensor peaks, as long as the enzymes that clear the PI(3)P and PI(4)P act sufficiently slowly. Thus the observed peak timing does not automatically guarantee that the catalytic rates are slow enough for the diffusion-based trigger mechanism.

The results of the model-based inference suggest that sensor dissociation is fast enough that the peak sensor concentration should be nearly simultaneous with the peak PIP concentration. The time interval between vesicle fission and the peak of the PI(3)P and P(4)P sensor signal should thus be a good measure of the time required to deplete the PI and PI(4,5)P_2, as assumed in the direct estimate of the total catalytic rate. When I discuss the PIP concentration trajectories below, I will explain which features of the data give rise to this conclusion.

As for the enzyme copy numbers, the direct estimate assumed that the enzymes operate at their maximum rate, with a new substrate molecule binding immediately after the previous reaction has finished. The accuracy of this approximation depends on the size of the Michaelis constant K_M in Equation (4.9). At 18 molecules/vesicle, the inferred Michaelis constant for the PI → PI(3)P reaction is still small enough that the enzyme can nearly reach its maximum reaction velocity, and so the rough estimate of enzyme number is not too far off. The estimated Michaelis constant for the PI(4,5)P_2 → PI(4)P reaction, however, is about 5300 molecules/vesicle, so the enzyme never comes close to saturation. To achieve the required total rate of PI(4)P synthesis, about 25 copies of the enzyme would be required. This result may reflect the limitations of the model, which assumes that there is only one kind of enzyme and that its local concentration is constant.

4.6.3 Sensor Affinities

Finally, the model predicts the affinities of the lipid-binding and clathrin-binding domains of the sensors for their substrates, which I have tabulated in Table 4.2. To obtain these predictions, I had to supply one more piece of information. The absolute cytosolic sensor concentrations N_{S0} are not constrained by the model, since they only appear as products with other free parameters k_{on}^L and k_{on}^C. To estimate the absolute scale for these concentrations, I will assume that binding of sensor molecules to their PIP substrates is diffusion-limited: whenever the lipid-binding part hits the correct phosphoinositide, it immediately binds. The absolute value of k_{on}^L is now set by the diffusion coefficient of the sensors in the cytosol, which should produce an on-rate of

	$N_{S0}k_{\text{on}}$ (s^{-1})	k_{off} (s^{-1})	K_D (nM)	Direct K_D
Clathrin	3 (2, 5) $\times 10^{-4}$	0.07 (0.06, 0.09)	200 (150, 400)	–
PI(4,5)P$_2$	0.03 (0.01, 0.06)	100 (40, 240)	3000 (2000, 4000)	2000 nM [62]
PI(3,4)P$_2$	0.09 (0.07, 0.1)	4 (3, 6)	40 (30, 60)	–
PI(3)P	0.07 (0.05, 0.10)	80 (40, 180)	1000 (500, 2000)	–
PI(4)P	0.11 (0.06, 0.19)	60 (30, 90)	500 (300, 800)	20 nM [88]

Table 4.2: Best-fit kinetic parameters and affinities with confidence bounds, computed as described in Section 4.5.3. Affinities were obtained by assuming that $N_{S0} = 1$ corresponds to a cytosolic concentration of 1 nM. Direct *in vitro* measurements reported in the literature for two of the sensors are listed for comparison.

about $100/\mu\text{M} \cdot s$ for proteins of this size [75]. This implies that $N_{S0} \sim 1$ nM, which is a small but plausible cytosolic concentration.

The affinities of the lipid-binding domains have been independently measured with various biochemical assays. The reported values for the PI(4,5)P$_2$ and PI(4)P sensors are listed in Table 4.2 along with the predictions of the model. The moderate-affinity PI(4,5)P$_2$ sensor shows reasonable agreement between the biochemical measurements and my *in vivo* estimate, but my estimate for the high-affinity PI(4)P sensor differs from the literature by an order of magnitude. This disagreement can be seen directly in the data by comparing the measured $k_{\text{off}} = 0.1$ s^{-1} in the biochemical assays to the rate at which the sensor signal is lost at the end of the averaged trajectory in Figure 4-8. The biochemical measurement implies that the time constant for the decay of the signal should be at least 10 s, while in fact it is less than 5 s.

This sort of discrepancy is common in biochemistry, because so many different factors can affect the apparent affinity of two molecules for each other in the complex environment of the cytosol. One possible cause for the discrepancy in this case is my use of a two-state model to describe the binding of the sensor to the phosphoinositide, in which the sensor is either free or bound. It would be more accurate to include three states: free, bound non-specifically to the membrane, and bound specifically to the sensor, as proposed for example in [57]. In this model, the apparent off-rate increases with decreasing phosphoinositide concentration. Since biochemical assays are typically performed at PIP concentrations much higher than what is found at the plasma membrane, the apparent affinities in the cell should seem weaker. It would be interesting to directly compare the off-rate from membranes with varying PIP concentrations to test this hypothesis.

Chapter 5

Accelerating Kinetics through Nonequilibrium Driving

The goal of Chapter 4 was to identify the mechanism that couples the destabilization of the clathrin coat to the fission of the vesicle from the plasma membrane. In the introduction, I alluded to the challenge of designing a material that can respond to such a trigger so rapidly. How does the cell maintain the mechanical robustness of the clathrin lattice and its ability to exert force on the membrane, while making it so easy to dissolve in response to a tiny signal? In this chapter, I explain why the combination of these two features seems so strange from a thermodynamic perspective, and show how it is more easily attained in a nonequilibrium material.

5.1 The Trade-off

The clathrin lattice is just one example out of many biological structures that combine dynamic responsiveness with persistent force generation and resilience. As noted in a recent review, this paradoxical union can be found at all biological length scales, from microtubule networks to macroscopic tissues [20]. Efforts are underway to replicate this phenomenon in synthetic systems, in order to create materials that can support mechanical loads but also “heal” themselves on a reasonable time scale when damaged [9].

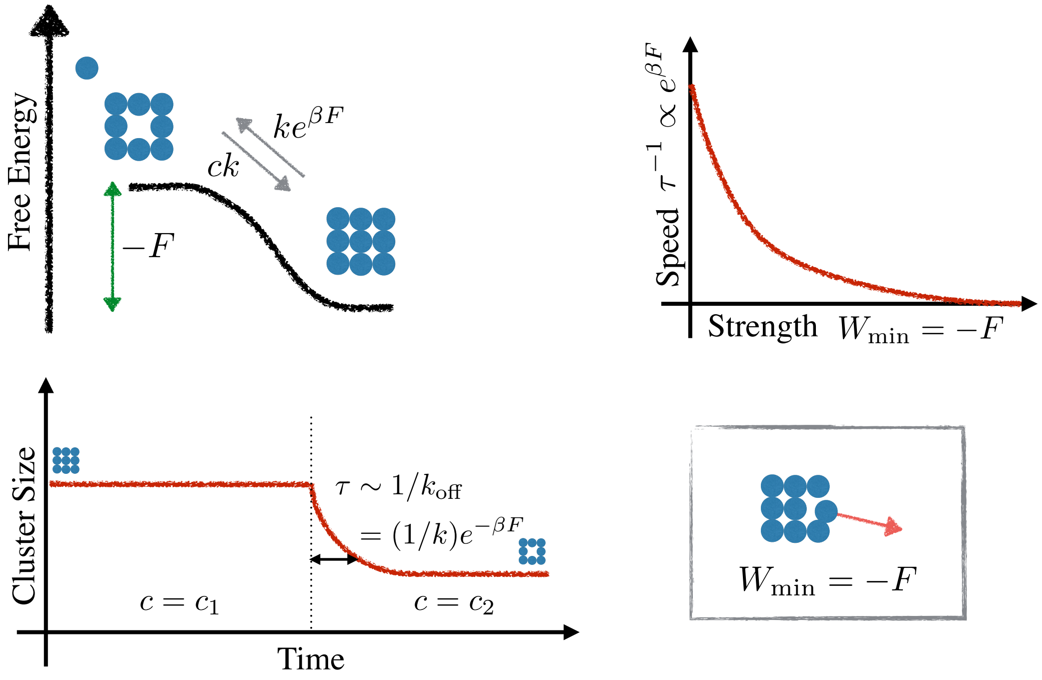

Figure 5-1: Color. Top left: The ratio of rates for adding and removing monomers from a structure is fixed via Equation (2.43) by the free energy F of the bound state and the concentration c of monomers in solution. Bottom left: If some parameter (like the concentration) is suddenly changed by a small amount, the structure will relax to a new equilibrium state over a time scale τ determined by the off-rate. Bottom right: The minimum amount of work W_{min} required to dislodge a single particle from the structure is equal to the free energy change $-F$. Top right: These definitions imply that the speed τ^{-1} is exponentially suppressed with increasing strength W_{min} if the basic rate k is held fixed.

To locate the source of the difficulty, I will consider the generic self-assembly process illustrated in Figure 5-1. A suspension of small particles (like proteins) at concentration c can stick to each other to spontaneously form a structure when c is large enough. Concretely, I assume that the particles can stochastically bind to a set of N pre-defined lattice sites in a Poisson process with rate ck for adding a particle to an unoccupied site. The rate $k_{\text{off}} = ke^{\beta F}$ for removal of a particle is then fixed by the microscopic reversibility relation for Markov jump processes (2.43) given in Chapter 2. The binding energy F (a negative number) depends on how many neighboring sites are occupied. I will require that $F = 0$ in the absence of neighbors, so that $k_{\text{off}} = k$ is the rate at which an unbound particle can diffuse out of the lattice site. This stipulation sets the units of volume in such a way that $c = 1$ is the concentration at which half the lattice sites would be filled in the absence of inter-particle interactions.

F controls the mechanical resilience of the material, since $W_{\text{min}} = -F$ is the minimum amount of work required to remove a particle from the structure. It also affects the rate at which the structure returns to its equilibrium size and configuration after a perturbation. The recovery rate is set by the mean time τ required for a particle to be spontaneously added or removed from a given lattice site. At equilibrium, the total rates for adding and removing are equal, and the rate per site is approximately equal to k_{off} when the sites are mostly occupied: $\tau \approx 1/k_{\text{off}} = e^{-\beta F}/k$.

This does not immediately imply anything about the relationship between τ and the strength $W_{\text{min}} = -F$, because k could also vary as a function of F. But in many self-assembly scenarios, the rate-limiting step for association of a new particle to the structure is diffusion to the target site. The association rate k per unit concentration will then be fixed by the viscosity of the medium and the size of the particle, independently of the local force fields and contact surface geometry that determine F. At a fixed particle size in a given medium, k can thus be regarded as a constant, generating an exponential trade-off between strength and recovery time. When $-F$ becomes much larger than $k_B T$, τ can be orders of magnitude larger than the natural timescale of the system $1/k$.

5.2 Active Dissociation Model

In this section I present a toy model inspired by the clathrin/auxilin/Hsc70 system that softens this trade-off by constantly dissipating chemical energy.

5.2.1 Transition Rates

I start with the basic model proposed in the previous section, and incorporate a chemical driving force into the model by adding a second association/dissociation pathway as illustrated in Figure 5-2. The monomers of actin filaments, microtubules, clathrin coats, and several other macromolecular assemblies can be found in at least two distinct internal states. There is an "active" state capable of binding strongly to the structure, and an "inactive" state that binds much more weakly. For simplicity, I require that both states have the same free energy in solution, and choose this as the zero of free energy. I will denote the free energy of an active particle in the structure as F_A, and that of an inactive particle as $F_I > F_A$. Both these free energies will depend on the number of neighbors in the structure. In the absence of coupling to a chemical energy source, the rate k_+ of transitions from the active to inactive state and k_- for the reverse transition are related by Equation (2.43):

$$\frac{k_-}{k_+} = e^{\beta(F_I - F_A)}. \tag{5.1}$$

The ratios of association and dissociation rates for particles in each of the two states includes a contribution from their concentrations c_A and c_I in the solution surrounding the lattice. The chemical potentials μ_A, μ_I of the reservoirs of active and inactive particles are equal to $k_B T \ln c_A$ and $k_B T \ln c_I$, respectively, because I have assumed that the internal free energies of both conformations are the same. Equation (2.43) thus implies:

$$\frac{k_{\text{on}}^{A}}{k_{\text{off}}^{A}} = e^{-\beta(F_A-\mu_A)} \tag{5.2}$$

$$\frac{k_{\text{on}}^{I}}{k_{\text{off}}^{I}} = e^{-\beta(F_I-\mu_I)}. \tag{5.3}$$

In Figure 5-2 I have chosen the units of concentration such that $c_A = 1$ is the value at which half the lattice sites would be occupied in the absence of interactions or driving ($F_A = k_+ = 0$), and have set the units of time in terms of the fixed diffusive time scale so that $k = k_{\text{on}}^{A}/c_A = 1$.

Coupling these internal state transformations to ATP or GTP hydrolysis drives cycles of association-inactivation-dissociation-activation, as illustrated in Figure 5-2. The coupling mechanism often introduces additional intermediate states: in the clathrin system, for instance, a clathrin triskelion in the structure first binds to auxilin, which in turn binds to Hsc70-ATP, and the ATP is finally hydrolyzed when the Hsc70 binds to the "inactive" conformation of the triskelion (cf. [78]). I will combine all these steps into one, so that their net effect is to modify k_+ and k_-. The ratio of rates is still given by Equation (2.43), but now includes the extra change in free energy due to the hydrolysis of ATP. This change in free energy includes contributions from the internal entropy of the complex, from the release of an inorganic phosphate into the phosphate reservoir, and from the dissipation of the potential energy stored in the Coulomb repulsion between negatively charged phosphates. To keep this free energy distinct from the neighbor-dependent free energies of the structural components, I will refer to it simply as the heat of the hydrolysis reaction Q_{hyd}. The rate ratio thus becomes

$$\frac{k_-}{k_+} = e^{\beta(F_I-F_A-Q_{\text{hyd}})}. \tag{5.4}$$

My decision to combine all the intermediate transitions into k_+ and k_- further implies that the ratio c_A/c_I is equal to the ratio [ATP]/[ADP] of ATP to ADP concentrations. This is easiest to see in systems like actin, which are always either bound to ATP or ADP. With the help of passive catalysts called nucleotide exchange factors, actin monomers in solution can exchange ATP for ADP and vice versa. If the internal free energies of the active/ATP-bound and inactive/ADP-bound states were equal, as I have assumed for this model, then the rates of exchange in each direction would be entirely determined by the ATP and ADP concentrations.

Since the net free energy change around a closed cycle must vanish, Equation (2.43) says that the sum of the log-ratios of forward to reverse rates around the cycle must be equal to the chemical work $\mathcal{W} = \Delta\mu = \mu_{\text{ATP}} - \mu_{\text{ADP}}$ associated with converting a molecule of ATP to ADP:

$$\begin{aligned}
\beta\Delta\mu &= \ln\frac{c_A}{k_{\text{off}}^A} + \ln\frac{k_+}{k_-} + \ln\frac{k_{\text{off}}^I}{c_I} && (5.5)\\
&= \ln\frac{c_A}{c_I} + \beta Q_{\text{hyd}} && (5.6)\\
&= \ln\frac{[\text{ATP}]}{[\text{ADP}]} + \beta Q_{\text{hyd}}. && (5.7)
\end{aligned}$$

5.2.2 Mean-Field Interactions

To complete the model, I need to specify how the particles interact with their neighbors in the lattice, which determines how F_A and F_I depend on the overall microstate of the structure. I will make the mean-field assumption that $F_A = -Jm$ is proportional to the fraction of occupied sites $m \equiv (N_A + N_I)/N$, where N_A is the total number of active particles in the structure N_I the number of inactive, and J sets the energy scale. The binding energy of the inactive particles must satisfy $F_I \leq 0$ so that the rate $k = 1$ of free diffusion from a lattice site remains the upper bound on the off-rate. As long as this is always satisfied, the binding energy difference $\Delta F \equiv F_I - F_A$ can be an arbitrary function of m, which will be eliminated from the final equations.

This model is now equivalent to N globally coupled copies of the three-state rotor illustrated in Figure 5-2, with a nonequilibrium driving force that provides a net drift in one direction around the cycle. If the inactive monomers bind so poorly that they make up a negligible percentage of the total structure occupancy at any given time

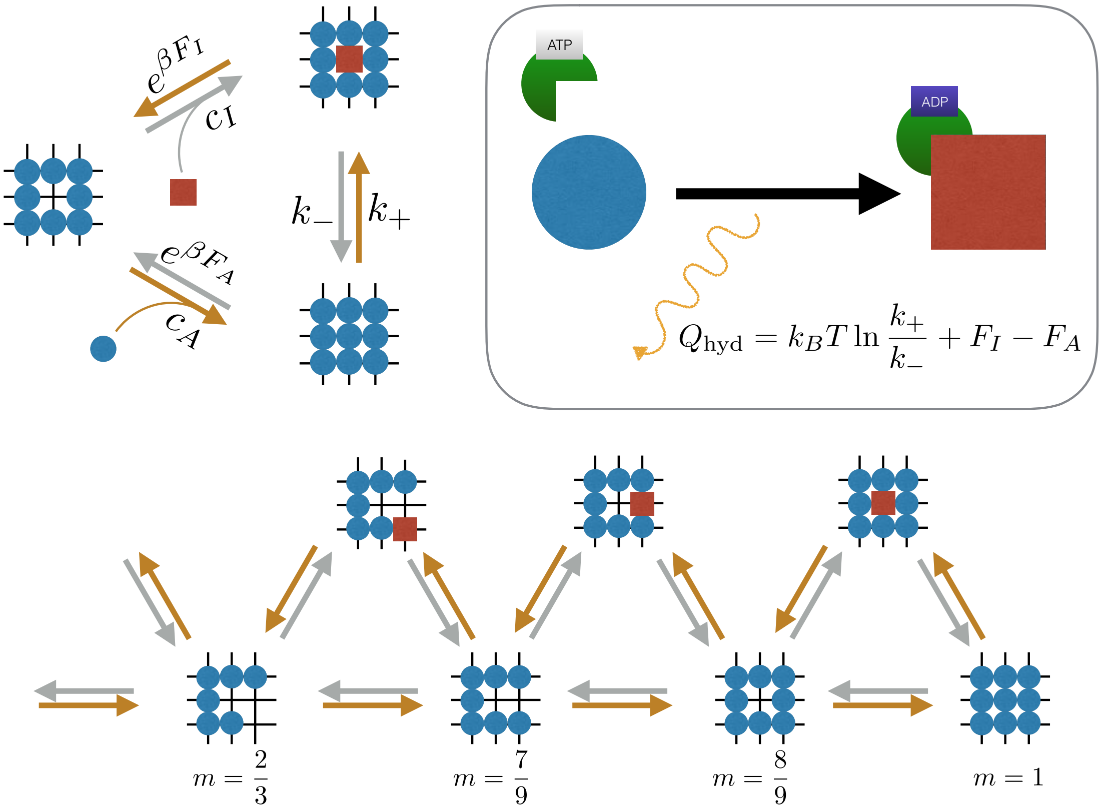

Figure 5-2: Color. Top left: Transition rates for expanded model with two possible internal states for each monomer. If the active and inactive monomer concentrations are kept away from their equilibrium values, the system relaxes to a nonequilibrium steady state with a net drift around the cycle. Top right: Schematic of Hsc70-mediated coupling of assembly dynamics to ATP hydrolysis. Hsc70 can only bind to the "inactive" conformation of the clathrin in the lattice (square), which has a lower overall affinity for its neighbors than the "active" conformation (circle). Binding to clathrin causes Hsc70 to rapidly hydrolyze its bound ATP molecule while clamping on to the clathrin with extremely high affinity. The rate of return to the active state is thereby suppressed by a factor of up to $e^{-\beta Q_{\text{hyd}}}$. Bottom: Simplified dynamics in which number of inactive monomers in the lattice is much smaller than the number of active monomers. The dynamics can now be stated entirely in terms of the fraction m of occupied lattice sites.

($N_I \ll N_A$), then a further simplification is possible: we can eliminate the inactive state, so that each site is either unoccupied or occupied by an active monomer. A sufficient condition on the rates for this approximation to hold is

$$c_I, k_+ \ll e^{\beta F_I} \leq 1 \tag{5.8}$$

where the second inequality follows from the fact that $F_I \leq 0$. This guarantees that the incoming rates are much smaller than at least one of the outgoing rates, so that the steady state will have much more probability concentrated in the states with no particle or a bound active particle in a given site than with an inactive particle. Alternatively, I could have made the ingoing rates smaller than the other outgoing rate k_-. But the physical mechanisms I am trying to understand with this model all contain a highly irreversible inactivation reaction with $k_+ \gg k_-$, which directly violates this assumption.

Now there are two ways for a site to change state: either through direct association/dissociation of an active monomer from the solution, or by transiently passing through the inactive form. The dissociation rate via the inactive conformation is simply the product of the inactivation rate k_+ and the probability $e^{\beta F_I}/(e^{\beta F_I} + k_-)$ that the site ends up with a different occupancy after the next jump (instead of returning to its starting point). Likewise, the association rate on this pathway is the product of the rate c_I for adding an inactive particle to the lattice and the probability $k_-/(e^{\beta F_I} + k_-)$ that the site exits this state in the right direction.

5.2.3 Coarse-Grained Rates

This approximation makes it possible to express the dynamics entirely in terms of the lattice occupancy m, as depicted in Figure 5-2, facilitating the computation of the steady-state distribution $p_{\mathrm{ss}}(m)$. The transition rate $w_{m+1/N,m}$ from m to $m + 1/N$ is the sum of the rates of all possible ways of accomplishing this transition: particles can be added to any of the $N(1-m)$ free sites, and they can be added to each site by either of the two pathways. Similarly, the $m + 1/N$ to m transition with rate

$w_{m,m+1/N}$ involves a sum over the two removal rates for all $N(m+1/N)$ particles currently in the structure:

$$w_{m+1/N,m} = N(1-m)\left(c_A + c_I \frac{k_-}{k_- + e^{-\beta(Jm-\Delta F)}}\right) \tag{5.9}$$

$$w_{m,m+1/N} = N\left(m+\frac{1}{N}\right)\left(e^{-\beta J(m+1/N)} + k_+ \frac{e^{-\beta(J(m+1/N)-\Delta F)}}{k_- + e^{-\beta(J(m+1/N)-\Delta F)}}\right), \tag{5.10}$$

where I have written the kinetics of the inactive state in terms of the free energy difference $\Delta F \equiv F_I - F_A$. Before proceeding, it will be useful to express both rates in terms of the thermodynamic quantities $\Delta\mu$ and Q_{hyd} using Equations (5.4) and (5.5), which imply

$$e^{\beta\Delta\mu} = \frac{c_A}{c_I} e^{\beta Q_{\text{hyd}}} \tag{5.11}$$

$$= \frac{c_A k_+}{c_I k_-} e^{\beta\Delta F}. \tag{5.12}$$

In terms of these new quantities, I have:

$$w_{m+1/N,m} = N(1-m)c_A(1+e^{-\beta\Delta\mu}q(m)) \tag{5.13}$$

$$w_{m,m+1/N} = N\left(m+\frac{1}{N}\right)e^{-\beta J(m+1/N)}(1+q(m+1/N)) \tag{5.14}$$

where

$$q(m) \equiv \frac{k_+}{e^{-\beta Jm} + k_+ e^{-\beta Q_{\text{hyd}}}} \tag{5.15}$$

contains all the dependence on k_+ and Q_{hyd}.

5.3 Steady-State Solution

In this section, I find the steady-state distribution and an approximate expression for the fluctuation dynamics of this model, which will allow me to compute the speed τ^{-1}, the strength $W_{\min}$ and the chemical work rate $\dot{\mathcal{W}}$.

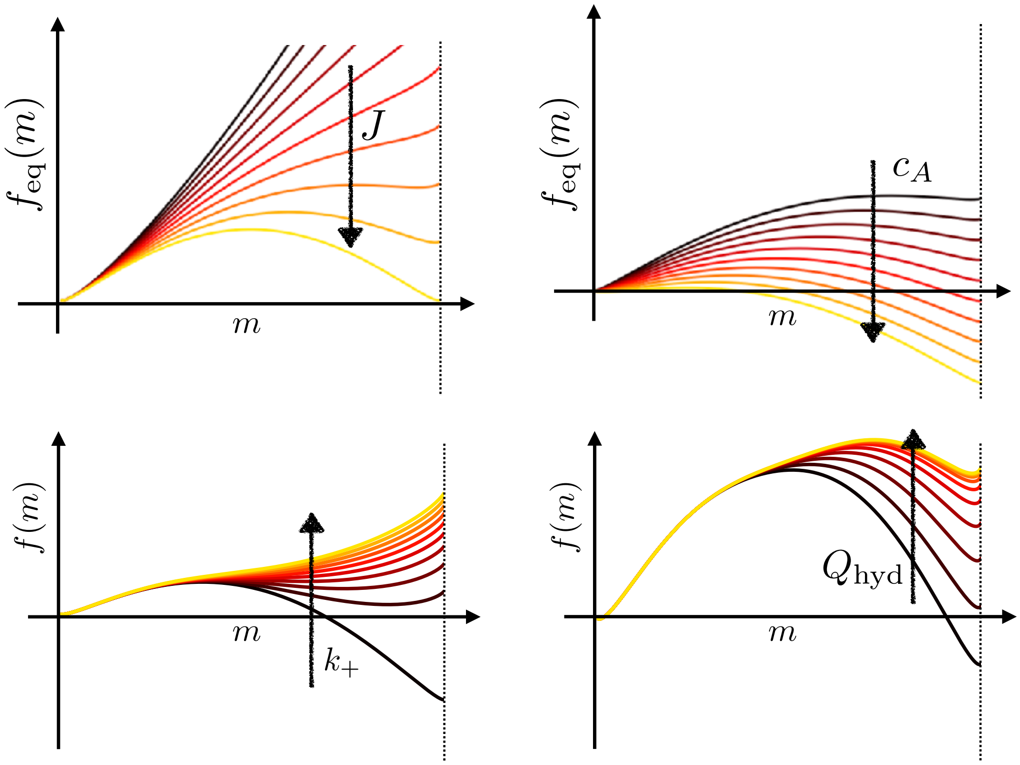

Figure 5-3: Color. Top: Equilibrium free energy density $f_{\rm eq}(m)$ at various values of the coupling J (left) and the concentration c_A (right). Bottom left: Nonequilibrium $f(m)$ at various values of the kinetic parameters k_+ and $Q_{\rm hyd}$ at fixed $\Delta\mu > 0$.

5.3.1 Stationary State

The coarse-grained dynamics are one-dimensional, with hard boundaries at $m = 0$ and $m = 1$, so they cannot support any steady currents. Even in the presence of a nonequilibrium driving force $\Delta\mu \neq 0$, the steady state must obey "detailed balance" in the sense that

$$w_{m+1/N,m}p_{\rm ss}(m) = w_{m,m+1/N}p_{\rm ss}(m + 1/N). \tag{5.16}$$

This means that the model could also describe an undriven system whose free energy landscape is given by the logarithm of $p_{\rm ss}$. But as we will see, the functional dependence of these energies on m does not resemble any readily identifiable physical situation.

From Equation (5.16), it is easy to compute $p_{\text{ss}}(m)$ up to an overall normalization constant. It will be convenient to perform this computation in terms of the derivative of the effective free energy per site $f(m)$, which I will define as

$$\beta f(m) \equiv - \lim_{N\to\infty} \frac{1}{N} \ln p_{\text{ss}}(m). \tag{5.17}$$

The derivative of this quantity can now be related to the ratio of rates:

$$\begin{align} \frac{d(\beta f)}{dm} &\equiv \beta \lim_{N\to\infty} \frac{f(m+1/N) - f(m)}{1/N} \tag{5.18} \\ &= \lim_{N\to\infty} \ln \frac{p_{\text{ss}}(m)}{p_{\text{ss}}(m+1/N)} \tag{5.19} \\ &= \lim_{N\to\infty} \ln \frac{w_{m,m+1/N}}{w_{m+1/N,m}} \tag{5.20} \\ &= \ln m - \ln(1-m) - \beta J m - \ln c_A + \ln \frac{1+q}{1+e^{-\beta\Delta\mu} q}. \tag{5.21} \end{align}$$

Integrating this expression, I find

$$f(m) = f_{\text{eq}}(m) + k_B T \int_0^m dm' \, \ln \frac{1+q}{1+e^{-\beta\Delta\mu} q} + \mathcal{N} \tag{5.22}$$

where $\mathcal{N}$ is a normalization constant, and $f_{\text{eq}}(m)$ is the equilibrium free energy

$$f_{\text{eq}}(m) = k_B T \left[m \ln m + (1-m)\ln(1-m) - m \ln c_A \right] - \frac{J}{2} m^2. \tag{5.23}$$

The remaining integral can be evaluated in terms of the dilogarithm function

$$\text{Li}_2(x) \equiv - \int dx \, \frac{\ln(1-x)}{x} \tag{5.24}$$

to give

$$\begin{align} f(m) = f_{\text{eq}}(m) + \frac{k_B T}{\beta J} \Bigg(&\text{Li}_2 \left[-k_+ (e^{-\beta Q_{\text{hyd}}} + e^{-\beta\Delta\mu}) e^{\beta J m} \right] \\ &- \text{Li}_2 \left[-k_+ (e^{-\beta Q_{\text{hyd}}} + 1) e^{\beta J m} \right] \Bigg) + \mathcal{N}'. \tag{5.25} \end{align}$$

The shapes of $f(m)$ and $f_{\rm eq}(m)$ for various parameter values are illustrated in Figure 5-3.

My goal is to compute τ, $W_{\rm min}$ and $\dot{\mathcal{W}}$ at the stable occupancy m^* that minimizes $f(m)$. This point is where all the probability concentrates in the $N \to \infty$ limit. Equation (5.20) in the above derivation implies that $df/dm = 0$ wherever the reverse transition rate $w_-(m) \equiv w_{m,m+1/N}$ equals the forward transition rate $w_+(m) \equiv w_{m+1/N,m}$:

$$w_+(m^*) = w_-(m^*) \tag{5.26}$$

$$(1 - m^*)c_A(1 + e^{-\beta\Delta\mu}q) = m^* e^{-\beta J m^*}(1 + q) \tag{5.27}$$

The two sides of this equation are plotted in Figure 5-4 for several different parameter values, and the solutions are the m values where the two lines cross. I have plotted the left hand side at three values of c_A with $J = 6k_BT$ fixed, to show how the number of solutions changes as c_A is varied. For small c_A values, the only solution is the global minimum near $m = 0$. As c_A increases, a new local minimum appears near $m = 1$, separated from the small-m solution by a local maximum. The three solutions coexist over a finite range of c_A values, and at some point in this range the value of $f(m)$ at large-m minimum becomes smaller than the value at the small-m minimum. This discontinuous shift in the location of the global minimum of $f(m)$ is an example of a first-order phase transition. As c_A increases further, the minimum near $m = 0$ and the local maximum eventually disappear.

The critical values c_A^* that bound the three-solution region are those for which the functions on the left and right sides of Equation (5.27) are tangent to each other at some point. This means that both the functions and their derivatives must be equal. In equilibrium, when $\Delta\mu = 0$ or $k_+ = 0$, the q-dependent terms cancel out or vanish, and these two conditions take on a simple form that can be solved analytically:

$$(1 - m^*)c_A^* = m^* e^{-\beta J m^*} \tag{5.28}$$

$$-c_A^* = (1 - \beta J m^*)e^{-\beta J m^*} \tag{5.29}$$

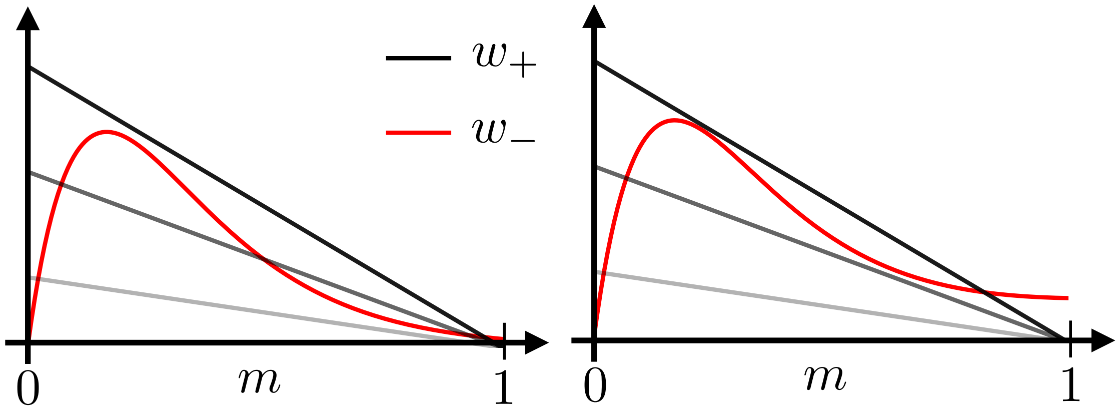

Figure 5-4: Color. Existence of solutions to Equation (5.27). The decreasing gray lines are equal to the association rate w_+ for various values of c_A, and the curved red line is the dissociation rate w_- at $\beta J = 6$. The stationary states of the dynamics defined by the rates (5.13-5.14) occur where these two lines cross. Left: Equilibrium. If c_A is too small, as in the bottom gray line, the only solution occurs near $m = 0$. Right: Nonequilibrium steady state, with the same J and same set of c_A values. The nonequilibrium terms destroy the large-m stationary state that existed at equilibrium for the middle line.

At fixed J, this is a set of two equations in two unknowns. For $J < J_c = 4k_BT$, there are no solutions, and no phase transition: the global minimum increases smoothly from $m = 0$ to $m = 1$ with increasing c_A. When $J > J_c$, eliminating the exponential term generates a solvable quadratic equation in m^*, satisfied by two (m^*, c_A^*) pairs:

$$m^* = \frac{1}{2} \pm \sqrt{\frac{1}{4} - \frac{1}{\beta J}} \tag{5.30}$$

$$c_A^* = \frac{m^*}{1 - m^*} e^{-\beta J m^*}. \tag{5.31}$$

The first-order phase transition must occur somewhere between these two c_A^* values. In the large J limit, the phase transition threshold is easy to find because the two minima are very close to $m = 0$ and $m = 1$. The logarithmic terms in the expression (5.23) for $f_{\text{eq}}(m)$ both vanish at these two points, so setting $f_{\text{eq}}(0) = f_{\text{eq}}(1)$ yields $c_A = e^{-\beta J/2}$.

Out of equilibrium, with nonzero $\Delta\mu$ and k_+, the $q(m)$ terms in Equation (5.27) make the solution much more complicated. But since the nonequilibrium correction term in Equation (5.22) is strictly positive for $\Delta\mu > 0$, the phase transition should happen at a larger c_A than in equilibrium.

Another effect of $\Delta\mu > 0$ comes in to play due to my assumption that $c_I \ll 1$, which ensures that inactive particles occupy a negligibly small fraction of the lattice. By Equations (5.11-5.12), $c_A = e^{\beta(\Delta\mu - Q_{\text{hyd}})}c_I$. The equilibrium ratio of concentrations is given by the exponential of the internal free energy change Q_{hyd} due to the hydrolysis reaction, and deviations from this ratio require a nonzero $\Delta\mu$. So if c_I is fixed to some suitably small value, say $c_I = 0.01$, then increasing $\Delta\mu$ allows a larger active concentration c_A for a given Q_{hyd}. Even though larger $\Delta\mu$ destabilizes the high-occupancy state at fixed c_A, this is more than compensated by the exponential increase of $c_A \propto e^{\beta\Delta\mu}$, and the net result of increasing $\Delta\mu$ is to make high-m states more probable.

5.3.2 Dynamics

To determine the speed τ^{-1} of relaxation from perturbations, I need to go beyond the steady-state distribution and also look at the dynamics of $m(t)$. The equation of motion for m at large but finite N can be approximated by the overdamped Langevin dynamics

$$\dot{m} = A(m) + B(m)\xi_t \tag{5.32}$$

with

$$\begin{aligned} A(m) &= \frac{1}{N}\left(w_{m+1/N,m} - w_{m-1/N,m}\right) && (5.33) \\ &= (1-m)c_A(1 + e^{-\beta\Delta\mu}q) - me^{-\beta J m}(1+q) && (5.34) \end{aligned}$$

and

$$B^2(m) = \frac{1}{N^2}\left(w_{m+1/N,m} + w_{m-1/N,m}\right) \tag{5.35}$$

$$= \frac{1}{N}\left[(1-m)c_A(1+e^{-\beta\Delta\mu}q) + me^{-\beta J m}(1+q)\right]. \tag{5.36}$$

In an equilibrium system with continuous degrees of freedom, the statistical force is proportional to the derivative of free energy and we would have $A \propto -df/dm$. But since the true microscopic dynamics of this model really consists of discrete jumps, $A(m)$ is not directly related to $f(m)$. This is clear from the expressions for these two quantities in terms of the rates, since $f(m)$ is equal to the ratio of forward to reverse w's while A is given by the difference. The usual continuum result is only guaranteed to hold in a neighborhood around the points m^* where $A = df/dm = 0$.

The rate at which a given site spontaneously changes occupancy can now be written in terms of the noise amplitude B as

$$\tau^{-1} = \frac{NB^2(m^*)}{2} \tag{5.37}$$

$$= (1-m^*)c_A(1+e^{-\beta\Delta\mu}q) \tag{5.38}$$

$$= m^* e^{-\beta J m^*}(1+q). \tag{5.39}$$

where I have used Equation (5.27) to simplify the expression in two different ways. In the steady state, the rates for adding and removing particles must balance each other, and so the response speed can be computed from either one.

Since m^* is close to 1 in the high-strength states of interest, the acceleration of the dynamics comes mainly from an increase in q, which was defined in Equation (5.15) as

$$q(m) \equiv \frac{k_+}{e^{-\beta J m} + k_+ e^{-\beta Q_{\text{hyd}}}}. \tag{5.40}$$

Recall that $k_+ \ll 1$ in order to ensure that inactive monomers take up a negligible fraction of occupied lattice sites. At fixed k_+, q increases with Q_{hyd}, up to an asymptotic value of

$$\lim_{Q_{\text{hyd}}\to\infty} q(m) = k_+ e^{\beta J m} \tag{5.41}$$

which gives

$$\lim_{Q_{\text{hyd}}\to\infty} \tau^{-1} = m^* k_+ + m^* e^{-\beta J m^*}. \tag{5.42}$$

In this limit, every inactivation reaction leads to rapid ejection from the lattice, so when J is large the effective off-rate is simply k_+. Even though $k_+ \ll 1$, an acceptable value of $k_+ = 0.01$ still accelerates the dynamics by a factor of 200 at $J = 10k_BT$.

As mentioned in Section 5.3.1 above, Q_{hyd} also affects the supply of active monomers c_A at fixed $c_I \ll 1$. Taking $Q_{\text{hyd}} \to \infty$ implies that $\Delta\mu \to \infty$ in order to keep $c_A = e^{\beta(\Delta\mu - Q_{\text{hyd}})} c_I$ above the phase transition threshold . But τ^{-1} remains close to the $Q_{\text{hyd}} \to \infty$ limit as long as $k_+ e^{-\beta Q_{\text{hyd}}} \ll e^{-\beta J}$.

5.3.3 Energetics

To determine the work rate $\dot{\mathcal{W}}$ according to Equation (2.90) from Chapter 2, I need to separate out the rates of change of m due to the two particle reservoirs at chemical potentials $\mu_A = k_BT \ln c_A$ and $\mu_I = k_BT \ln c_I$:

$$\dot{m}^A = A^A(m) + B^A(m)\xi_t^A \tag{5.43}$$

$$\dot{m}^I = A^I(m) + B^I(m)\xi_t^I \tag{5.44}$$

with

$$A^A = (1-m)c_A - me^{-\beta J m} \tag{5.45}$$

$$A^I = (1-m)c_A e^{-\beta\Delta\mu} q - me^{-\beta J m} q \tag{5.46}$$

$$(B^A)^2 = \frac{1}{N}[(1-m)c_A + me^{-\beta J m}] \tag{5.47}$$

$$(B^I)^2 = \frac{1}{N}\left[(1-m)c_A e^{-\beta\Delta\mu} q + me^{-\beta J m} q\right]. \tag{5.48}$$

In terms of these quantities, the chemical work rate is

$$\dot{\mathcal{W}} = N(\mu_A \dot{m}^A + \mu_I \dot{m}^I). \tag{5.49}$$

As discussed in Section 2.6.2, this equation with the Langevin expressions (5.43-5.44) correctly describes the first two cumulants of the distribution of work rates, but is insufficient for the computation of $\Phi_{\rm ex}$ except near the steady state when $\Delta\mu \to 0$.

In the limit of small $\Delta\mu$, the nonequilibrium correction to the steady-state distribution can be computed using Equations (2.73) and (2.97) (with $\epsilon = 0$) from Chapter 2. This gives

$$f(m) = f_{\rm eq}(m) - \frac{1}{N}\mathcal{W}_{\rm ex}(m) + \mathcal{N} \tag{5.50}$$

where $\mathcal{N}$ is a constant independent of m and the excess work is

$$\mathcal{W}_{\rm ex}(m) = N(\bar{\mu} - \mu_A)(m - m^*). \tag{5.51}$$

I have subtracted off μ_A from $\bar{\mu}$ because the chemical work of the reservoir of active particles is already included in $f_{\rm eq}(m)$ as defined in Equation (5.23). The average chemical potential $\bar{\mu}$ is defined in Equation (2.98), giving:

$$\begin{aligned}\bar{\mu} &= \frac{\mu_A \frac{dA^A}{dm} + (\mu_A - \Delta\mu)\frac{dA^I}{dm}}{\frac{dA^A}{dm} + \frac{dA^I}{dm}} && (5.52)\\ &= \mu_A - \Delta\mu \frac{q}{1+q} + O(\Delta\mu^2). && (5.53)\end{aligned}$$

The resulting prediction

$$f(m) = f_{\rm eq}(m) + \Delta\mu \frac{q}{1+q}(m - m^*) + \mathcal{N}' \tag{5.54}$$

agrees with the first-order Taylor expansion of Equation (5.22) about $\Delta\mu = 0$ near $m = m^*$.

For $\Delta\mu$ of order $k_B T$ and larger, the accuracy of this prediction breaks down. But

Equation (5.49) still correctly describes the mean and typical fluctuations of the work rate. The mean work rate at $m = m^*$ can be written in a relatively compact form using the fact that $A_A = -A_I$ at m^*:

$$\langle \dot{\mathcal{W}} \rangle_{m^*} = N\Delta\mu \left[\frac{1+q}{1+e^{-\beta\Delta\mu}q} - 1 \right] m^* e^{-\beta J m^*}. \tag{5.55}$$

Combining Equations (5.39) and (5.55) generates an expression for the chemical work rate in terms of the speed τ^{-1}:

$$\langle \dot{\mathcal{W}} \rangle_{m^*} = N\Delta\mu \frac{(e^{\beta\Delta\mu} - 1)(\tau^{-1} - m^* e^{-\beta J m^*})}{\tau^{-1} + (e^{\beta\Delta\mu} - 1) m^* e^{-\beta J m^*}} m^* e^{-\beta J m^*}. \tag{5.56}$$

At a given J and τ^{-1}, the work rate is determined by $\Delta\mu$, increasing from zero as $\Delta\mu$ increases. The minimum work rate at a given strength and speed is thus determined by the minimum $\Delta\mu$ required to maintain the existence of a high-occupancy steady state m^*.

5.4 Cost of Acceleration

5.4.1 Results

With the expressions for the steady state m^*, the speed τ^{-1} and the work rate $\langle \dot{\mathcal{W}} \rangle$ in hand, I can now proceed to investigate how the constant supply of chemical work affects the speed-strength trade-off.

To do this, I fix $c_I = 0.01$ and $k_+ = 0.01$ at values compatible with the assumption of negligible inactive population on the lattice, as discussed in Section 5.2.2. The concentration c_A of active monomers is now controlled by $\Delta\mu$ and Q_{hyd} via $c_A = e^{\beta(\Delta\mu - Q_{\text{hyd}})} c_I$ as in Section 5.3.1. The minimal work rate compatible with a given c_I and Q_{hyd} is found when $\Delta\mu$ is just barely large enough to keep c_A on the assembled side of the phase transition presented in Section 5.3.1. In the first panel of Figure 5-5 I plot W_{min}, τ^{-1} and $\langle \dot{\mathcal{W}} \rangle$ at this minimal $\Delta\mu$ value as I sweep over J and Q_{hyd}. In the second panel, I show the full $f(m)$ for a subset of the parameter values included

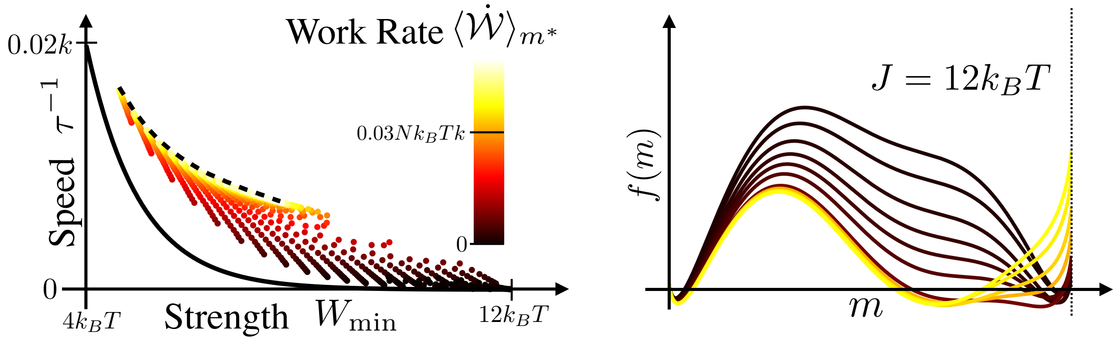

Figure 5-5: Color. Left: Minimum work rate as a function of speed and strength at $k_+ = 0.01$, $c_I = 0.01$. The solid black line is the equilibrium trade-off with $k_+ = 0$, and the dotted line is the maximum speed at the given k_+ value, as given by Equation (5.42). Right: Plots of $f(m)$ at the parameter values corresponding to the rightmost set of colored dots in the first panel, which all have $J = 12k_BT$. The chemical potential difference $\Delta\mu$ is tuned to the phase transition threshold for each curve, where the probabilities of the low-m and high-m local minima are equal.

in the sweep, highlighting the fact that $\Delta\mu$ is tuned to give the low-m and high-m local minima the same probability.

The solid black line in the first panel is the equilibrium trade-off from Section 5.1. The dotted line is the maximum possible speed $\tau^{-1} = k_+m^* + m^*e^{-Jm^*}$ as given by Equation (5.42). This speed is achieved in limit of infinite Q_{hyd}, and therefore infinite $\langle\dot{\mathcal{W}}\rangle$. The colored dots between the two lines are the results obtained from the J, Q_{hyd} sweep as just described. As Q_{hyd} increases at fixed J, τ^{-1} increases while the strength $W_{\min} = Jm^*$ decreases. This is the reason why the dots are arranged in slanted lines.

A work rate of about $\langle\dot{\mathcal{W}}\rangle \approx 0.03k_BTNk$ is sufficient to nearly achieve the maximum τ^{-1} at a given strength $W_{\min}$. This work rate is equal to $5k_BT$ per lattice site per dissociation event at the maximum speed in the large J limit $\tau^{-1} \approx 0.006k$. For comparison, the chemical work corresponding to conversion of a single ATP molecule to ADP under typical cellular conditions is about 20 k_BT [75]. This maximum speed is much slower than the fixed diffusion rate k, but it is much faster than the equilibrium speed. At a strength of $W_{\min} = 8k_BT$, supplying $5k_BT$ of chemical work per dissociation event accelerates τ^{-1} by a factor of 20.

5.4.2 Discussion

This model is a simple example of a generic mechanism for combining strength with rapid relaxation dynamics in a self-assembling structure. The mechanism relies on component parts that can exist in at least two distinct internal states. In the “active” state, they stick to each other, and spontaneously assemble. In the “inactive” state, they do not stick, and spontaneously disassemble. A nonequilibrium steady state is set up when the inactive state has a much lower internal free energy than the active state, and the concentration of inactive monomers is kept low by some active process. Then a steady cycle is set up in which sticky monomers temporarily assemble, and then disassemble as they switch to the weakly binding state. My results indicate that significant acceleration can be achieved at a given target strength even if the internal state changes slowly compared to the diffusive timescale for freely entering or leaving a binding site.

The concentrations of active and inactive monomers play a crucial role in this phenomenon. It is not enough simply to have component parts that spontaneously lose their stickiness some time after entering the structure. If the inactive particles that dissociate into the surrounding bath are not removed, and replaced with active particles, the active particles will eventually run out, leaving a solution of inactive particles that do not stick to each other. This is what I mean by a “nonequilibrium” self-assembly process: the structure depends on the constant activity of some (chemical) work source, and dissolves when that activity ceases.

Under my current set of assumptions and constraints, increasing k_+ allows higher speeds to be achieved at lower work rates. But if k_+ becomes too large, a significant fraction of the particles on the lattice will be in the inactive state, where they bind their neighbors much more weakly. This decreases the strength of the structure, since these weak binders take up space on the lattice without contributing much to the total binding energy. The effect is more dramatic at lower dimensions, because a cluster of inactive particles can span the whole structure and effectively break it into several disconnected pieces. Since the mechanical integrity of the structure requires that most of its components be active, the small k_+ regime is where the speed-strength trade-off becomes most relevant. It would be interesting to look at a more general solution not subject to the inequality (5.8) to determine more precisely how the strength breaks down at higher k_+ values. But many important features of the breakdown of mechanical integrity in this regime depend on structural features like the spanning clusters just mentioned, which are not present in the mean-field model and are not fully captured by my measure of strength $W_{\min}$. Progress in this direction requires numerical simulations or experiments on specific systems.

This model displays several striking features that are worth exploring in more depth. As the coupling strength J increases, the lattice occupancy m^* at the minimum dissipation rate for a given speed decreases. Asymptotically, it appears that the occupancy is inversely proportional to the coupling strength, so that their product Jm^* (the "strength" of the structure) approaches a constant plateau in the large J limit. The degree to which a structure's turnover dynamics can be sped up while preserving its strength depends on where this plateau occurs. But it is not yet clear why this happens, and which parameters are most responsible. This system also displays rich phase behavior at high J, apparently containing an additional first-order phase transition between two mostly-occupied states in addition to the original assembly transition. These features need to be investigated in more detailed models, and in lower dimensions, to determine whether they are merely artifacts of the mean-field approximation.

The results of Chapter 2 may prove useful for obtaining intuition about these more complex models. My current model is simple enough that the exact steady-state distribution can be obtained analytically, and so I had no use for the calculation based on excess work contained in Equation (5.54). But in a model that does not admit of exact results, this approach may provide a basis for new approximation schemes.

Chapter 6

Conclusions

Over the course of this dissertation, I have addressed two fundamental questions about driven steady states: How far can ideas from equilibrium thermodynamics be extended into this nonequilibrium regime? And what new material properties become available when the time-reversal symmetry of thermal equilibrium is broken?

In Chapter 2, I determined the range of validity of a variational procedure (2.73) that generalizes the idea of free energy minimization to predict the properties of near-equilibrium steady states. I showed that this prediction remains accurate beyond the traditional linear response regime of vanishingly small nonequilibrium driving force. The driving force can be made arbitrarily strong, as long as the fluctuation dynamics of the observables of interest remain well described by a linear overdamped Langevin equation. The quantity minimized in this generalization is obtained by subtracting the mean external work done on the way to a fluctuation from the equilibrium free energy. This mean work can be expressed up to a normalization constant in terms of the instantaneous work rate and a relaxation time. My result thus provides a route for generating physical intuition about the behavior of complex driven systems in terms of these familiar quantities.

When the fluctuation dynamics of the observables are nonlinear, the accuracy of my variational prediction depends on the size of the nonlinearity. I showed that the simplified variational principle remains accurate as long as the nonlinearity changes the work on the way to a typical fluctuation by significantly less than $k_B T$.

I illustrated these results with the example of a driven Brownian colloid in Chapter 3. This system naturally exhibits a strong dependence of relaxation time on the strength of the driving force, which allows Equation (2.73) to accurately predict the nonlinear response of the shear stress to an applied shear flow. I numerically simulated the dynamics of such a colloid, and extracted both the mean and the variance of the work done on the way to typical fluctuations of the shear stress. I used these measurements to confirm that the generalization of free energy minimization provides a good prediction of the actual mean shear stress in the limit of large system size. At very high values of the shear rate, where the prediction begins to depart slightly from the real value, the correction term involving the work variance fully accounted for the error.

The thermodynamic perspective of Equation (2.73) allowed me to capture the full nonlinear response of the shear stress with a phenomenological model of the dependence of the relaxation time on shear rate, which is readily generalizable to systems with more complex microscopic details. In the absence of shear, the relaxation time depends on how fast the particles can randomize their positions through Brownian motion, which is determined by the diffusion coefficient and the number density of particles. The Green-Kubo relation of linear response theory relates the viscosity of the suspension to this equilibrium relaxation time. But the imposed shear flow can accelerate the randomization by convectively stirring the particles, resulting in a relaxation time that asymptotically decreases as the inverse of the shear rate. My variational principle generalizes the Green-Kubo relation to this strongly driven regime, showing how the decreased relaxation time generates a decrease in viscosity.

I then turned to the second question, inspired by the surprising properties of the driven protein structures involved in clathrin-mediated endocytosis. This is a complex process involving many physically interesting features, including self-assembling functional structures, active disaggregation mechanisms and membrane-based information processing. Chapter 4 describes the essential features of this process, and presents the results of an experimental investigation of the regulation of the disaggregation mechanism.

We wanted to measure the changes in concentration of information-bearing membrane components throughout the process, using fluorescent "sensor" proteins that specifically bind to these components and to the clathrin coat. But as often happens in biology, many of the parameters values (such as binding affinities) required to extract the quantities of interest from the sensor data could not be independently measured. Even though some affinity measurements have been made on the isolated component parts, the effective values change considerably when the protein is inserted into the crowded and heterogeneous environment of the cell.

Most of my effort in the collaboration was dedicated to systematically accounting for this uncertainty. I first reduced the uncertainty as much as possible by integrating the available prior knowledge into a mathematical model of the concentration changes and sensor binding. After finding the parameter values that best fit the data, I quantified the remaining uncertainty by calculating the sensitivity of the sensor signal to changes in parameters. The uncertainties in the binding affinities and enzymatic conversion rates were small enough to provide some guidance for future experiments investigating *in vivo* binding kinetics or searching for the enzymes that trigger uncoating.

In Chapter 5, I explored the novel physical properties of a class of nonequilibrium structures that includes the clathrin coats of my experimental collaboration, as well as networks of actin or microtubules and synthetic self-healing materials that mimic their behavior. Clathrin coats display a remarkable combination of strength and speed, sustaining sufficient mechanical force to bend the cell membrane, but dissolving in seconds when the vesicle is complete. I first investigated the physical reason for the intuitive surprise we feel when observing such phenomena, by identifying the physical quantities associated with the relevant sense of "speed" and "strength." After explaining how these properties become incompatible in thermal equilibrium when the basic time scale of the dynamics is constrained, I developed a simple stochastic model to show how this trade-off can be softened by an active disassembly pathway, and computed the dissipation rate required to attain a given degree of dynamic acceleration at a given strength.

Over the course of these investigations, I have not only developed new tools for understanding complex driven systems, but have also gained a deeper grasp of what this "understanding" looks like. The central challenge is to identify which of the myriad parameters characterizing the exact state of a given system are actually relevant for controlling the properties of interest. I faced this challenge most directly in my experimental collaboration, where I combined qualitative arguments and recently developed computational methods to estimate the sensitivity of our data set to arbitrary perturbations in the full 16-dimensional parameter space. But this was also the underlying goal of my extended linear response theory, which expressed the properties of a nonequilibrium steady state in terms of the free energy, the relaxation time and the work rate. In this regime, microscopic properties of the individual components can only affect the observable features of the whole macroscopic system via these three quantities. My driven self-assembly model is also aimed in this direction, since it provides a tractable mathematical platform for teasing out the factors that control strength and turnover speed in different regimes.

I have developed these last two lines of research to the point where they can provide some initial guidance for sorting through the bewildering array of parameters in experimental or numerical studies of technologically interesting systems – including biological entities and synthetic driven materials. Feedback from these applications is essential for developing the ideas in their most productive direction. By gradually refining our concepts in response to successes or failures in predictive control, we may finally find a comfortable home for living matter within the umbrella of physics.

Bibliography

[1] F. Aguet, C. Antonescu, M. Mettlen, S. Schmid, and G. Danuser. Advances in analysis of low signal-to-noise images link dynamin and AP2 to the functions of an endocytic checkpoint. *Dev. Cell*, 26(3):279–291, 2013.

[2] T. Balla. Phosphoinositides: tiny lipids with giant impact on cell regulation. *Physiol. Rev.*, 93(3):1019–137, 2013.

[3] U. Basu, C. Maes, and K. Netočný. Statistical forces from close-to-equilibrium media. *New J. Phys.*, 17(11), 2015.

[4] G. K. Batchelor. The stress system in a suspension of force-free particles. *J. Fluid Mech.*, 41:545, 1970.

[5] G. K. Batchelor. The effect of Brownian motion on the bulk stress in a suspension of spherical particles. *J. Fluid Mech.*, 83:97–117, 1977.

[6] R. Belousov and E. G. D. Cohen. Second-order fluctuation theory and time autocorrelation function for currents. *Phys. Rev. E*, 94:062124, 2016.

[7] L. Bertini, A. De Sole, D. Gabrielli, G. Jona-Lasinio, and C. Landim. Macroscopic fluctuation theory. *Rev. Mod. Phys.*, 87(2):593–636, 2015.

[8] T. Böcking, F. Aguet, S. C. Harrison, and T. Kirchhausen. Single-molecule analysis of a molecular disassemblase reveals the mechanism of Hsc70-driven clathrin uncoating. *Nat. Struct. Mol. Biol.*, 18(3):295–301, 2011.

[9] J. Boekhoven, W. E. Hendriksen, G. J. M. Koper, R. Eelkema, and J. H. van Esch. Transient assembly of active materials fueled by a chemical reaction. *Science*, 349:1075–1079, 2015.

[10] L. Boltzmann. *Theoretical physics and philosophical problems: selected writings*. Reidel Pub. Co., Dordrecht, 1974.

[11] J. M. Brader. Nonlinear rheology of colloidal dispersions. *J. Phys. Condens. Matter*, 22(36):363101, 2010.

[12] J. F. Brady and G. Bossis. Stokesian {D}ynamics. *Annu. Rev. Fluid Mech.*, 20(1):111–157, jan 1988.

[13] J. R. Brown and K. R. Auger. Phylogenomics of phosphoinositide lipid kinases: perspectives on the evolution of second messenger signaling and drug discovery. *BMC Evol. Biol.*, 11:4, 2011.

[14] J. G. Carlton and P. J. Cullen. Coincidence detection in phosphoinositide signaling. *Trends Cell Biol.*, 15(10):540–547, 2005.

[15] V. Y. Chernyak, M. Chertkov, and C. Jarzynski. Path-integral analysis of fluctuation theorems for general Langevin processes. *J. Stat. Mech. Theory Exp.*, 2006:P08001–P08001, 2006.

[16] M. Colangeli, C. Maes, and B. Wynants. A meaningful expansion around detailed balance. *J. Phys. A Math. Theor.*, 44(9):095001, jan 2011.

[17] G. E. Crooks. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Phys. Rev. E*, 60(3):2721–2726, 1999.

[18] G. E. Crooks. *Excursions in Statistical Dynamics.* PhD thesis, University of California at Berkeley, 1999.

[19] G. E. Crooks. Path-ensemble averages in systems driven far from equilibrium. *Phys. Rev. E*, 61:2361–2366, 2000.

[20] S. Dumont and M. Prakash. Emergent mechanics of biological structures. *Mol. Biol. Cell*, 25:3461, 2014.

[21] A. Einstein. Investigations on the theory of the Brownian movement. *Ann. der Phys.*, 17:549, 1905.

[22] A. Einstein. The theory of the opalescence of homogeneous fluids and liquid mixtures near the critical state. *Ann. Phys.*, 33:1275–1298, 1910.

[23] J. L. England. Statistical physics of self-replication. *J. Chem. Phys.*, 139:121923, 2013.

[24] D. L. Ermark and J. A. McCammon. Brownian dynamics with hydrodynamic interactions. *J. Chem. Phys.*, 69:1352, 1978.

[25] D. Evans. *Statistical mechanics of nonequilibrium liquids.* Cambridge University Press, Cambridge, 2008.

[26] E. Frey and K. Kroy. Brownian motion: A paradigm of soft matter and biological physics. *Ann. Phys.*, 14:20–50, 2005.

[27] M. Fuchs and M. E. Cates. Integration through transients for Brownian particles under steady shear. *J. Phys. Condens. Matter*, 17:S1681, 2005.

[28] T. Fujiwara, K. Ritchie, H. Murakoshi, K. Jacobson, and A. Kusumi. Phospholipids undergo hop diffusion in compartmentalized cell membrane. *J. Cell Biol.*, 157(6):1071–1081, 2002.

[29] C. W. Gardiner. *Stochastic methods.* Springer-Verlag, Berlin, 4th edition, 2009.

[30] T. R. Gingrich, J. M. Horowitz, N. Perunov, and J. L. England. Dissipation bounds all steady-state current fluctuations. *Phys. Rev. Lett.*, 116:120601, 2016.

[31] D. J. Griffiths and C. Inglefield. *Introduction to Electrodynamics.* Prentice Hall, Upper Saddle River, New Jersey, 2005.

[32] R. Guan, D. Han, S. C. Harrison, and T. Kirchhausen. Structure of the PTEN-like region of auxilin, a detector of clathrin-coated vesicle budding. *Structure*, 18(9):1191–1198, 2010.

[33] R. N. Gutenkunst, F. P. Casey, J. J. Waterfall, C. R. Myers, and J. P. Sethna. Extracting falsifiable predictions from sloppy models. *Ann. N. Y. Acad. Sci.*, 1115:203–211, 2007.

[34] R. N. Gutenkunst, J. J. Waterfall, F. P. Casey, K. S. Brown, C. R. Myers, and J. P. Sethna. Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput. Biol.*, 3(10):e189, 2007.

[35] T. Harada. Macroscopic expression connecting the rate of energy dissipation with the violation of the fluctuation response relation. *Phys. Rev. E*, 79(3):030106, 2009.

[36] T. Hatano and S.-i. Sasa. Steady-state thermodynamics of Langevin systems. *Phys. Rev. Lett.*, 86(16):1–4, 2001.

[37] D. T. Haynie and C. P. Ponting. The N-terminal domains of tensin and auxilin are phosphatase homologues. *Protein Sci.*, 5:2643–2646, 1996.

[38] K. He, R. Marsland, S. Upadhyayula, E. Song, R. Gaudin, M. Ma, and T. Kirchhausen. Dynamics of phosphoinositide conversion in clathrin-mediated endocytic traffic. Manuscript submitted for publication. 2017.

[39] J. M. Horowitz. Diffusion approximations to the chemical master equation only have a consistent stochastic thermodynamics at chemical equilibrium. *J. Chem. Phys.*, 143(4):044111, 2015.

[40] J. M. Horowitz and J. M. R. Parrondo. Entropy production along nonequilibrium quantum jump trajectories. *New J. Phys.*, 15(8):085028, 2013.

[41] L. F. Horowitz, W. Hirdes, B.-C. Suh, D. W. Hilgemann, K. Mackie, and B. Hille. Phospholipase C in living cells: activation, inhibition, Ca2+ requirement, and regulation of M current. *J. Gen. Physiol.*, 126(3):243–262, 2005.

[42] Y. Ishikawa, M. Maeda, M. Pasham, F. Aguet, S. K. Tacheva-Grigorova, T. Masuda, H. Yi, S.-U. Lee, J. Xu, J. Teruya-Feldstein, M. Ericsson, A. Mullally, J. Heuser, T. Kirchhausen, and T. Maeda. Role of the clathrin adaptor PICALM in normal hematopoiesis and polycythemia vera pathophysiology. *Haematologica*, 100(4):439–451, 2015.

[43] M. Iwata and S.-i. Sasa. Theoretical analysis for critical fluctuations of relaxation trajectory near a saddle-node bifurcation. *Phys. Rev. E*, 82:011127, 2010.

[44] C. Jarzynski. A nonequilibrium equality for free energy differences. *Phys. Rev. Lett.*, 78:11, 1996.

[45] C. Jarzynski. Hamiltonian derivation of a detailed fluctuation theorem. *J. Stat. Phys.*, 98:21, 1999.

[46] C. Jarzynski. Rare events and the convergence of exponentially averaged work values. *Phys. Rev. E*, 73(4):46105, 2006.

[47] C. Jarzynski. Equalities and inequalities: irreversibility and the Second Law of Thermodynamics at the nanoscale. *Annu. Rev. Condens. Matter Phys.*, 2:329–351, 2011.

[48] Y. Kanaoka, S. H. Kimura, I. Okazaki, M. Ikeda, and H. Nojima. GAK: a cyclin G associated kinase contains a tensin/auxilin-like domain. *FEBS Lett.*, 402:73–80, 1997.

[49] M. Kardar. *Statistical Physics of Particles*. Cambridge University Press, Cambridge, 2007.

[50] T. Kirchhausen. Imaging endocytic clathrin structures in living cells. *Trends Cell Biol.*, 19:596–605, 2009.

[51] T. Kirchhausen and S. C. Harrison. Protein organization in clathrin trimers. *Cell*, 23(3):755–761, 1981.

[52] T. Kirchhausen, D. Owen, and S. C. Harrison. Molecular structure, function, and dynamics of clathrin-mediated membrane traffic. *Cold Spring Harb. Perspect. Biol.*, 6:a016725, 2014.

[53] T. S. Komatsu and N. Nakagawa. Expression for the stationary distribution in nonequilibrium steady states. *Phys. Rev. Lett.*, 100(3):30601, 2008.

[54] T. S. Komatsu, N. Nakagawa, S.-i. Sasa, and H. Tasaki. Representation of nonequilibrium steady states in large mechanical systems. *J. Stat. Phys.*, 134:401, 2009.

[55] T. S. Komatsu, N. Nakagawa, S.-i. Sasa, and H. Tasaki. Entropy and nonlinear nonequilibrium thermodynamic relation for heat conducting steady states. *J. Stat. Phys.*, 142:127–153, 2011.

[56] M. Krüger and M. Fuchs. Fluctuation dissipation relations in stationary states of interacting Brownian particles under shear. *Phys. Rev. Lett.*, 102:135701, 2009.

[57] T. G. Kutateladze, D. G. S. Capelluto, C. G. Ferguson, M. L. Cheever, A. G. Kutateladze, G. D. Prestwich, and M. Overduin. Multivalent mechanism of membrane insertion by the FYVE domain. *J. Biol. Chem.*, 279(4):3050–3057, 2004.

[58] G. Lan, P. Sartori, S. Neumann, V. Sourjik, and Y. Tu. The energy-speed-accuracy trade-off in sensory adaptation. *Nat. Phys.*, 8(5):422–428, 2012.

[59] A. H. Lang, C. K. Fisher, T. Mora, and P. Mehta. Thermodynamics of statistical inference by cells. *Phys. Rev. Lett.*, 113(14):148103, 2014.

[60] J.-O. Lee, H. Yang, M.-M. Georgescu, A. Di Cristofano, T. Maehama, Y. Shi, J. E. Dixon, P. Pandolfi, and N. P. Pavletich. Crystal structure of the PTEN tumor suppressor. *Cell*, 99(3):323, 1999.

[61] M. A. Lemmon. Membrane recognition by phospholipid-binding domains. *Nat. Rev. Mol. Cell Biol.*, 9(2):99–111, 2008.

[62] M. A. Lemmon, K. M. Ferguson, R. O'Brien, P. B. Sigler, and J. Schlessinger. Specific and high-affinity binding of inositol phosphates to an isolated pleckstrin homology domain. *Proc. Natl. Acad. Sci.*, 92(23):10472–10476, 1995.

[63] H. A. Lorentz. A general theorem concerning the motion of a viscous fluid and a few consequences derived from it. *Versl. Konigl. Akad. Wetensch. Amst.*, 5:168–175, 1896.

[64] D. G. Luchinsky, P. V. E. McClintocky, and M. I. Dykman. Analogue studies of nonlinear systems. *Rep. Prog. Phys.*, 61:889–997, 1998.

[65] C. Maes and K. Netočný. Rigorous meaning of McLennan ensembles. *J. Math. Phys.*, 51:015219, 2010.

[66] U. M. B. Marconi, A. Puglisi, L. Rondoni, and A. Vulpiani. Fluctuation-dissipation: Response theory in statistical physics. *Phys. Rep.*, 461:111–195, 2008.

[67] R. Marsland III and J. England. Far-from-equilibrium distribution from near-steady-state work fluctuations. *Phys. Rev. E*, 92:052120, 2015.

[68] R. Marsland III and J. England. Limits of predictions in thermodynamic systems. Manuscript submitted for publication. 2017.

[69] R. H. Massol, W. Boll, A. M. Griffin, and T. Kirchhausen. A burst of auxilin recruitment determines the onset of clathrin-coated vesicle uncoating. *Proc. Natl. Acad. Sci.*, 103(27):10265–10270, 2006.

[70] M. A. D. Matteis and A. Godi. PI-loting membrane traffic. *Nat. Cell Biol.*, 6(6):487–492, 2004.

[71] J. McLennan. Statistical mechanics of the steady state. *Phys. Rev.*, 115(6):1405–1409, 1959.

[72] P. Mehta, A. H. Lang, and D. J. Schwab. Landauer in the age of synthetic biology: energy consumption and information processing in biochemical networks. *J. Stat. Phys.*, 162:1153, 2016.

[73] P. Mehta and D. J. Schwab. Energetic costs of cellular computation. *Proc. Natl. Acad. Sci.*, 109:17978–17982, 2012.

[74] C. J. Merrifield, M. E. Feldman, L. Wan, and W. Almers. Imaging actin and dynamin recruitment during invagination of single clathrin-coated pits. *Nat. Cell Biol.*, 4(9):691–698, 2002.

[75] U. Moran, R. Phillips, and R. Milo. SnapShot: Key Numbers in Biology. *Cell*, 141:1262, 2010.

[76] A. Murugan, D. A. Huse, and S. Leibler. Discriminatory proofreading regimes in nonequilibrium systems. *Phys. Rev. X*, 4:021016, 2014.

[77] C. R. Myers, R. N. Gutenkunst, and J. P. Sethna. Python unleashed on systems biology. *Comput. Sci. Eng.*, 9(3):34, 2007.

[78] B. Nguyen, D. Hartich, U. Seifert, and P. D. L. Rios. Thermodynamic bounds on the ultra- and infra-affinity of Hsp70 for its substrates. *arXiv Prepr.*, 1702.01649, 2017.

[79] Y. Oono and M. Paniconi. Steady state thermodynamics. *Prog. Theor. Phys. Suppl.*, 130:29–44, 1998.

[80] N. Perunov, R. A. Marsland, and J. L. England. Statistical physics of adaptation. *Phys. Rev. X*, 6:021036, 2016.

[81] M. Polettini, G. Bulnes-Cuetara, and M. Esposito. Conservation laws and symmetries in stochastic thermodynamics. *Phys. Rev. E*, 94:052117, 2016.

[82] R. Rao and L. Peliti. Thermodynamics of accuracy in kinetic proofreading: dissipation and efficiency trade-offs. *J. Stat. Mech. Theory Exp.*, 2015:P06001, 2015.

[83] T. F. Roth and K. R. Porter. Yolk protein uptake in the oocyte of the mosquito *Aedes Aegypti* L. *J. Cell Biol.*, 20(2), 1964.

[84] M. S. P. Sansom, P. J. Bond, S. S. Deol, A. Grottesi, S. Haider, and Z. A. Sands. Molecular simulations and lipid-protein interactions: potassium channels and other membrane proteins. *Biochem. Soc. Trans.*, 33(5):916, 2005.

[85] P. Sartori, L. Granger, C. F. Lee, and J. M. Horowitz. Thermodynamic costs of information processing in sensory adaptation. *PLoS Comput. Biol.*, 10:e1003974, dec 2014.

[86] S.-i. Sasa. Possible extended forms of thermodynamic entropy. *J. Stat. Mech. Theory Exp.*, 2014(1):P01004, 2014.

[87] E. M. Schmid and H. T. McMahon. Integrating molecular and network biology to decode endocytosis. *Nature*, 448:883–888, 2007.

[88] S. Schoebel, W. Blankenfeldt, R. S. Goody, and A. Itzen. High-affinity binding of phosphatidylinositol 4-phosphate by Legionella pneumophila DrrA. *EMBO Rep.*, 11(8):598–604, 2010.

[89] E. Schrodinger. *What is Life?: The Physical Aspect of the Living Cell.* Cambridge University Press, Cambridge, 1948.

[90] U. Seifert. Configurations of fluid membranes and vesicles. *Adv. Phys.*, 46(1):13–137, 1997.

[91] U. Seifert. Stochastic thermodynamics, fluctuation theorems and molecular machines. *Rep. Prog. Phys.*, 75:126001, 2012.

[92] R. Sousa. Structural mechanisms of chaperone mediated protein disaggregation. *Front. Mol. Biosci.*, 1:12, 2014.

[93] R. Sousa and E. M. Lafer. The role of molecular chaperones in clathrin mediated vesicular trafficking. *Front. Mol. Biosci.*, 2:26, 2015.

[94] T. Speck and U. Seifert. Extended fluctuation-dissipation theorem for soft matter in stationary flow. *Phys. Rev. E*, 79:040102(R), 2009.

[95] H. Touchette. The large deviation approach to statistical mechanics. *Phys. Rep.*, 478:1–69, 2009.

[96] M. K. Transtrum, B. B. Machta, K. S. Brown, B. C. Daniels, C. R. Myers, and J. P. Sethna. Perspective: Sloppiness and emergent theories in physics, biology, and beyond. *J. Chem. Phys.*, 143:010901, 2015.

[97] E. Ungewickell, H. Ungewickell, S. E. H. Holstein, R. Lindner, K. Prasad, W. Barouch, B. Martin, L. E. Greene, and E. Elsenberg. Role of auxilin in uncoating clathrin-coated vesicles. *Nature*, 378:632, 1995.

[98] M. Zhang and G. Szamel. Effective temperatures of a driven, strongly anisotropic Brownian system. *Phys. Rev. E*, 83:061407, 2011.

Appendix A

Microscopic Reversibility in the Langevin Equation

In the main text, I described the physical basis of microscopic reversibility in general terms, and expressed it in terms of standard expressions for reservoir entropies. But any given application of these results will involve additional modeling assumptions, which can sometimes capture the phenomenon of interest quite well while distorting its thermodynamic properties (cf. [39]). Before applying results based on the microscopic reversibility relation (2.24) to the statistics of a particular model containing its own definitions of work and energy, it is important to verify explicitly that this relation is consistent with the equations of the model.

In this appendix, I perform this verification for two different interpretations of the Langevin Equation. I first consider the full dynamics of Brownian motion contained in Equations (2.44) and (2.45), with a δ-correlated random force as defined in Equations (2.46-2.48). Then I examine the coarse-grained versions underlying the simulation of Chapter 3 and the general discussion of flow-driven systems in Chapter 2.

The core mathematical result required for applying the microscopic reversibility relation to a Langevin equation is the explicit expression for the trajectory probabilities in terms of the equation parameters. This result is independent of the physical interpretation, and is best expressed in generic terms to avoid confusion with the physics. Consider the following Langevin equation for the evolution of a d-dimensional vector

$\mathbf{x}$:

$$\dot{\mathbf{x}} = \mathbf{A}_t(\mathbf{x}_t) + BM\boldsymbol{\xi}_t \tag{A.1}$$

where the deterministic part $\mathbf{A}_t(\mathbf{x}_t)$ is a function of $\mathbf{x}_t$ that can also depend explicitly on time. The matrix M is included to admit the possibility of situations like Equations (2.44-2.45) where some of the dynamical variables are not directly coupled to the noise term. This matrix is equal to the identity matrix, except that some of the elements can be set to zero to remove the $\boldsymbol{\xi}$-dependence from the corresponding equations. I will denote by $\mathbf{x}'$ a reduced state vector containing only those variables that are directly subject to noise, and by $\mathbf{A}'_t(\mathbf{x}_t)$ the corresponding force vector. For simplicity, I will take B to be a scalar in this section.

The probability of observing a given trajectory $\mathbf{x}_0^{\mathcal{T}}$, given that the system is initialized at the correct value $\mathbf{x}_0$, is simply the probability of observing the noise realization $\boldsymbol{\xi}_0^{\mathcal{T}}$ that generates this trajectory when Equation (A.1) is integrated. The result can be expressed as (cf. [26, 43]):

$$p[\mathbf{x}_0^{\mathcal{T}}|\mathbf{x}_0] = \exp\left[-\int_0^{\mathcal{T}} dt\, \left(\frac{1}{2B^2}[\dot{\mathbf{x}}' - \mathbf{A}'_t]^2 + \frac{1}{2}\nabla\cdot\mathbf{A}'_t\right)\right] \tag{A.2}$$

where the product $\mathbf{A}'_t \cdot \dot{\mathbf{x}}$ is performed under the Stratonovich interpretation [29].

A.1 Langevin Equation for Brownian Motion

For the case of Brownian motion (2.44-2.45) in d dimensions, the probability of a trajectory $\mathbf{x}_0^{\mathcal{T}} = (\mathbf{p}_0^{\mathcal{T}}, \mathbf{q}_0^{\mathcal{T}})$ is found by appropriately substituting for $\mathbf{A}'_t$, B and $\mathbf{x}$ in Equation (A.1):

$$p[\mathbf{x}_0^{\mathcal{T}}|\mathbf{x}_0, \lambda_0^{\mathcal{T}}] = \exp\left[-\int_0^{\mathcal{T}} dt\, \left(\frac{1}{2k^2}\left[\dot{\mathbf{p}}_t - \mathbf{f}(\mathbf{q}_t, \lambda_t) + \frac{b}{m}\mathbf{p}_t\right]^2 - d\frac{b}{2m}\right)\right]. \tag{A.3}$$

The reverse trajectory probability is found by transforming $\mathbf{p} \to -\mathbf{p}$:

$$p[\hat{\mathbf{x}}_0^{\mathcal{T}} | \mathbf{x}_{\mathcal{T}}^*, \hat{\lambda}_0^{\mathcal{T}}] = \exp\left[- \int_0^{\mathcal{T}} dt \left(\frac{1}{2k^2} \left[\dot{\mathbf{p}}_{\mathcal{T}-t} - \mathbf{f}(\mathbf{q}_{\mathcal{T}-t}, \lambda_{\mathcal{T}-t}) - \frac{b}{m} \mathbf{p}_{\mathcal{T}-t} \right]^2 - d \frac{b}{2m} \right) \right] \tag{A.4}$$

$$= \exp\left[- \int_0^{\mathcal{T}} dt \left(\frac{1}{2k^2} \left[\dot{\mathbf{p}}_t - \mathbf{f}(\mathbf{q}_t, \lambda_t) - \frac{b}{m} \mathbf{p}_t \right]^2 - d \frac{b}{m} \right) \right] \tag{A.5}$$

where I have changed the variable of integration from t to $\mathcal{T} - t$ and inverted the limits of integration in the second line. Combining these expressions yields:

$$\frac{p[\hat{\mathbf{x}}_0^{\mathcal{T}} | \mathbf{x}_{\mathcal{T}}^*, \hat{\lambda}_0^{\mathcal{T}}]}{p[\mathbf{x}_0^{\mathcal{T}} | \mathbf{x}_0, \lambda_0^{\mathcal{T}}]} = \exp\left[\frac{b}{mk^2} \Delta \mathbf{p}^2 - \frac{2b}{mk^2} \int_0^{\mathcal{T}} dt \, \mathbf{f} \cdot \mathbf{p}_t \right]. \tag{A.6}$$

Now I split the force into an externally supplied part and a part due to an internal potential energy landscape:

$$\mathbf{f} = \mathbf{f}_{\text{ext}}(\mathbf{q}, \lambda) - \nabla U(\mathbf{q}, \lambda). \tag{A.7}$$

With these definitions, I can rearrange (A.6) into a more suggestive form:

$$\frac{p[\hat{\mathbf{x}}_0^{\mathcal{T}} | \mathbf{x}_{\mathcal{T}}^*, \hat{\lambda}_0^{\mathcal{T}}]}{p[\mathbf{x}_0^{\mathcal{T}} | \mathbf{x}_0, \lambda_0^{\mathcal{T}}]} = \exp\left[-\frac{2b}{k^2} \int_0^{\mathcal{T}} dt \left(\mathbf{f}_{\text{ext}} \cdot \frac{\mathbf{p}_t}{m} + \frac{\partial U}{\partial \lambda} \dot{\lambda}_t \right) + \left(\frac{b}{mk^2} \Delta \mathbf{p}^2 + \frac{2b}{k^2} \Delta U \right) \right], \tag{A.8}$$

where I have used the chain rule $dU = \nabla U \cdot \dot{\mathbf{q}} \, dt + \partial_\lambda U \dot{\lambda}_t \, dt$ to express the ∇U term in terms of $\partial_\lambda U$ and ΔU. This is allowed thanks to the Stratonovich interpretation of the product stipulated above. The integral on the right-hand side is equal to β times the work done and the final term in parentheses to β times the change in energy when

$$k^2 = 2k_B T b. \tag{A.9}$$

I have thus confirmed that Equations (2.44-2.45) satisfy the microscopic reversibility relation for generic choices of $\mathbf{f}_{\text{ext}}$ and U, including the possibility of external manipulation of control parameters λ, and found the value of k that makes this happen.

Note that the expression for the work

$$W = \int_0^{\mathcal{T}} \left(\mathbf{f}_{\text{ext}} \cdot \frac{\mathbf{p}_t}{m} + \frac{\partial U}{\partial \lambda} \dot{\lambda}_t \right) \tag{A.10}$$

includes a new term $\mathbf{f}_{\text{ext}} \cdot \mathbf{p}_t/m$ that was not present in the original definition of work within the Hamiltonian derivation (2.26). This term accounts for the possibility of a non-conservative force, which drags the particle in a loop while continually dissipating heat. When actually realized in experiments, such forces are always produced by periodic variations in some control parameters λ. But representing the net effect of sufficiently rapid variations as a constant nonconservative force greatly simplifies the analysis.

A.2 Colloid Simulation with Externally Imposed Flows

Consider the general form of the overdamped Langevin equation of Equations (3.1) and (3.2) from Chapter 3, describing the motion of N identical colloidal particles with drag coefficient b in d dimensions:

$$b\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + \sqrt{2k_BTb}\boldsymbol{\xi}_t. \tag{A.11}$$

The dN-dimensional vector $\mathbf{x}$ contains the positions of all N particles, and the force vector $\mathbf{f}(\mathbf{x})$ mediates interactions among them.

The trajectory probabilities of (A.11) can be computed using the general form (A.1), yielding:

$$p[\mathbf{x}_0^{\mathcal{T}}|\mathbf{x}_0] \propto \exp\left[-\int_0^{\mathcal{T}} dt \left(\frac{(\dot{\mathbf{x}} - \mathbf{f}/b)^2}{4k_BT/b} + \frac{1}{2}\nabla \cdot \frac{\mathbf{f}}{b} \right) \right]. \tag{A.12}$$

I now consider the case where $\mathbf{f} = -\nabla U + b\mathbf{u}$ includes a conservative force $\mathbf{f}_c = -\nabla U$ and a contribution $b\mathbf{u}$ from the local solvent velocity $\mathbf{u}$.

Applying the time-reversal operation to the trajectory probability expression reverses the signs of both $\dot{\mathbf{x}}$ and $\mathbf{u}$. In a real experiment, this reversal would come (for example) from reversing the magnetic field in an electric motor driven by sinusoidal voltage oscillations. This justifies my continued use of the symbols $\lambda_0^{\mathcal{T}}$ and $\hat{\lambda}_0^{\mathcal{T}}$ to distinguish between the forward and reverse dynamics. The left-hand side of the microscopic reversibility relation (2.24) now reads:

$$\frac{p[\hat{\mathbf{x}}_0^{\mathcal{T}}|\mathbf{x}_{\mathcal{T}}^*, \hat{\lambda}_0^{\mathcal{T}}]}{p[\mathbf{x}_0^{\mathcal{T}}|\mathbf{x}_0, \lambda_0^{\mathcal{T}}]} = \exp\left[\frac{1}{k_B T}\int_0^{\mathcal{T}} dt\,\nabla U \cdot \dot{\mathbf{x}} - \frac{1}{k_B T}\int_0^{\mathcal{T}} dt\,\nabla U \cdot \mathbf{u}\right] \tag{A.13}$$

By the chain rule, the first term in the exponential is simply the change in energy ΔU over the trajectory (the control parameters driving the motor do not affect the interparticle potential, so there is no $\partial_\lambda U$ term). To understand the second term, note that my assumption of overdamped dynamics implies that the total force on each particle is always zero. The force exerted by the solvent must exactly cancel the force due to interparticle interactions or external fields, and thus equals ∇U. I have already defined that the local speed of the solvent is equal to $\mathbf{u}$, and so the work done by the imposed flow field is

$$\mathcal{W} = \int_0^{\mathcal{T}} dt\,\nabla U \cdot \mathbf{u}. \tag{A.14}$$

Note that the work rate $\dot{\mathcal{W}} = \nabla U \cdot \mathbf{u}$ is fully determined by the particle positions $\mathbf{x}$ (which fix U and its derivatives), and is independent of the velocities.

Using the First Law of thermodynamics (2.32), I find

$$Q = -\Delta U + \mathcal{W} = -\int_0^{\mathcal{T}} dt\,\nabla U \cdot \dot{\mathbf{x}} + \int_0^{\mathcal{T}} dt\,\nabla U \cdot \mathbf{u}. \tag{A.15}$$

But this is exactly equal to $-k_B T$ times the exponent in Equation (A.13). I have thus verified the microscopic reversibility equation (2.24) with $\Delta S_e = Q/T$:

$$\frac{p[\hat{\mathbf{x}}_0^{\mathcal{T}}|\mathbf{x}_{\mathcal{T}}^*, \hat{\lambda}_0^{\mathcal{T}}]}{p[\mathbf{x}_0^{\mathcal{T}}|\mathbf{x}_0, \lambda_0^{\mathcal{T}}]} = e^{-\frac{Q}{k_B T}}. \tag{A.16}$$

The simple shear flow field investigated in Chapter 3 is $\mathbf{u} = \sum_i \dot{\gamma} y \hat{x}_i$, where $\hat{x}_i$ is the unit vector in the x direction for particle i. The work is then given by

$$\mathcal{W} = \dot{\gamma} \sum_i \int_0^{\mathcal{T}} dt \, \frac{\partial U}{\partial x_i} y_i \tag{A.17}$$

which is exactly what I found in Equations (3.6) and (3.7) using the result of Appendix D for the force required to move the top wall of the container.

Appendix B

Dual Processes in Multiple Dimensions

The theory of Chapter 2 is based on the statistics of trajectories that generate a given fluctuation. These statistics are conveniently expressed in terms of a "dual" dynamics, in which the ensemble of trajectories *initialized* at a given point $\mathbf{X} = \mathbf{X}_0$ is the time-reverse of the ensemble of trajectories *ending* at $\mathbf{X}_0$ in the original dynamics. I can state this requirement mathematically by requiring that the conditional trajectory probabilities $p^\dagger[\mathbf{X}_0^{\mathcal{T}}|\mathbf{X}_0]$ satisfy:

$$p_{\mathrm{ss}}(\mathbf{X}_0)p[\mathbf{X}_0^{\mathcal{T}}|\mathbf{X}_0] = p_{\mathrm{ss}}(\mathbf{X}^*_{\mathcal{T}})p^\dagger[\hat{\mathbf{X}}_0^{\mathcal{T}}|\mathbf{X}^*_{\mathcal{T}}]. \tag{B.1}$$

I will restrict my attention to the case $\mathbf{X}^* = \mathbf{X}$ where the phase space region corresponding to $\mathbf{X}$ is symmetric under time reversal.

For a general multidimensional Langevin equation

$$\dot{\mathbf{X}} = \mathbf{F}(\mathbf{X}) + B(\mathbf{X})\boldsymbol{\xi}_t \tag{B.2}$$

with steady-state distribution

$$p_{\mathrm{ss}}(\mathbf{X}) = e^{-\phi(\mathbf{X})} \tag{B.3}$$

the dual dynamics are given by [36, 15]:

$$\dot{\mathbf{X}} = -\mathbf{F} - B^2\nabla\phi + B\boldsymbol{\xi}. \tag{B.4}$$

The fluctuation trajectories are found by flipping the direction of time, so that

$$\dot{\mathbf{X}} = \mathbf{F} + B^2\nabla\phi + B\boldsymbol{\xi}. \tag{B.5}$$

and the initial condition is changed into a final condition, requiring that all trajectories *end* at the same point $\mathbf{X} = \mathbf{X}_{\mathcal{T}}$.

Now I specialize to the linear case $\mathbf{F} = -A\mathbf{X}$, where A is a constant matrix independent of $\mathbf{X}$. To find the dual dynamics, I need the steady-state distribution so I can compute $\nabla\phi$. To simplify the notation, I first rescale $\mathbf{X}$ according to the noise strength for each degree of freedom. The equation of motion for the rescaled variables $\hat{\mathbf{X}} = B^{-1}\mathbf{X}$ becomes

$$\dot{\hat{\mathbf{X}}} = -B^{-1}AB + \boldsymbol{\xi} \equiv -\hat{A} + \boldsymbol{\xi}. \tag{B.6}$$

I can now find the steady state by solving the Fokker-Planck Equation

$$\begin{aligned} 0 = \partial_t p_{\text{ss}} &= -\nabla^T\left(-\hat{A}\hat{\mathbf{X}}p_{\text{ss}} - \frac{1}{2}\nabla p_{\text{ss}}\right) && \text{(B.7)} \\ &= -\nabla^T p_{\text{ss}}(\hat{\mathbf{X}})\left(-\hat{A}\hat{\mathbf{X}} + \frac{1}{2}\nabla\phi\right) && \text{(B.8)} \end{aligned}$$

where ∇^T is a row vector formed by the partial derivative operators ∂_i. If $\hat{A}$ is symmetric, the only solution is $\nabla\phi = 2\hat{A}\mathbf{X}$, and so the dual dynamics are simply

$$\dot{\hat{\mathbf{X}}} = \hat{A}\hat{\mathbf{X}} + \boldsymbol{\xi}. \tag{B.9}$$

In general, the calculation is slightly more complicated, because $\mathbf{X}$ may relax to zero while circulating around the origin instead of traveling there directly. In that case, the fluctuation ensemble should reverse the sign of the direct part of the relaxation while maintaining the same direction of circulation. To find the dual dynamics in this case, it is helpful to write Equation (B.7) in terms of ϕ:

$$(\nabla\phi)^T \left(-\hat{A}\hat{\mathbf{X}} + \frac{1}{2}\nabla\phi\right) = -\mathrm{Tr}(\hat{A}) + \frac{1}{2}\nabla^2\phi. \tag{B.10}$$

Since the equations of motion are linear and the noise is Gaussian, $\mathbf{X}$ is a sum of Gaussian random variables, and thus always follows a Gaussian distribution itself. This means that the logarithm ϕ of the steady-state distribution must be quadratic in $\hat{\mathbf{X}}$:

$$\phi = \hat{\mathbf{X}}^T \hat{C} \hat{\mathbf{X}} \tag{B.11}$$

where $\hat{C}$ is a positive definite symmetric matrix (if it had an antisymmetric part, this would vanish in the evaluation of the quadratic form). So Equation (B.10) becomes

$$2\mathbf{X}^T \hat{C}(-\hat{A} + \hat{C})\hat{\mathbf{X}} = \mathrm{Tr}(-\hat{A} + \hat{C}). \tag{B.12}$$

Since the right-hand side does not depend on $\hat{\mathbf{X}}$, the only way to make this true for all $\hat{\mathbf{X}}$ is to make both sides vanish, which implies that the symmetric part of the matrix between the $\hat{\mathbf{X}}^T$ and $\hat{\mathbf{X}}$ on the left hand side must vanish. $\hat{C}$ must therefore satisfy these two equations:

$$\hat{C}(-\hat{A} + \hat{C}) + (-\hat{A} + \hat{C})^T \hat{C} = 0 \tag{B.13}$$

$$\mathrm{Tr}(-\hat{A} + \hat{C}) = 0. \tag{B.14}$$

The first condition can be simplified to

$$\hat{A} + \hat{C}^{-1}\hat{A}^T\hat{C} = 2\hat{C}. \tag{B.15}$$

Now the equation for the fluctuation trajectories can be found by plugging in to

Equation (B.5):

$$\dot{\hat{\mathbf{X}}} = (-\hat{A} + 2\hat{C})\hat{\mathbf{X}} + \boldsymbol{\xi} \tag{B.16}$$

$$= \hat{C}^{-1}\hat{A}^T\hat{C}\hat{\mathbf{X}} + \boldsymbol{\xi}. \tag{B.17}$$

Using $\hat{C} = BCB$ to switch the first term back to the original variables yields

$$\hat{C}^{-1}\hat{A}^T\hat{C}\hat{\mathbf{X}} = B^{-1}C^{-1}B^{-1}BA^TB^{-1}BCBB^{-1}B\mathbf{X} = B^{-1}C^{-1}A^TC\mathbf{X}. \tag{B.18}$$

Thus I conclude that the ensemble leading from the steady state to the fluctuation $\mathbf{X}(0)$ is described by

$$\dot{\mathbf{X}} = C^{-1}A^TC\mathbf{X} + B\boldsymbol{\xi}, \tag{B.19}$$

where $C^{-1}A^TC$ is denoted in the main text as $\tilde{A}$.

This can be further simplified if $\hat{A}$ is normal, i.e., if $\hat{A}\hat{A}^T = \hat{A}^T\hat{A}$. To see this, I first decompose $\hat{A} = \hat{A}_S + \hat{A}_A$ into a symmetric part $\hat{A}_S = \hat{A}_S^T$ and an antisymmetric part $\hat{A}_A = -\hat{A}_A^T$, without loss of generality. Then $\hat{A}\hat{A}^T = \hat{A}^T\hat{A}$ implies that the symmetric and antisymmetric parts commute: $[\hat{A}_S, \hat{A}_A] = 0$. I now guess that $\hat{C} = \hat{A}_S$. $[\hat{A}_S, \hat{A}_A] = 0$ now implies $[\hat{C}, \hat{A}_A] = 0$ and therefore also $[\hat{C}, \hat{A}] = 0$. Using this fact in the first of the two conditions (B.13-B.14) and noting that antisymmetric matrices are traceless, I can easily verify that both are satisfied, confirming that the steady-state distribution is indeed given in the rescaled variables by $\hat{C} = \hat{A}_S$. Inserting this into Equation (B.16) and proceeding through the rest of the steps eliminates the change of basis in the first term of (B.19), yielding

$$\dot{\mathbf{X}} = A^T\mathbf{X} + B\boldsymbol{\xi}. \tag{B.20}$$

The normality of $\hat{A}$ is determined by the physical properties of the system, and cannot in general be achieved by a simple change of basis. Even when A can be made normal by an appropriate (non-unitary) basis change, another non-unitary basis change is required to transform from A to $\hat{A}$, which can make $\hat{A}$ non-normal.

Appendix C

Perturbative Calculations of Work Statistics for Nonlinear Macroscopic Dynamics

In this chapter, I compute $\mathcal{W}_{\text{ex}}(X)$ and $\Phi_{\text{ex}}(X)$ for the one-dimensional nonlinear macroscopic dynamics (2.79) to first order in the small parameter ϵ that controls the size of the nonlinearity. The calculation is slightly different depending on the choice of thermodynamic interpretation. I first consider the case where the work comes from an externally imposed flow field. The other forms of driving, including nonconservative forces, time-varying conservative forces, energy/matter transport and chemical reactions, all share the same basic structure for the core computation, and are considered together in the second subsection.

C.1 Mathematical Preliminaries

The ensemble of fluctuation trajectories $X^{0}_{-\mathcal{T}}$ that end in a given state X_0 is fully contained in Equation (2.79), independent of the interpretation. Combining Equations

(2.80), (2.81) and (2.82) yields for X_t:

$$X_t = e^{At}(X_0 + f_t) - \frac{\epsilon}{2}\int_t^0 dt'\, e^{A(t-t')}[e^{At'}(X_0 + f_{t'})]^2 + O(\epsilon^2) \tag{C.1}$$

where $f_t = -\int_t^0 e^{-At'} B\xi_{t'} dt'$. The only random term in this expression that is not Gaussian is $f_{t'}^2$. It turns out that this term will only enter the calculations as part of a X_0-independent term or at order ϵ^2. So the X_0-dependent terms at order ϵ in $\mathcal{W}_{\text{ex}}$ and Φ_{ex} depend only on the average $\langle X_t \rangle_{X_0}$ and two-point function $\langle X_t X_{t'} \rangle_{X_0}^c \equiv \langle X_t X_{t'} \rangle_{X_0} - \langle X_t \rangle_{X_0} \langle X_{t'} \rangle_{X_0}$. I now proceed to compute these quantities in terms of the given parameters.

Since $\langle f_t \rangle = 0$, the conditional average is

$$\langle X_t \rangle_{X_0} = e^{At} X_0 - \frac{\epsilon}{2A} e^{At}(1 - e^{At}) X_0^2 + O(\epsilon^2) + \mathcal{N} \tag{C.2}$$

where the constant $\mathcal{N}$ contains the $f_{t'}^2$ term, which is independent of X_0.

The two-point function is

$$\begin{aligned}\langle X_t X_{t'} \rangle_{X_0}^c = \epsilon X_0 \left[\int_{t'}^0 dt''\, e^{A(t+t'+t'')} \langle f_t f_{t''} \rangle + \int_t^0 dt''\, e^{A(t+t'+t'')} \langle f_{t'} f_{t''} \rangle \right] \\ + e^{A(t+t')} \langle f_t f_t' \rangle + \epsilon \mathcal{N}' + O(\epsilon^2)\end{aligned} \tag{C.3}$$

where $\mathcal{N}'$ is another constant independent of X_0.

Further simplification requires computing $\langle f_t f_{t'} \rangle$. I first consider the case $|t'| \geq |t|$:

$$\begin{aligned}\langle f_t f_{t'} \rangle &= B^2 \int_t^0 ds \int_{t'}^0 du\, e^{-A(s+u)} \langle \xi(s) \xi(u) \rangle \\ &= B^2 \int_t^0 ds \int_{t'}^t du\, e^{-A(s+u)} \delta(s-u) \\ &\quad + B^2 \int_t^0 ds \int_t^0 du\, e^{-A(s+u)} \delta(s-u) \\ &= B^2 \int_t^0 ds\, e^{-2As} = \frac{B^2}{2A}(e^{-2At} - 1).\end{aligned} \tag{C.4}$$

If $|t'| \leq |t|$, this becomes

$$\langle f_t f_{t'} \rangle = \frac{B^2}{2A}(e^{-2At'} - 1). \tag{C.5}$$

Combining the two answers yields

$$\langle f_t f_{t'} \rangle = \frac{B^2}{2A}(e^{-2At_m} - 1) \tag{C.6}$$

where t_m is equal to whichever of t, t' has the smaller absolute value. Inserting this back into the expression for the two-point function of X_t, I find

$$\begin{aligned}\langle X_t X_{t'} \rangle^c_{X_0} = \epsilon X_0 \frac{B^2}{2A^2} &\left[-e^{-A(t-2t')} - 4e^{A(t'+t)} + 3e^{At'} + e^{A(t+2t')} + e^{A(t'+2t)}\right] \\ &+ \frac{B^2}{2A}(e^{-A|t-t'|} - e^{A(t+t')}) + \epsilon \mathcal{N}' + O(\epsilon^2).\end{aligned} \tag{C.7}$$

Enforcing symmetry under $t \to t'$ yields

$$\begin{aligned}\langle X_t X_{t'} \rangle^c_{X_0} = \epsilon X_0 \frac{B^2}{2A^2} &\left[-e^{-A(|t-t'|-t_M)} - 4e^{A(t'+t)} + 3e^{At_M} + e^{A(t+2t')} + e^{A(t'+2t)}\right] \\ &+ \frac{B^2}{2A}(e^{-A|t-t'|} - e^{A(t+t')}) + \epsilon \mathcal{N}' + O(\epsilon^2).\end{aligned} \tag{C.8}$$

where t_M is whichever of t, t' has the larger absolute value.

For some of the calculations, I will need to use the fact that the integral of the expression in brackets over all times is:

$$\begin{aligned}\int_{-\infty}^{0} dt \int_{-\infty}^{0} dt' \Big[&-e^{-A(|t-t'|-t_M)} - 4e^{A(t'+t)} \\ &+ 3e^{At_M} + e^{A(t+2t')} + e^{A(t'+2t)}\Big] = \frac{2}{A^2}.\end{aligned} \tag{C.9}$$

I will also make use of the correlator between X_t and $\xi_{t'}$:

$$\langle X_t \xi_{t'} \rangle_{X_0} = \epsilon X_0 \int_t^0 e^{A(t+t'')} \langle f_{t''} \xi_{t'} \rangle dt'' + e^{At} \langle f_t \xi_{t'} \rangle + \epsilon \mathcal{N}'' + O(\epsilon^2) \tag{C.10}$$

$$= -\epsilon X_0 B \int_t^0 e^{A(t+t''-t')} \Theta(t'-t'') dt'' - B e^{A(t-t')} \Theta(t'-t) + \epsilon \mathcal{N}'' + O(\epsilon^2) \tag{C.11}$$

$$= -B\Theta(t'-t) e^{A(t-t')} [\epsilon X_0 \frac{1}{A}(e^{At'} - e^{At}) + 1] + \epsilon \mathcal{N}'' + O(\epsilon^2) \tag{C.12}$$

where $\mathcal{N}''$ again is independent of X_0 and $\Theta(x)$ is the Heaviside step function, equal to zero for $x < 0$ and 1 for $x \geq 0$.

Because I will only be looking at the Gaussian part of the fluctuations, all higher-order correlations can be expressed in terms of the two-point function and the average. In particular, I will need the fact that

$$\langle X_t, X_{t'}^2 \rangle_{X_0}^c \equiv \langle X_t X_{t'}^2 \rangle_{X_0} - \langle X_t \rangle_{X_0} \langle X_{t'}^2 \rangle_{X_0} \tag{C.13}$$

$$= 2 \langle X_{t'} \rangle_{X_0} \langle X_t X_{t'} \rangle_{X_0}^c \tag{C.14}$$

and that

$$\langle X_t^2, \xi_{t'} \rangle_{X_0}^c = 2 \langle X_t \xi_{t'} \rangle_{X_0} \langle X_t \rangle_{X_0}. \tag{C.15}$$

C.2 Driving by Imposed Flow

When the system is driven by an imposed flow field as discussed in Appendix A above, the macroscopic variable can be chosen as the excess current $J - J_{\mathrm{ss}}$ beyond the steady-state mean value J_{ss}. Then the work is $\mathcal{W} = V\mathcal{F} \int_{-\mathcal{T}}^0 dt\, X_t + \mathcal{W}_0$ where $\mathcal{W}_0$ is a constant, independent of the trajectory $X_{-\mathcal{T}}^0$.

The excess work defined in Equations (2.59) and (2.57) can be found by integrating the exponentials in Equation (C.2):

$$\begin{aligned}\mathcal{W}_{\text{ex}}(X_0) &= V\mathcal{F}\int_{-\infty}^{0} dt\,(\langle X_t\rangle_{X_0} - \langle X_t\rangle_{\text{ss}}) \\ &= \frac{V\mathcal{F}}{A}X_0 + \epsilon\frac{V\mathcal{F}}{4A^2}X_0^2 + O(\epsilon^2) \\ &= \mathcal{W}_{\text{ex}}^{(0)}(X_0) + \epsilon\frac{V\mathcal{F}}{4A(\mathcal{F})^2}(X_0^2 - \langle X_0^2\rangle_{\text{ss}}) + O(\epsilon^2), \end{aligned} \qquad \text{(C.16)}$$

where $\mathcal{W}_{\text{ex}}^{(0)}(X_0)$ is the answer obtained in Equation (2.71) under the linear dynamics alone.

To compute the higher cumulants, I first note that the non-Gaussian term $\epsilon f_{t'}^2$ in Equation (C.1) can only be part of an X_0-dependent term in $\Delta\langle\mathcal{W}^n\rangle^c$ when it is multiplied by some nonzero power of the $\epsilon X_0 f_{t'}$ term. It therefore contributes only to the $O(\epsilon^2)$ part of the expression and to the overall normalization. The remaining part of the work can be expressed as a sum of independent Gaussian random variables ξ_t, so it is itself a Gaussian random variable and has no nonzero cumulants beyond the variance.

The excess variance can be computed from the expression for the two-point function (C.8) and its integral obtained above:

$$\Delta\langle\mathcal{W}^2\rangle^c(X_0) = V^2\mathcal{F}^2\int_{-\infty}^{0} dt\int_{-\infty}^{0} dt'\,(\langle X_t X_{t'}\rangle_{X_0}^c - \langle X_t X_{t'}\rangle_{\text{ss}}^c) \qquad \text{(C.17)}$$

$$= V^2\mathcal{F}^2\epsilon\frac{B^2}{A^4}X_0 + O(\epsilon^2). \qquad \text{(C.18)}$$

Since the higher cumulants are higher-order in ϵ, the excess fluctuations are:

$$\Phi_{\text{ex}}(X) = \epsilon\frac{\beta^2V^2\mathcal{F}^2B^2}{2A^4}X + O(\epsilon^2) \qquad \text{(C.19)}$$

$$= \tilde{\epsilon}\beta\mathcal{W}_{\text{ex}}(X) + O(\tilde{\epsilon}^2) \qquad \text{(C.20)}$$

so that $\tilde{\epsilon} = \epsilon\frac{\beta V\mathcal{F}B^2}{2A^3}$ is the appropriate dimensionless version of ϵ that controls how quickly the expansion of $\ln p_{\text{ss}}$ about the linearized dynamics converges.

C.3 Driving by Thermal/Chemical/Mechanical forces

The work rate for the chemical driving of Equation (2.95) takes the form

$$\dot{\mathcal{W}} = a_t X + \epsilon c_t X^2 + b_t \xi_t + \dot{\mathcal{W}}_0 \tag{C.21}$$

where a_t, b_t and c_t are independent of X. The work rates for thermal and mechanical driving take this same form to first order around the linearized dynamics. I have included the subscript t in order to set up a framework that includes the possibility of periodically varying driving forces, although I do not directly apply this formalism to such cases in the present work.

The excess work is

$$\mathcal{W}_{\text{ex}}(X_0) = \int_{-\infty}^{0} dt\, [a_t(\langle X_t\rangle_{X_0} - \langle X_t\rangle_{\text{ss}}) + \epsilon c_t(\langle X_t^2\rangle_{X_0} - \langle X_t^2\rangle_{\text{ss}})]. \tag{C.22}$$

For time-independent thermal or chemical driving, the integral of the first term is identical to what was computed in the previous section, and the second term reduces to

$$\langle X_t^2\rangle_{X_0} = e^{2At} X_0^2 + O(\epsilon) \tag{C.23}$$

since $\langle X_t^2\rangle_{X_0}^c$ is $O(\epsilon)$. This gives:

$$\mathcal{W}_{\text{ex}}(X_0) = \frac{a}{A} X_0 + \epsilon \left(\frac{a}{4A^2} + \frac{c}{2A} \right) X_0^2 + O(\epsilon^2) + \mathcal{N} \tag{C.24}$$

where $\mathcal{N}$ contains terms independent of X_0.

The X_0-dependent parts of the work fluctuations are Gaussian up to order ϵ for the same reasons stated in the previous section. The excess variance is given by

$$\begin{aligned}\Delta\langle \mathcal{W}^2\rangle^c(X_0) = \int_{-\infty}^{0} dt \int_{-\infty}^{0} dt'\, (a_t a_{t'} \langle X_t X_{t'}\rangle_{X_0}^c + 4\epsilon a_t c_{t'} \langle X_{t'}\rangle_{X_0} \langle X_t X_{t'}\rangle_{X_0}^c \\ + 2a_t b_{t'} \langle X_t \xi_{t'}\rangle_{X_0} + 4\epsilon c_t b_{t'} \langle X_t \xi_{t'}\rangle_{X_0} \langle X_t\rangle_{X_0}) + O(\epsilon^2) + \mathcal{N}\end{aligned} \tag{C.25}$$

where I have expressed all higher-order correlations in terms of the average and two-point functions, as explained in Section C.1 above.

For constant driving, plugging in the expressions from Section C.1 gives:

$$\begin{aligned}\Phi_{\text{ex}}(X_0) &= \frac{\beta^2}{2}\Delta\langle\mathcal{W}^2\rangle^c(X_0) + O(\epsilon^2) &\text{(C.26)}\\ &= \frac{\beta^2}{2}\epsilon\left(\frac{a^2B^2}{A^4} + \frac{2acB^2}{A^3} - \frac{abB}{A^3} - \frac{3cbB}{A^2}\right)X_0 + O(\epsilon^2) &\text{(C.27)}\\ &= \beta\mathcal{W}_{\text{ex}}(X_0)\tilde{\epsilon} + O(\epsilon^2). &\text{(C.28)}\end{aligned}$$

where the new dimensionless expansion parameter is

$$\tilde{\epsilon} = \beta\epsilon\left(\frac{aB^2}{2A^3} + \frac{cB^2}{A^2} - \frac{bB}{2A^2} - \frac{3cbB}{2aA}\right). \tag{C.29}$$

Appendix D

Physical Justification of Mean Wall Stress

Formulas for determining the particle contribution to the shear stress of a colloidal suspension have been known for a long time, and received an especially careful treatment by G.K. Batchelor in the 1970's [4, 5]. The established literature mainly deals with the *mean* shear stress, either averaged over an infinite ensemble of systems or over an infinitely large system. The statistical uniformity of the system can then be invoked to argue that the mean stress over a typical 2-D slice through the system is equal to the mean stress averaged over the whole system volume. Although the wall is not a *typical* 2-D slice, because the boundary condition modifies the particle distribution, the fact that there is no mean net force on any part of the system when it is in steady state implies that the mean stress on all parallel 2-D slices must be the same. The average over an infinite system volume must therefore also be equal to the average over an infinite wall [4].

For the purpose of this paper, it is not enough to know the ensemble- or infinite-system-averaged mean. I need to look at the fluctuations about the mean in order to apply my procedure for empirically determining the mean renormalized work and the equilibrium free energy as a function of the shear stress. Therefore I need to go back through the derivation, and examine the *instantaneous* value of the shear stress at the wall in a suspension of a *finite* number of particles.

In this appendix, I prove that the instantaneous shear stress exerted by the fluid on the moving wall of the shear apparatus described in Chapter 3, averaged over the moving wall area, is

$$\sigma_{xy}^{\text{wall}} = \sigma_{xy}^{I} + \sigma_{xy}^{0} \tag{D.1}$$

where σ_{xy}^{I} is defined by

$$\sigma_{xy}^{I} \equiv \frac{1}{2V} \sum_{i \neq j} \hat{\mathbf{x}} \cdot \mathbf{F}_{ij} \Delta y_{ij}. \tag{D.2}$$

and σ_{xy}^{0} is independent of the particle positions.

I start by giving some necessary background on the behavior of shear stress in low-Re Newtonian fluids. To make this proof accessible to readers less familiar with hydrodynamics, I then map to a mathematically analogous problem in electrostatics (which turns out to be a homework problem from Griffiths' *Electricity and Magnetism* [31, problem 3.44a]). After presenting the solution to this electrostatics problem, I finally map back to hydrodynamics to obtain my final result.

D.1 Stress in Newtonian Fluids

The shear stress σ_{xy} is an off-diagonal component of the 3-by-3 stress tensor σ. σ is defined at each point in the fluid such that $\hat{\mathbf{n}} \cdot \sigma$ is the force per unit area exerted from below on a surface element at that location with unit normal vector $\hat{\mathbf{n}}$. By "from below," I mean from the side opposite to the direction of the normal vector. I will focus on the x column $\sigma \cdot \hat{\mathbf{x}}$ to obtain a vectorial quantity that will be easier to visualize.

By the definition of the stress tensor above, the x-component of the force on a region Ω of fluid is given by

$$F_x = -\int_{\partial\Omega} d\mathbf{A} \cdot \sigma \cdot \hat{\mathbf{x}} \tag{D.3}$$

$$= -\int_{\Omega} dV \nabla \cdot \sigma \cdot \hat{\mathbf{x}} \tag{D.4}$$

where $d\mathbf{A}$ is an infinitesimal area element of the boundary $\partial\Omega$ pointing along the outward normal direction, and dV is an infinitesimal volume element. I add the minus sign because I am computing the force on this surface from the outside. The second line results from the divergence theorem. Since this holds for every possible region Ω, I conclude that the integrand of equation (D.4) is equal to minus the x component of force per unit volume f_x exerted by the surrounding fluid on an infinitesimal volume element:

$$\nabla \cdot \sigma \cdot \hat{\mathbf{x}} = -f_x. \tag{D.5}$$

Finally, I must invoke the assumption that the solvent in which the particles are suspended is a Newtonian fluid, which implies

$$\sigma \cdot \hat{\mathbf{x}} = -\eta_0 \nabla u_x \tag{D.6}$$

where u_x is the x-component of the fluid velocity field, and η_0 is the (constant) viscosity of the solvent. Combining this with the previous equation gives us the set of equations

$$\eta_0 \nabla^2 u_x = f_x \tag{D.7}$$

$$\sigma \cdot \hat{\mathbf{x}} = -\eta_0 \nabla u_x \tag{D.8}$$

that together fully determine $\sigma \cdot \hat{\mathbf{x}}$ for a given set of boundary conditions.

D.2 Mapping to Electrostatics

Equations (D.7) and (D.8) suggest a mapping to electrostatics. $\eta_0 u_x$ is the analog to the electric potential ϕ, $\sigma \cdot \hat{\mathbf{x}}$ is the analog to the electric field $\mathbf{E}$, and $-f_x$ is the analogue to the charge density ρ. With these mappings, the mathematics of the problem are identical to electrostatics, and I can do everything in terms of $\mathbf{E}$, ϕ and ρ until I map back at the end.

The only remaining piece of setup is to map the boundary conditions and the "charge distribution." The non-slip boundary condition requires that every part of the fluid in contact with a non-rotating rigid surface must share the same velocity. Since the electric potential ϕ is the analog of the x-component of velocity, this implies that non-rotating surfaces behave like perfect conductors - they are always equipotentials. In particular, the constraint that the bottom wall is fixed and the top wall moves at constant velocity v implies that the walls of the shear cell become parallel conducting plates separated by a distance d, with fixed electric potential difference $\Delta\phi$. The problem of determining the total force on the walls is thus equivalent to determining the induced charge on these conducting plates.

The particles, however, are allowed to rotate. Their boundary conditions are therefore more complicated, involving the other columns of the stress tensor. Specifically, I have

$$\mathbf{u} = \Omega \times \mathbf{r}_\perp + \mathbf{u}_{\mathrm{cm}} \tag{D.9}$$

for all points on the surface of the sphere, where $\mathbf{r}_\perp$ is the vector pointing from the center of the sphere to the surface point, projected onto a plane perpendicular to the angular velocity vector Ω. Ω and the center-of-mass velocity $\mathbf{u}_{\mathrm{cm}}$ are free parameters that must be adjusted so as to be consistent with equations (D.7) and (D.8). The resulting restriction on σ is

$$\sigma = -\eta_0 \nabla \left(\Omega \times \mathbf{r}_\perp + \mathbf{u}_{\mathrm{cm}} \right). \tag{D.10}$$

To determine the charge distribution, I use my assumption of low Re to require the total force on any volume element to vanish. In the electrostatic analogy, this implies that the solvent is uncharged, and all charge must reside at the walls or on the particles. The interparticle repulsion exerts force on each particle that must be canceled by the friction of the fluid in order to satisfy the requirement of zero total force. This implies that the total "charge" on each particle must be $q_i = \sum_{j \neq i} \mathbf{F}_{ji} \cdot \hat{\mathbf{x}}$, where $\mathbf{F}_{ji}$ is the force exerted on particle i by particle j. The distribution of this total charge over the surface of each sphere is not fixed in advance, however, and must be determined by solving equations (D.7) and (D.8) (along with the corresponding equations for the other components of the stress tensor) with the boundary conditions just described. The decision to "ignore hydrodynamic interactions" mentioned in the main text allows us to greatly simplify the problem of determining these distributions, by solving the equations for each particle individually, with boundary condition $\mathbf{E} \to -(\Delta\phi/d)\hat{\mathbf{y}}$ far from the sphere. This approximation ignores the effect of the other particles and of the induced wall charge on the charge distribution over each sphere. The solutions obtained under this approximation are independent of the particle positions, which will be important later on.

D.3 Obtaining the Induced Charge on the Conducting Plate

My problem is thus reduced to determining the induced charge on a pair of conducting parallel plates at fixed electric potential due to a given charge distribution inside.

I start by splitting the charge on the plates into two parts, following the strategy of Batchelor in his treatment of the effect of particle interactions on mean shear stress [5]. The derivation will resemble Batchelor's in many ways, despite the electrostatic language, but adds a new element by considering the wall stress due to a given *instantaneous* configuration of particles as opposed to an ensemble average of all possible configurations.

The first part of the charge is the part required to maintain the electric potential difference $\Delta\phi$ in the absence of any additional charges between the plates: $Q_0 = A\Delta\phi/d$ on the top and $-Q_0$ on the bottom. To find the remaining charge, I can solve for the case where the two plates are grounded. When I add up the two charge distributions, the resulting field is guaranteed to produce the desired constant electric potential difference. The case of grounded plates is problem 3.44a in Griffiths, as mentioned above, and I will follow his method to solve it [31].

Griffiths starts by having the student derive a relation known as Green's Reciprocity Theorem. (This theorem is closely related to a result due to Lorentz in hydrodynamics, which Batchelor employs in his analysis [63].) Consider two distinct charge distributions $\rho_1(\mathbf{r})$ and $\rho_2(\mathbf{r})$, which produce electric fields $\mathbf{E}_1(\mathbf{r})$ and $\mathbf{E}_2(\mathbf{r})$, with electric potentials $\phi_1(\mathbf{r})$ and $\phi_2(\mathbf{r})$. Now use the Maxwell Equation $\nabla \cdot \mathbf{E} = \rho$ and the definition of electric potential $\mathbf{E} = -\nabla\phi$ to obtain

$$\begin{aligned}\int dV \mathbf{E}_1 \cdot \mathbf{E}_2 &= -\int dV \nabla\phi_1 \cdot \mathbf{E}_2 = \int dV \phi_1 \nabla \cdot \mathbf{E}_2 \\ &= \int dV \phi_1 \rho_2 \qquad \text{(D.11)} \\ &= -\int dV \nabla\phi_2 \cdot \mathbf{E}_1 = \int dV \phi_2 \nabla \cdot \mathbf{E}_1 \\ &= \int dV \phi_2 \rho_1 \qquad \text{(D.12)}\end{aligned}$$

where I are integrating over all space, and have used integration by parts to switch the ∇ from ϕ to $\mathbf{E}$.

I thus obtain Green's Reciprocity Theorem:

$$\int dV \phi_1 \rho_2 = \int dV \phi_2 \rho_1. \qquad \text{(D.13)}$$

Now I use this relation to compute the induced charge on my plates. I will start by computing the induced charge due to a point charge q at location $\mathbf{r} = (x, y, z)$. I will work in coordinates where the bottom plate is at $y = 0$ and the top is at $y = d$.

To apply the Reciprocity Theorem, I choose for ρ_1 the charge distribution we're interested in, with the point charge between the grounded parallel plates. I define Q_+ as the total induced charge on the top plate and Q_- as the total induced charge on the bottom plate. For ρ_2, I choose a charge distribution with conducting plates in the same locations, but with the top plate fixed at electric potential ϕ_0 above the bottom one, and with no charge in the space between them. The LHS of the Reciprocity Theorem vanishes, because $\phi_1 = 0$ whenever ρ_2 is nonzero. The RHS has a contribution from the charge distribution on the top plate, and a contribution from the particle. If the plates are infinite, then the potential a distance y above the bottom plate in scenario 2 is exactly $(y/d)\phi_0$. This will still be a good approximation in a finite system for charges that are not too close to the edges of the system, which will be true for the charges on the vast majority of the spheres when the number of spheres is large. Thus I obtain:

$$0 = \phi_0 Q_+ + \phi_0 \frac{y}{d} q. \tag{D.14}$$

Solving for Q_+, I find

$$Q_+ = -\frac{y}{d} q. \tag{D.15}$$

Now I again use the linearity of my equations to obtain the total induced charge by summing up the contributions from all the infinitesimal charge elements in the distribution. A convenient way to perform this sum is to split up the charge distribution on each sphere into two parts: a spatially uniform part equal to the mean surface charge on the sphere, and spatially varying part that integrates to zero over each sphere surface.

D.3.1 Contribution of Variations about the Mean

I start by computing the contribution of the second part of the charge distribution. Since this part of the charge sums to zero on each sphere, every positive charge δq has a corresponding negative charge $-\delta q$ somewhere else on the sphere. The net induced charge from each such pair is

$$\delta Q_+ = \frac{\delta q}{d}(y_- - y_+) \tag{D.16}$$

where y_- and y_+ are the coordinates of the $+\delta q$ and $-\delta q$ charges, respectively. Now recall that by ignoring hydrodynamic interactions, I can solve for the charge distribution over each sphere without knowing its position relative to the plates or the other particles. Furthermore, the linearity of the governing equations implies that the variations about the mean charge density are independent of the size of the mean. This implies that the y-distance $y_- - y_+$ between any pair of charges on a single sphere is independent of the spatial configuration of the particles and of the total charge q_i of the particle in question.

Summing over all pairs of charges from all the spheres in the sample, I define the quantity

$$Q_H = \sum \delta Q_+ \tag{D.17}$$

as the total induced charge due to the variations about the mean charge on the surface of the spheres. This quantity is independent of the particle positions, and just adds a constant offset to the total charge. The H subscript stands for "hydrodynamic," because this contribution comes purely from the friction of the flow field around each particle.

D.3.2 Contribution of the Mean Charge

To complete my calculation, I must compute the charge induced on the plate by a given configuration of uniformly charged spheres. Since the field of a uniformly charged sphere is equivalent to the field of a point charge (for points outside the surface of the sphere), I can simply evaluate the point charge solution derived above for every particle, and add them all up. I thus find

$$Q_I = -\sum_i \frac{y_i}{d} q_i. \tag{D.18}$$

Combining the above results, I find that the total induced charge on the top plate is $Q = Q_I + Q_0 + Q_H$, with Q_I the only term that depends on the particle positions.

D.4 Mapping Back to Hydrodynamics

I can now map back into the original variables (recalling that charge is equivalent to minus the force exerted by the fluid) in order to obtain the total force exerted by the fluid on the moving wall of the shear apparatus:

$$F_{\text{wall}} = \sum \frac{y_i}{d} \left(\sum_{j \neq i} \hat{\mathbf{x}} \cdot \mathbf{F}_{ji} \right) + F_0 + F_H. \tag{D.19}$$

I can simplify this expression by using the fact that $\mathbf{F}_{ji} = -\mathbf{F}_{ij}$:

$$F_{\text{wall}} = \frac{1}{2d} \sum_{i \neq j} \hat{\mathbf{x}} \cdot \mathbf{F}_{ij} \Delta y_{ij} + F_0 + F_H. \tag{D.20}$$

Finally, I can divide through by the area A of the wall to obtain the mean shear stress exerted on the wall by the fluid:

$$\sigma_{xy}^{\text{wall}} = \sigma_{xy}^I + \sigma_{xy}^0 + \sigma_{xy}^H \tag{D.21}$$

where

$$\sigma_{xy}^I = \frac{1}{2V} \sum_{i \neq j} \hat{\mathbf{x}} \cdot \mathbf{F}_{ij} \Delta y_{ij} \tag{D.22}$$

and the other two terms are independent of the particle positions. For notational simplicity, I combine them into one term in the main text, which I call σ_{xy}^0.