Title
Nonequilibrium Fluctuations and Information Processing in Mesoscopic Complex Systems

Author
Semaan, Mikhael T.

Publication Date
2022

Peer reviewed|Thesis/dissertation

Nonequilibrium Fluctuations and Information Processing
in Mesoscopic Complex Systems

By

MIKHAEL T. SEMAAN
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Physics

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

James P. Crutchfield, Chair

Daniel L. Cox

Xiaoli Dong

Committee in Charge

2022

For my Mother and Father:
Georgina, of Saïd and Violette
and
Toufic, of Jamil and Laurice.

You furnished my pack,
walked me to the forest's edge,
and bade me go.

Contents

Nonequilibrium Fluctuations and Information Processing
in Mesoscopic Complex Systems

Abstract

Natural systems are regularly "far from equilibrium:" transient, dynamical, and in constant interaction with their environments. These systems are often complex, comprising many mutually-interacting parts and giving rise to emergent behavior at observable scales—the human body, forest ecosystems, and the planet's climate are but a few pertinent examples. In cases like these, traditional techniques for analysis fall short: we cannot feasibly deduce the observed behavior either from individual, microscopic degrees of freedom (writing a system Hamiltonian, for example) nor from macroscopic statistical treatments as in equilibrium thermodynamics. How can we understand this class of systems? How do they absorb and dissipate energy; how do they store, process, and produce information; broadly, how do they *function*?

This dissertation provides a suite of mathematical and computational tools for probing these and related questions. Specifically: it extends ideas from stochastic thermodynamics to develop a so-called trajectory class fluctuation theorem for nonequilibrium complex systems, freed from many previously restrictive assumptions; derives a stochastic dynamical "first law" linking nonequilibrium behavior to information transduction; and applies these theoretical results to a collection of example systems: the ion channels of mammalian neural membranes and autonomous Maxwellian ratchets.

The contributions are organized as follows: Ch. 2 introduces key ideas from the languages of information theory, stochastic processes, and computational mechanics. Ch. 3 then continues by introducing *fluctuation theorems* and the core language of stochastic thermodynamics, before extending it with a nonequilibrium steady-state trajectory class fluctuation theorem and applying the theory to uncover surprising implications for the ion channels of mammalian neural membranes. Ch. 4 then combines thermodynamic and information-processing paradigms to derive generalized information processing first and second laws, applying them to the limits on how autonomous Maxwellian ratchets can leverage information to perform useful work. Finally, Ch. 5 closes and suggests rich ongoing and future lines of inquiry.

Acknowledgments

This dissertation is built upon a vast, nearly-unfathomable network of mentorship, support, and love, without which it would not be possible. I cannot hope to really do justice in acknowledging the people behind that network, but I will make an attempt here.

First, I must of course thank my PhD advisor, James P. Crutchfield. Jim: for your faith in my abilities even when I lacked it, for your relentless scientific and professional support, for being a living example of the spark of wonder's persistence through a scientific career, and for your—frankly—bottomless patience, thank you. I am proud to be a part of this intellectual tradition, and it is my greatest academic hope to carry it forward. I am inspired, every day, by the lenses through which I now view nature's wonders.

To my groupmates, cohort-mates, and fellow graduate students at UC Davis—intellectual guideposts, collaborators, colleagues, several of whom I've tricked into lasting friendships—great thanks is due. In particular:

Alex Jurgens, for your tremendous and unyielding support, friendship, guidance, and for dealing with infinite state machines so I could incessantly ask about them, thereby keeping us both sane ("sane?") during quarantine.

Sam Loomis, for your patience, empathy, professional guidance, and boundless mathematical assistance.

David Gier, for your quantum mechanical insight, friendship, support, and truly first-rate sarcasm.

Adam Rupe, for your thermodynamic and computational-mechanical expertise, willingness to share it, and academic and professional guidance.

Michael O. Flynn, for your unconditional kindness and encouragement, sharp physical insight, and for always being willing to chalkboard a physics problem.

Ryan James, for your early and constant mentorship, information-theoretic knowledgebase, and example-setting docstring practices.

Gregory Wimsatt, Alec Boyd, Paul Riechers, and Kyle Ray: for your stochastic-thermodynamic eyes, welcoming onboarding, lively and exciting research environment, constructive skepticism, and gracious engagement with my work.

Thanks also to the many additional friends, colleagues, mentors, and teachers at UC Davis who contributed to a lively, supportive, and inspiring academic environment: Rylai Melody and Molly Luginbuhl, Zaki Jawaid, Adam Kunesh, Ariadna Venegas-Li, Dany Masante, Korana Burke, Meredith Tromble, Nicolas Brodu, Fabio Anza, Xincheng Lei, Anastasiya Salova, Jeffrey Emenheiser, Sarah Marzen, Rose Baunach, Luke McClintock, Isaiah Santistevan, May Palace, Siddharth Vadnerkar, Jacob Cohen, Steven Macauda, Sonya Vyas, Lauren Downing, Steve Frediani, Gina Gilson, Xiaoli Dong, Dina Zhabinskaya, Nemanja Kaloper, Veronika Hubeny, Richard Scalettar, Rena Zieve, Susan P. Harrison, the entire GGE class taking 200A/B in 2018-19, the UCD TKD team, and so many more.

My decisions to pursue both research and physics—neither of which I understood people could really *do* at all prior—were borne of a series of fortunate events, friends, and mentors at California State University, Long Beach. Special thanks to my mentors:

Galen Pickett, for making my first college physics experience surprising and formative; for entertaining a young undergraduate who *really couldn't add the minor, I have to finish engineering in four years* in office hours well after his term in your class completed; for being a lasting and consistent guide and mentor through graduate admissions and beyond. Thank you.

Jiyeong Gu, for agreeing to take me into your lab despite no subject knowledge or clue whatsoever what experimental research entailed, for training and advising me through two years of undergraduate research, for your constant patience and kindness, and for teaching *the best* E&M. How many people get to say E&M was their *favorite* class?

Maryam Moussavi, for your vigilant faith in me, first-rate advising, kindness, and incredible teaching. Also, for sneakily—in ways I am still comprehending—planting the seeds of my fascination with chaos and nonlinearity. Your elective course changed my life.

To my peers and friends at Long Beach—Dany Atallah, Kayla Bollinger, Debbie Tonne, Diana González, Chris Kim, Jaylen Wimbish, Kevin Cano, Jason Jung, Rachel Yarbrough Jordan, Celeste Manughian-Peter, The Table, and many, many others—thank you, from the bottom of my heart, for your camaraderie, support, and for some of the best times of my life.

Also crucial to my decision to do research at all was the ample opportunity I had to pursue it as an undergraduate. To the Wireless Health Institute REU program at UCLA, and especially to

Wesley Uehara, Mani Srivastava and NESL, Prashanth Swaminathan, Matthew Newbill, and Dillon Stadther; and to the University of Toledo Physics REU program, especially Rick Irving, Jacques G. Amar, and my cohort-mates there; thank you.

A special thanks to the the teachers and mentors throughout my pre-collegate life, who recognized and nurtured my curiosity and character: Richere Barbeau, Eduardo Galaviz, Josette Bean, Tina Carlson, Roger Rios, Miguel Soto, Russ Cramm, Garr Nelson, Karen Williams, Cindy Ekk, Myra Awad, Kevin Dodge, David Minahan, and many more.

To the truly ridiculous Semaan and Saghbini clans: thank you; words cannot express how important the support of my family, and of the broader Lebanese culture which raised me, has been throughout my life's journey. It takes a village. Of course, I assume you all already know this, and so you'll forgive my not trying to enumerate all the Georges and Tonys. Love you all; I wouldn't be here without you. Thanks especially to my immediate family: Mom and Dad, I will never let you forget I'm a card-carrying *faylasouf* now. You did this to yourselves. Jamil, Gabriel, Catherine: you three continue to light my way. I love you.

To Brooke Gates, my wonderful partner: thank you for putting up with the erratic, indescribable ups and downs of a maniac pursuing a PhD. Your support through these years cannot be measured; neither can the joy your presence has made them. I love you, I am proud of you, and thank you for everything.

Common Acronyms

ESS	Equilibrium Steady State
NESS	Nonequilibrium Steady State
SUS	System Under Study
FT	Fluctuation Theorem
IFT	Integral Fluctuation Theorem
DFT	Detailed Fluctuation Theorem
TCFT	Trajctory Class Fluctuation Theorem
CTMC	Continuous-Time Markov Chain
ODE	Ordinary Differential Equation
IPSL	Information Processing Second Law
IPFL	Information Processing First Law
NEDS	Nonequilibrium Dynamical State
AS4C	Asymmetric Stochastic 4-Cycle
HMC	Hidden Markov Chain

CHAPTER 1

Introduction

Pattern and structure abound in nature. In many cases they are immediately apparent to human perception: a forest is recognizably a forest; a flock of birds recognizably a flock; a snowflake a snowflake; the cycle of day and night so deeply familiar that our inner rhythms evolved to its account. Yet not all of nature's patterns are so obvious: it took humans millenia to discover the quantitative pattern by which objects resist changes to their motion, centuries more for Poincaré to to realize that even the predetermined is generically beyond prediction [**1**], and discovering pattern in complex systems—broadly, those for which "more is different" [**2**]—remains an ongoing research program across scales and domains [**3**, **4**, **5**, **6**, **7**].

Importantly, *discovering* hidden pattern is distinct from *recognizing* known ones—the latter is the purview of supervised or partially supervised machine learning, a task at which its many methods are wildly successful. The former requires a concrete, operable notion of what pattern and structure *are*, independent of the choice of basis and without *a priori* knowledge to guide the search. For broad classes of complex systems—namely, those described by stationary and ergodic stochastic processes—a constructive solution is provided by *computational mechanics* [**4**]. The central approach of computational mechanics is to view complex systems as communication channels in the sense of Shannon information theory [**8**], which store, process, and produce information to function. Measures emerge which unambiguously characterize (i) a systems's *intrinsic randomness*, identified as the rate at which it actively produces information; and (ii) its *intrinsic memory*, given by the information stored in its hidden predictive states. Together, these quantities capture the disctinct but interrelated dynamics between randomness and organization.

Yet a complex system exists on a *physical* substrate: as it stores, processes, and produces information, it must also absorb, dissipate, and generate energy to function. As ubiquitous as emergent pattern in complex systems is their operation "far from equilibrium:" even apparently "steady" systems—the basal state of a human body, stable population structure in an ecological

community, an electronic circuit—rely on energetic gradients maintained by their myriad microscopic degrees of freedom to exhibit their stationary, apparently stable behavior.

Indeed, these two paradigms of viewing complex systems—as computers processing and communicting information, and as thermodynamic systems far from equilibrium absorbing and dissipating energy—are related. Were they not, the "neat-fingered being" of Maxwell's demon would provide a way to leverage information to extract useful work "for free," violating the second law of thermodynamics [**9**]. Nearly a century after Maxwell's thought experiment, Landauer [**10**] resolved the apparent paradox by bounding the amount of energy required to erase the information stored in a single bit (the binary unit of information theory). Succinctly, gaining knowledge about a system's internal configuration—even in the abstract—costs energy.

A central question emerges, inspired by the successes of Landauer and of the tools of computatoinal mechanics: What are the *detailed* fundamental trade-offs between energy and information processing in complex systems? How does a system's information processing determine its nonequilibrium thermodynamic behavior, and vice-versa? Are there unifying themes, or laws, as there are for mechanics and thermodynamics? If so, what character do they take?

This dissertation builds on and extends a growing body of work in this vein. Specifically, it does so by leveraging the philosophical and operational approaches of information theory, computational mechanics, and nonequilibrium thermodynamics towards describing complex systems which are

(1) not stationary, but actively driven by contact with an external environment;
(2) arbitrarily far from thermal equilibrium, even in steady conditions—exhibiting *nonequilibrium steady states*—and
(3) in explicit contact with thermodynamic and information-bearing degrees of freedom, facilitating an exchange between mechanical, thermal, and information-bearing physical resources.

It does so in several steps.

First, Ch. 2 gives the necessary technical background for concepts of information theory and computational mechanics on stationary stochastic processes. In particular, it begins by defining a stationary and ergodic stochastic process and several key measures of information theory, including Shannon entropy and its rate, mutual information, and Kullback-Liebler divergence between two

distributions. Finally, it defines the key object of computational mechanics—the ϵ-machine, a stationary and ergodic process's minimally-sized, maximally-predictive representation (equivalently, its minimal sufficient statistic)—and demonstrates how this representation enables principled calculation of a system's intrinsic randomness and intrinsic structure.

Ch. 3 then begins by shifting away from stationary processes and instead considering systems which actively interact with a dynamic environment, towards the goal of characterizing the nonequilbrium thermodynamics of mesoscopic complex systems—in particular, systems which exhibit *nonequilibrium steady states*. *Stochastic thermodynamics*, concerned with binding nonequilibrium statistical fluctuations to equilibrium or steady-state observables, here enters the stage as a natural way to describe the fluctuating dynamics of non-stationary stochastic processes, even absent a well-defined thermal environment.

Sec. 3.2 introduces the core stochastic dynamical functionals of such a driven, nonequilibrium process, used extensively throughout this work. Then Sec. 3.3.1 first reviews the *fluctuation theorems* of stochastic thermodynamics, before culminating in this work's first primary theoretical result: a *trajectory class fluctuation theorem* for generically nonequilibrium steady-state complex systems.

Secs. 3.4 and 3.5 then interpret and apply this new theoretical toolkit to two biological examples: the voltage-gated sodium and potassium ion channels of mammalian neural membranes. These channels, instrumental for generating and propagating neuronal action potentials, are shown to exhibit novel biophysical functionality when driven by a realistic, biophysically-plausible action potential signal, including an extra dimension of thermodynamic second law violation accessible only to nonequilibrium steady-state systems which must be accounted for to avoid apparent violations in preceding nonequilibrium theories.

Ch. 4 then considers the *informational* dynamics of such a driven, nonequilibrium steady-state complex system, introducing as an example class the autonomous Maxwellian ratchet in Sec. 4.2.4: Sec. 4.3 derives an "information processing first law" for transitions between nonequilibrium dynamical states, which describes thermodynamic changes to a system's information content in exactly the same way the thermodynamic first law does for changes to its energy. Sec. 4.4 continues the development by demonsrating how the combination of the information processing first law and

fluctuation theorems of stochastic thermodynamics tighten and generalize two recent "information processing second laws" for autonomous Maxwellian ratchets.

By way of example, Sec. 4.5 demonstrates the new theory on a tunably nonequilibrium steady-state Maxwellian ratchet, dubbed the "asymmetric stochastic 4-cycle," where we find quantitatively how this system must siphon part of its potential for work extraction towards the housekeeping entropic costs of maintaining nonequlibrium steady states.

Taken together, this dissertation—through combining and extending stochastic thermodynamics and information theory with techniques from computational mechanics—builds and demonstrates concrete theoretical and computational results towards fully characterizing the information-bearing thermodynamics of mesoscopic, highly nonequilibrium complex systems.

CHAPTER 2

Background

While Chs. 3 and 4 develop all the necessary thermodynamic and model class background and are therefore self-contained, they both make repeated use of and reference to measures of Shannon information theory (at the level of Ref. [**11**]) and of the theory of computational mechanics, an established but active and growing body of research literature [**4**, **12**, **13**, **14**, **15**, **16**, **17**, **18**, **19**, **19**, **20**, **21**, **22**, **23**]. To this end, we review the requisite notions of both here, to contextualize their appearance in Chs. 3 and 4. As such, this chapter makes no attempts at either mathematical rigor or comprehensiveness of review, instead focusing on those pieces of each body of theoretical background which make appearances in subsequent chapters.

2.1. Probability and Information Theory Primer

Consider a system which can stochastically realize a state x from a set $\mathcal{X}$ of possible values, with the probability distribution over states given by random variable X[1]. The probability that upon observation we find the system in state $x \in \mathcal{X}$ is denoted

$$\Pr(X = x) . \tag{2.1}$$

Throughout this text, a countable space $\mathcal{X}$ is assumed unless otherwise specified.

Since probabilities take values between 0 and 1, without loss of generality define a new quantity called the *surprisal* s:

$$s(X = x) \doteq -\log \Pr(X = x) , \tag{2.2}$$

with the convention that $0 \log 0 = 0$. The surprisal is so called because the *less* likely the outcome of an observation—colloquially, the more surprising its result—the larger s's value. Its standard

[1]The notation used throughout denotes individual realizations with lowercase, random variables with uppercase, and sample spaces with calligraphic styling.

units vary by base of the logarithm: in information theory often it is taken to base 2, in which case surprisal is measured in *bits*. With natural logarithms, surprisal takes units of *nats*.

Surprisal is also called *self-information*: more surprising outcomes communicate more information, while a certain outcome communicates no information—this results in its lower bound at 0 where the probability of a realization is 1. Unlike probability, however, surprisal is not upper-bounded: as the probability of an event approaches 0, the surprisal at observing it approaches $+\infty$.

This small change in perspective from probability to surprisal forms the basis for converting from probability to information theories. In point of fact, the *Shannon entropy* or *Shannon information*—a fundamental quantity of information theory—is itself simply the average surprisal over all possible outcomes: for countable spaces $\mathcal{X}$,

$$H[X] \doteq \sum_{x \in \mathcal{X}} \Pr(X = x)\, s(X = x). \tag{2.3}$$

The Shannon entropy of a random variable X thus quantifies the average surprisal over observations of X.

2.1.1. Bivariate Information Theory.

For two independent random variables X and Y, Shannon information is additive[2]:

$$H[X, Y] = H[X] + H[Y] \quad \text{iff} \quad X \text{ and } Y \text{ are independent}, \tag{2.4}$$

where X, Y denotes the joint random variable formed by X and Y, and so

$$\begin{aligned} H[X, Y] &\doteq \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} \Pr(X = x \cap Y = y)\, s(X = x \cap Y = y) \\ &= -\sum_{x \in \mathcal{X}, y \in \mathcal{Y}} \Pr(X = x \cap Y = y) \log \Pr(X = x \cap Y = y). \end{aligned} \tag{2.5}$$

When X and Y are not independent, the *conditional Shannon entropy* is defined:

$$H[Y \mid X] \doteq H[X, Y] - H[X], \tag{2.6}$$

[2]Indeed this was originally an axiom for its construction [**8**], though phrased differently.

the average uncertainty remaining in Y after conditioning on X.

Information theory's fundamental unit of pairwise correlation is the *mutual information*

$$I[X:Y] \doteq H[X,Y] - H[Y \,|\, X] - H[X \,|\, Y], \tag{2.7}$$

which is symmetric and (like Shannon entropy) strictly nonnegative. It is the average uncertainty remaining in the joint distribution of two random variables after accounting for each's conditioning on the other.

Thus far we have measures of information for (i) an individual random variable, (ii) a random variable conditoined on another, (iii) the joint distribution of two random variables, and finally (iv) mutual information between two random variables. It remains to define a "distance" between two distributions. There are many choices here [**24**,**25**,**26**], but we will introduce only the *Kullback-Leibler divergence* between random variables X_1 and X_2 defined on the same sample space $\mathcal{X}$:

$$D_{\mathrm{KL}}[X_1 \,\|\, X_2] \doteq \sum_{x \in \mathcal{X}} \Pr(X_1 = x) \log \frac{\Pr(X_1 = x)}{\Pr(X_2 = x)}. \tag{2.8}$$

Unlike mutual information, this divergence is antisymmetric: $D_{\mathrm{KL}}[X_1 \,\|\, X_2] = -D_{\mathrm{KL}}[X_2 \,\|\, X_1]$. Neither does it satisfy the triangle inequality: $D_{\mathrm{KL}}[X_1 \,\|\, X_3] \nleq D_{\mathrm{KL}}[X_1 \,\|\, X_2] + D_{\mathrm{KL}}[X_2 \,\|\, X_3]$. For these reasons it is not a metric in the sense of metric space geometry, and so is not a strict "distance" on the space of probability distributions. Still, it sees extensive use measuring the difference between one probability distribution and another with the same support, and appears particularly often in stochastic thermodynamics as discussed in Chs. 3 and 4.

2.2. Discrete-Time Stochastic Process

By *discrete-time stochastic process* (shortened in this dissertation to simply "process") is meant a joint random variable

$$X_{m:n} \doteq X_m X_{m+1} \ldots X_{n-1} X_n, \tag{2.9}$$

composed of a sequence of individual random variables indexed by time (m, n we customarily take to be integers). A specific realization of the process—the outcome of a time-series measurement

of the system—is also called a *trajectory* $x_{m:n}$, and the state space $\mathcal{X}$ is called the *alphabet* of the process.

The process itself is determined completely by the probability distribution over the entire *joint* random variable—the distribution over all possible sequences of measurements. Thus we neither assume nor require that the individual random variables are independent or identically-distributed—indeed, we make no assumption about the individual random variables at all except that they are drawn from the same alphabet.

2.2.1. Stationary and Ergodic Bi-Infinite Processes.

In stochastic thermodynamics finite-length processes as defined previously are extensively used. However, it is also of interest (and convenience) to consider the limit of a bi-infinite stochastic process $X_{-\infty:\infty} \doteq X_:$ (infinite indices are suppressed), which allows us to define two additional restrictions which are crucial to computational mechanics.

The first of these is stationarity: a stochastic process is *stationary* if its statistics to all orders are unchanged by a shift in time:

$$\begin{aligned} &X_: \text{ stationary iff } \forall t \in \mathbb{Z}, L \in \mathbb{Z}^+, x_{0:L} \in \mathcal{X}^{N+1}, \\ &\quad \Pr\left(X_{0:L} = x_{0:L}\right) = \Pr\left(X_{t:t+L} = x_{0:L}\right). \end{aligned} \tag{2.10}$$

Stationarity allows us to identify a process with its *word probabilities* $\{\Pr\left(X_{0:N} = x_{0:N}\right)\}$ for all word lengths.

The second of these is ergodicity: in this setting, we call a process *ergodic* if, given a particular realization $x_{0:M}$, the empirical estimate for any length-L word probability converges almost surely to the true word probability in the $M \to \infty$ limit. Put simply, a single infinitely long observation is sufficient to determine the "true" statistics of the underlying process.

2.2.2. Information in Stochastic Processes.

As a stochastic process is, at root, a joint random variable, the techniques and measures of information theory are readily applied to it [**12**, **27**]. In particular, for a stationary process, we can consider the trivariate splitting of the "past" $X_{:-1}$, the "present" X_0, and the "future" X_1 :. This

allows us to define the *Shannon entropy rate* h_μ for a stationary process,

$$h_\mu \doteq H[X_0 \,|\, X_{:-1}] \tag{2.11}$$

$$= \lim_{N\to\infty} \frac{H[X_{0:N-1}]}{N}, \tag{2.12}$$

the asymptotic rate at which new measurements produce *new* information—not accounted for by conditioning on the past. This is identified with the *intrinsic randomness* of a process: that randomness still present even after conditioning on infinitely-long past obsevations, or equivalently after considering temporal correlations of all lengths [**4**].

Of common interest is the special case when a process's future depends only on finite-length observations of the past—when considering only finite-order temporal correlation is sufficient. We say a process has *Markov order* n, if n is the smallest positive integer such that for any semi-infinite trajectory $x_{:1}$

$$\Pr\left(X_1 = x_1 \mid X_{:0} = x_{:0}\right) = \Pr\left(X_1 = x_1 \mid X_{-n:0} = x_{-n:0}\right). \tag{2.13}$$

Typically a process is called simply a "Markov process" (or, equivantly referring to a representation of such a process, a "Markov chain") if the Markov order is 1.

2.3. Computational Mechanics

At the core of computational mechanics, whose aim it is to study the intrinsic structure of systems described by stationary and ergodic stochastic processes, is the *predictive equivalence relation* $\sim_\epsilon$, which groups together infinite pasts $x_{:0}$ which make the same predictions over infinite futures $x_{1:}$:

$$x_{:0} \sim_\epsilon x'_{:0} \iff \forall x_{1:},\, \Pr\left(X_{1:} = x_{1:} \mid x_{:0}\right) = \Pr\left(X_{1:} = x_{1:} \mid x'_{:0}\right). \tag{2.14}$$

Equivalently, we may consider the ϵ function which maps individual pasts their equivalence class. The equivalence classes generated by $\sim_\epsilon$ are called *causal states* or *predictive states*. The set of causal states together with the induced state-to-state dynamic between them form computational mechanics' ϵ-machine: the unique, minimally-sized, maximally-predictive representation of the underlying stochastic process.

When it can be constructed, the ϵ-machine provides for direct calculation of a process's entropy rate—a process' intrinsic randomness—and additionally provides for defining a process's *intrinsic structure* as the *statistical complexity* C_μ. Let $\mathcal{S}$ be the set of causal states of an ϵ-machine, and S be the random variable supported by $\mathcal{S}$ and distributed according to the asymptotic dynamics of the machine (that is: the stationary distribution over its internal states). Then

$$C_\mu = H[S], \tag{2.15}$$

quite literally the information stored in the causal states. While statistical complexity can be defined for any effective state-space model of a process with an associated transition dynamic, the guaranteed uniqueness and minimality of the ϵ-machine representation makes this a canonical measure for intrinsic structure, independent of the process' representation.

These notions of information and structure in hand, we now turn to the nonequilibrium thermodynamics of complex systems.

CHAPTER 3

Homeostatic and Adaptive Energetics: Nonequilibrium Fluctuations Beyond Detailed Balance in Voltage-Gated Ion Channels

Much of the content in this chapter appears in press as Ref. [**28**]: M. T. Semaan and J. P. Crutchfield, *Phys. Rev. E* **106**, 044410.

Stochastic thermodynamics has largely succeeded in characterizing both equilibrium and far-from-equilibrium phenomena. Yet many opportunities remain for application to mesoscopic complex systems—especially biological ones—whose effective dynamics often violate detailed balance and whose microscopic degrees of freedom are often unknown or intractable. After reviewing *excess* and *housekeeping* energetics—the adaptive and homeostatic components of a system's dissipation—we extend stochastic thermodynamics with a trajectory class fluctuation theorem for nonequilibrium steady-state, nondetailed-balanced complex systems. We then take up the neurobiological examples of voltage-gated sodium and potassium ion channels to apply and illustrate the theory, elucidating their nonequilibrium behavior under a biophysically plausible action potential drive. These results uncover challenges for future experiments and highlight the progress possible understanding the thermodynamics of complex systems—*without* exhaustive knowledge of every underlying degree of freedom.

3.1. Introduction

Nonequilibrium phenomena pervade nature: In their many forms, energy gradients send hurricanes and wildfires to ravage, volcanoes to form and erupt, life to emerge. Mesoscopic complex systems—a planetary climate, forest ecosystems, the human body—consist of microscopic degrees of freedom that are inaccessible, intractable, or simply unknown. In point of fact, the human body's biochemistry relies essentially on out-of-equilibrium dynamics to function, adapt, and maintain

homeostasis; its myriad degrees of freedom are only ever partially accessible. Similarly, mesoscopic and complex systems provide fertile grounds for honing and applying tools to analyze real-world nonequilibrium processes.

Describing energetic fluxes in complex systems—developing a suitable mesoscopic nonequilibrium thermodynamics—remains an ongoing challenge: mathematics and physics difficulties continue to hinder deeper understanding of how these systems operate and function. The following leverages and extends tools from stochastic thermodynamics and information theory to address these challenges. To demonstrate the techniques, it takes up two suitably complex, mesoscopic neurobiological systems: voltage-gated ion channels.

3.1.1. Nonequilibrium steady states.

A system is typically called nonequilibrium in two distinct senses. The first, and most common, refers to nonequilibrium *processes*—say, induced by rapid environmental driving—wherein a system evolves through a series of transient configurations. When the environmental drive remains fixed, such a system remains out of equilibrium as it *relaxes* to some stationary distribution over its states, determined by the environmental parameters. If that stationary distribution corresponds to a thermodynamic equilibrium, we say the system possesses an *equilibrium steady state* (ESS), irrespective of its (perhaps highly nonequilibrium) driven, transient dynamics.

The second sense refers not to the transient behavior but to the nature of the stationary distributions: a *nonequilbrium steady-state* (NESS) system is one whose steady states are themselves out of thermodynamic equilibrium. This is simply achieved by contact with two heat baths at different temperatures. Rayleigh-Bénard convection [**29**] exemplifies this phenomenon: the temperature gradient between the top and bottom boundaries ensures a constant flux of energy through the fluid, from the hotter to the cooler, even when the gradient remains fixed indefinitely. In this case it is not enough to identify the energetic fluxes due to the system's transient dynamics; we must also identify the energy required to maintain steady-state conditions in the first place.

NESSs appear even without multiple heat baths. For example, by optically dragging a bead through viscous fluid [**30**]—an experimental realization of nonconservative force-driven Langevin

dynamics—by coarse-graining microstates [**31**]; or by contact with reservoirs of distinct electrochemical potentials—the case in virtually all common electrical circuits via Joule heating [**32**]. They emerge as well in the voltage-gated ion channels we consider.

A first attempt to give NESS systems a full thermodynamic framing defined the *housekeeping* heat $Q_{\rm hk}$ as the portion of the total heat Q that maintains NESS conditions [**33**]. (In this, the total heat is that energy exchanged between a system and its thermal environment, often idealized as a fixed-temperature bath.) What remains is the energy exchanged owing to the system's relaxation to steady state, termed the *excess* heat $Q_{\rm ex}$:

$$Q = Q_{\rm ex} + Q_{\rm hk}. \tag{3.1}$$

Contrast this with an equilibrium system's steady states, which by definition exchange no net energy with the thermal environment. In this setting, $Q_{\rm hk} = 0$ and so *all* dissipated heat is excess: $Q_{\rm ex} \to Q$. In other words, $Q_{\rm ex}$ in the NESS setting carries the same meaning as *total* heat Q in the ESS setting, and vice versa.

3.1.2. Approach.

Equilibrium thermodynamics and equilibrium statistical mechanics prove insufficient to analyze nonequilibrium processes [1]. That said, recent advances in stochastic thermodynamics now successfully describe fluctuations in a variety of far-from-equilibrium systems. This has been done both in the first sense (relaxation to ESSs) [**36**, **37**, **38**, **39**] and in the second (NESSs) [**40**, **41**, **42**]. Ref. [**43**] gives a recent review.

The following applies and extends these advances to analyze two complex neurobiological systems: voltage-gated sodium and potassium ion channels [**44**]—biophysical systems that originally motivated introducing master equations for NESSs [**45**]. This elucidates, for the first time, their nonequilibrium behavior under the realistic, dynamic environmental drive of an action potential spike. In doing so, a toolkit emerges whose validity extends to a host of other mesoscopic complex systems—even those for which a purely energetic interpretation is impossible or problematic—provided a relatively small set of constraints on their effective dynamics.

[1] With important exceptions, particularly in the first sense, under small perturbations from equilibrium [**34**, **35**].

Our development unfolds as follows. First, Sec. 3.2 lays out the relevant notation for our model classes and introduces appropriate *excess* thermodynamic functionals for describing them, ending with Sec. 3.2.3 which elaborates on the relationship between housekeeping heat, (ir)reversibility, and detailed balance. Sec. 3.3 reviews *fluctuation theorems*, which bind nonequilibrium thermodynamic fluctuations to steady-state quantities. It closes in Sec. 3.3.3 with our primary theoretical result: the first full trajectory class fluctuation theorem valid for NESS systems.

Moving to applications, Sec. 3.4 introduces our example neurobiological systems: voltage-gated sodium and potassium ion channels embedded in neural membranes. Sec. 3.5 then applies the techniques developed in the preceding theory to the channels, illustrating and comparing their responses under realistic action potential spikes.

These results serve three roles. First, they show how the trajectory class fluctuation theorem evades the divergences implied by real-world systems with one-way only transitions. Second, they quantitatively demonstrate how failing to account for housekeeping dissipation violates related fluctuation theorems, suggesting an important direction for experimental effort. Finally, despite marked differences between the ion channels' steady-states, the results show how to directly compare the channels' *excess* energetics. This both circumvents implied housekeeping divergences and allows for meaningful comparisons between their *adaptive* responses to the same environmental stimulus.

3.2. Preliminaries

The central object here is the finite-length *controlled stochastic process* $X_{0:N} \doteq X_0 X_1 X_2 \ldots X_N$, where $X_i \in \mathcal{X}$ is the random variable corresponding to the state of a system under study (SUS) at times $\{t_i \in \mathbb{R} : i = 0, \ldots, N\}$. We call a specific realization $x_{0:N}$ a *trajectory*. The process' dynamics are not stationary; rather, they are driven by a *protocol* $\lambda_{0:N}$. Fig. 3.1 illustrates the scheme.

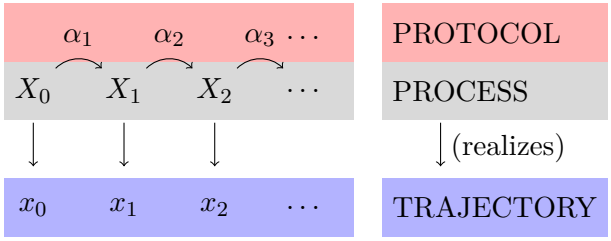

FIGURE 3.1. Interaction between the stochastic process $X_{0:N}$, protocol $\lambda_{0:N}$, and realized (observed) trajectories $x_{0:N}$.

We place the following constraints on the SUS.

(1) Each system parameter λ_i leads to a stationary and ergodic stochastic process, realized by holding the protocol fixed indefinitely at λ_i. This implies a unique would-be steady-state distribution $\boldsymbol{\pi}_{\lambda_i}$ associated to each λ_i.

(2) The state and protocol spaces are of *even parity*, in the sense that we do not negate their values under time reversal, defined precisely later. Sec. 3.2.3 discusses removing this assumption.

We emphasize these are *all* that is required for the main theoretical result and for meaningful definitions of the excess and housekeeping functionals. Importantly, we do not require dynamics of any particular form or possessing any particular structure—Markovian, Langevin, detailed-balanced, Hamiltonian, master equation, coupled to ideal baths, and so on—beyond that specified by the two conditions above. We do not require states to be *microscopic*; they can correspond to arbitrary or unknown coarse grainings. With this in mind, even the discrete succession of events is flexible. In particular, from any continuous-time dynamic we may generate a corresponding discrete-time one for appropriately small time steps.

While one cannot, in this most general setting, determine *energetics*, the fluctuation theorems introduced hold independently and exactly—and at any level of system description. We state the fluctuation theorems in this setting for two primary reasons: first, for clarity of derivation; second, with an eye toward future applications beyond thermodynamic systems to generally nonstationary stochastic processes.

3.2.1. The thermodynamic system.

That said, generality can hinder ease of application. To this end, when presenting the theoretical tools we frequently return to the relevant example "thermal system" of Fig. 3.2. This is a SUS coupled to an ideal heat bath at inverse temperature $\beta = 1/k_{\mathrm{B}}T$, an ideal work reservoir parameterized by λ, and an *auxiliary reservoir* representing the otherwise unaccounted-for degrees of freedom. Furthermore, we assign to each SUS state x an energy $E_\lambda(x)$. Finally, while the example system does not assume (order-1) Markov dynamics, it *does* assume no dynamical dependency on times before t_0. That is, the system's initial preparation is sufficient to determine the stochastic dynamics during the protocol.

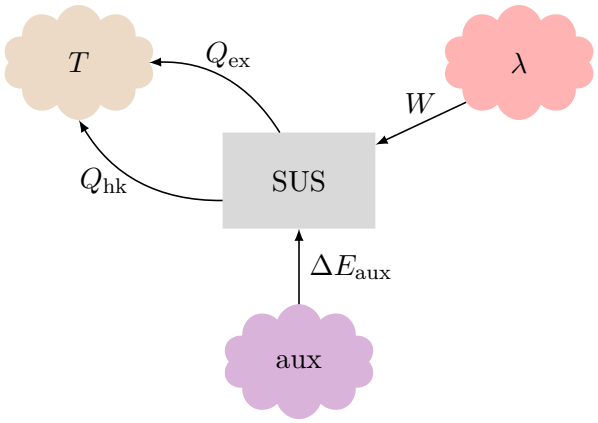

FIGURE 3.2. A *thermal system* and its interactions with various baths. While heat and work reservoirs (labeled with temperature T and parameter λ, respectively) are ideal, by design we assume nothing about the auxiliary or "aux" reservoir, and so label its energetic contribution by ΔE_{aux} to avoid confusion with well-defined terms like heat and work. The labels and arrow directions indicate energetic fluxes to-from the system. Notably, we allow for nonequilibrium steady states and functionally split the total heat Q into the *excess* heat Q_{ex}—corresponding to *adaptive* dissipation—and *housekeeping* heat Q_{hk}—referring to *homeostatic* dissipation.

These additional restrictions allow identifying the *energies* associated to each dynamical functional—introduced shortly. While these constraints are minimal, they still allow the SUS to be a coarse-grained representation. In general, this implies that the energetic fluxes are bounds rather than strict equalities [**31**].

Of paramount importance—and missing from most idealized thermal schemes—is the presence of the auxiliary bath. The heat and work reservoirs are each proxies for distinct *kinds* of coarse-grained degrees of freedom with distinct internal structures: the heat reservoir is an infinitely large source of purely thermal energy; the work reservoir is an entropyless source of energy, whose role is to set the SUS's energetic landscape via parameters λ.

In contrast, there are no restrictions on the auxiliary reservoir's structure. It is unnecessary when describing the SUS's effective state (or energy) at any particular time. Partly, the auxiliary reservoir stands in for coarse-graining out unknown degrees of freedom by unknown schemes. In the case of Rayleigh-Bénard convection, the auxiliary reservoir is a heat bath at a different temperature T_2. To take another example, in an information ratchet scheme, the auxiliary reservoir may represent the information tape interacting with the ratchet system [**46**, **47**, **48**].

Yet the relevance of the auxiliary bath goes beyond this: specifically, Sec. 3.2.3 shows that it is a *necessary* source for maintaining NESS conditions in this idealized picture.

Altogether, Fig. 3.2 captures a large class of mesoscopic physical, chemical, biological, and engineered systems that exhibit nonequilibrium steady states, but that have additional structure and are in contact with at least one thermal environment. We note an important distinction: the "heats" to which we refer in the thermal system context are always associated to the heat bath defined in Fig. 3.2, and "works" associated to λ. We avoid calling any fluxes between the auxiliary bath and the system heat or work, since we place no *a priori* restriction on its structure.

Subsequent sections develop tools for calculating the associated heats and works and for bounding their nonequilibrium fluctuations. Practically, this suggests experimental calorimetry and introduces a valuable way to *calibrate* effective models—such as, for example, those of the ion channels we take up later.

3.2.2. Excess energetics.

For an ESS system in contact with a single heat bath, the familiar First Law defines work W and heat Q as distinct contributions to its energy change over the course of a protocol [**37**]:

$$\begin{aligned} \Delta E &= \underbrace{\int \partial_\lambda E_\lambda(x)\, \mathrm{d}\lambda}_{W} + \underbrace{\int \partial_x E_\lambda(x)\, \mathrm{d}x}_{-Q}. \\ &\doteq \end{aligned} \tag{3.2}$$

That is, W denotes a difference in system energy owing to a change in protocol—a change in the overall energy *landscape*—and Q denotes the difference owing to a change in system state—a dissipative signature of its *adaptation* to environmental conditions.

In contrast, the general setting may not provide a meaningful notion of energy. Worse, even in the restricted case of Fig. 3.2, we can no longer define total heat for a NESS system by Eq. (3.2): it leads to contradiction.

To see this, consider a fixed protocol at λ and a system poised already in the distribution $\boldsymbol{\pi}_\lambda$. By definition then $W = 0$ and so $\langle W \rangle = 0$, where $\langle \cdot \rangle$ denotes a weighted average over all possible

paths. We also have $\langle \Delta E \rangle = \Delta \langle \pi_\lambda | E_\lambda \rangle = 0$, where:

$$\langle \pi_\lambda | E_\lambda \rangle \doteq \int_{x \in \mathcal{X}} \pi_\lambda(x) E_\lambda(x). \tag{3.3}$$

Yet we cannot have $\langle Q \rangle = 0$, for by Eq. (3.1):

$$\begin{aligned} \langle Q \rangle &= \langle Q_{\mathrm{hk}} \rangle \\ &\neq 0 \ , \end{aligned} \tag{3.4}$$

since this is a NESS system. In other words, the observed housekeeping flux, leaving no signature in the state energies, must come from somewhere *outside* of the system coupled to single ideal heat and work reservoirs. This is precisely what the auxiliary bath provides: in this case $\Delta E_{\mathrm{aux}} = Q_{\mathrm{hk}}$.

However, there is an alternative to energy. Since to every parameter λ is an associated steady-state distribution $\boldsymbol{\pi}_\lambda$, we can define the *steady-state surprisal*:

$$\phi_\lambda(x) \doteq -\ln \pi_\lambda(x) \ ; \tag{3.5}$$

so called since it is the Shannon self-information [**11**] of observing state x under the distribution $\boldsymbol{\pi}_\lambda$.

Taking the surprisal's state average under this distribution yields its Shannon entropy (or, for continuous-state spaces, its differential entropy):

$$\begin{aligned} \langle \pi_\lambda | \phi_\lambda \rangle &\doteq \int_{x \in \mathcal{X}} \pi_\lambda(x) \phi_\lambda(x) \\ &= H[\boldsymbol{\pi}_\lambda] \ . \end{aligned} \tag{3.6}$$

To see how the surprisal relates to energy, consider the canonical ensemble of statistical mechanics—the ESS version of Fig. 3.2, with $Q_{\mathrm{hk}} = 0$ and so no auxiliary bath—where $\boldsymbol{\pi}_\lambda$ is the Boltzmann distribution [**49**]. Then:

$$\begin{aligned} -\ln \pi_\lambda(x) &= \beta(E_\lambda(x) - F_\lambda^{\mathrm{eq}}) \\ &= \phi_\lambda(x) \ , \end{aligned} \tag{3.7}$$

where F_λ^{eq} is the equilibrium free energy (the familiar logarithm of the canonical partition function).

Eq. (3.7) motivates yet another moniker for $\phi_\lambda(x)$: the *nonequilibrium potential*. In this sense, steady-state surprisal is analogous to a generalized energy. However, it remains a meaningful characterization of a system's steady-state distribution—via its information-theoretic interpretation—even when energy is not meaningful.

Leveraging this, an analogous First Law for $\phi_\lambda(x)$ defines the *excess* heat and work:

$$\begin{aligned}\Delta\phi &= \underbrace{\int \partial_\lambda \phi_\lambda(x)\, \mathrm{d}\lambda}_{} + \underbrace{\int \partial_x \phi_\lambda(x)\, \mathrm{d}x}_{} \\ &\doteq \qquad \mathcal{W}_{\text{ex}} \qquad\qquad - \mathcal{Q}_{\text{ex}}.\end{aligned} \tag{3.8}$$

As with their nonexcess counterparts, these quantities characterize distinct *dynamical* contributions to a change in steady-state surprisal: $\mathcal{W}_{\text{ex}}$ capturing that due to a changing protocol, which sets the steady-state probability landscape; $\mathcal{Q}_{\text{ex}}$ monitoring a system's *adaptation* to its environment.

For Fig. 3.2's thermal system, these conveniently convert to energies: $\mathcal{W}_{\text{ex}} \to \beta W_{\text{ex}}$ and $\mathcal{Q}_{\text{ex}} \to \beta Q_{\text{ex}}$. And, they agree with other standard formulations of excess thermodynamic functionals [**40**, **42**]. Using Eq. (3.7) and taking the ESS limit of Boltzmann-distributed steady-states yields: (i) $\mathcal{Q}_{\text{ex}} \to \beta Q$—with equilibrium steady states, all dissipated heat is excess—and (ii) $\mathcal{W}_{\text{ex}} \to \beta(W - \Delta F)$, leading to its classification as an *excess environmental entropy production* [**42**].

We stress, though, that the excess work and heat—and the steady-state surprisal—retain dynamical meaning independent of Boltzmann or even energetic assumptions. In this way, Eq. (3.7) is a guidepost for thermodynamic interpretation. It is not, however, a strict equivalence. In point of fact, as we will see, sodium channels (as with other NESS systems) lack a well-defined steady-state free energy [**42**]. Nevertheless, Eq. (3.8) describes—tractably—two functionally distinct aspects of their response to dynamic environments.

3.2.3. Detailed balance and housekeeping.

The housekeeping heat remains. Recall that it corresponds to energy dissipated to maintain NESSs, as in Eq. (3.1). Phenomenologically, Eq. (3.1) provided a satisfactory answer. However, our excess heat definition only required would-be steady-state distributions exist. The definition of total heat, in contrast, depended explicitly on well-defined state energies. This difference led to problems with NESSs.

The upshot is that a more general definition of housekeeping heat is called for. In particular, it should depend only on the stochastic dynamics and, when added to excess heat, it should give a reasonable generalization of *total* heat. Naturally, we also require interpretability and that it reduces to the corresponding well-understood thermodynamic terms in the appropriate limits.

To these ends, but in a slightly more general form than previously reported, we define *housekeeping heat* to explicitly allow non-Markovian dynamics:

$$\mathcal{Q}_{\text{hk}} \doteq \ln \frac{\Pr\left(X_{1:N} = x_{1:N} \quad \middle| X_0 = x_0 \;\; ; \lambda_{0:N}\right)}{\Pr\left(X_{1:N} = x_{N-1:0} \,\middle|\, X_0 = x_N \;\; ; \lambda_{N:0}\right)} + \ln \prod_{i=0}^{N-1} \frac{\pi_{\lambda_{i+1}}(\;\; x_i \;\;)}{\pi_{\lambda_{i+1}}(x_{i+1})}. \tag{3.9}$$

Observe that the first term is a log-ratio of conditioned path probabilities. The denominator is the numerator's time reversal: the probability of obtaining the reversed path $x_{N:0}$ conditioned on starting in the state x_N and subject to the reversed protocol $\lambda_{N:0}$. The second term is exactly $-\mathcal{Q}_{\text{ex}}$ by the discrete form of Eq. (3.8). And so, by identifying $\mathcal{Q} \doteq \mathcal{Q}_{\text{ex}} + \mathcal{Q}_{\text{hk}}$, housekeeping heat is a component of the generalized *total heat* $\mathcal{Q}$.

In the (single heat bath) thermal example, one recovers units of energy as $\mathcal{Q} \to \beta Q$ and $\mathcal{Q}_{\text{hk}} \to \beta Q_{\text{hk}}$. And, the resulting total heat is consistent with formulations based on *microscopic reversibility* [**38**]. Equivalently, we could have started with this microscopic reversibility condition for even state spaces and arrived at the appropriate housekeeping heat.

With this in mind, consider *relaxing* the even state space assumption. Doing so and keeping the appropriate microscopic reversibility condition allows for an analogous splitting of housekeeping heat—modified so that the denominator's terms are negated where required—and excess heat, consistent with previous considerations of odd-parity NESS systems [**50**, **51**]. While we note that an analogue to our Eq. (3.24) holds, we do not treat this further here.

Now, consider a Markov dynamic of order 1. That is, conditioning on the previous time step fully characterizes the probability distribution over futures. Then, the first term reduces to the logarithm of a product of one-step conditional probabilities. And, $\mathcal{Q}_{\text{hk}}$ tracks the degree of *detailed-balance violation* over the trajectory. This is in agreement with existing definitions [**40**, **42**, **52**]. Concretely, detailed-balanced dynamics imply $\mathcal{Q}_{\text{hk}} = 0$ for every trajectory. If any trajectory yields $\mathcal{Q}_{\text{hk}} \neq 0$, the dynamic is necessarily nondetailed-balanced.

Finally, recall that by definition $\mathcal{Q}_{\text{hk}} = 0$ for an ESS system. Taken together with assuming an even state space—ensuring correct "reverse" probabilities—the Markov condition says, succinctly:

$$\textit{nondetailed-balanced dynamics} \iff \textit{nonequilibrium steady states}.$$

Recall that the Markov condition is appropriate for many microscopically-modeled thermal systems such as overdamped Langevin dynamics, as well as for a host of biological systems like the ion channels we consider later.

Nonzero housekeeping heat actually *necessitates* including an auxiliary reservoir for a complete picture. Recall Fig. 3.2. This follows since a NESS system, even fully relaxed to its stationary distribution, constantly dissipates housekeeping heat to the thermal reservoir. (And does so at an average rate of $\mathrm{d}\left\langle Q_{\text{hk}}\right\rangle/\mathrm{d}t$.) Yet, with the protocol parameter fixed, no work (or excess work) is done: $W = 0$ by Eq. (3.2). The system's average energy does not change, though, since the parameter and individual state uniquely set its energies: $\mathrm{d}\left\langle E\right\rangle/\mathrm{d}t = 0$.

The conclusion is that energy flux through the system, observed in the housekeeping dissipation to the thermal reservoir, must come from *somewhere* not otherwise described by the ideal constructs. In other words, in the thermodynamically-interpretable setting, nondetailed-balanced dynamics are signatures of unaccounted-for degrees of freedom. In this way, the constructions in Eqs. (3.8) and (3.9) provide the tools to isolate this *homeostatic* part of a system's energetic fluxes, so called for its role maintaining homeostatic (steady-state) conditions.

We close by calling out a feature on direct display in Eq. (3.9). While placing minimal restrictions on the *dynamics*, problems arise when any path is strictly irreversible, in the sense that a nonzero-probability forward trajectory is associated a zero-probability reverse. Then, $\mathcal{Q}_{\text{hk}}$ diverges. And this seemingly forbids dynamics in finite state spaces with one-way-only transitions.

In the thermodynamic interpretation, such a transition costs infinite dissipation. And, with this realization, usually a model's mesoscopic nature comes to bear. Indeed, Ref. [**42**] in its related sodium channel analysis remarks that "more careful experimental effort should be done to bound the actual housekeeping entropy production in these ion channels". The following section demonstrates that a new *trajectory class fluctuation theorem* provides a tool for analyzing such experiments and circumvents the divergence while still placing strong bounds on fluctuations.

3.3. Fluctuations and Free Energy

So far, we defined the generalized quantities $\mathcal{W}_{\text{ex}}$, $\mathcal{Q}_{\text{ex}}$, and $\mathcal{Q}_{\text{hk}}$ and elucidated their meanings outside the ESS regime. As with their ESS counterparts, though, they depend on the specific path a system takes through its state space under a particular protocol. A suite of statistical tools called *fluctuation theorems* (FTs) tie such nonequilibrium behaviors to equilibrium (or steady-state, more generally) quantities. They come in three primary flavors: (i) *integral* FTs (IFTs) concern weighted averages over all possible trajectories, (ii) *detailed* FTs (DFTs) fix the relationship between a specific path and its associated reversal, and (iii) *trajectory class* FTs (TCFTs) interpolate between the two [**53**].

The remainder of this section compares and contrasts these, discusses their relation to free energy, and concludes with a TCFT for NESS systems. This sets the stage for analyzing the two ion channels' thermodynamic responses—expressed in terms of excess work, excess heat, and housekeeping heat—to complex environmental signals.

3.3.1. Fluctuation theorems.

Integral and detailed FTs each exhibit complementary tradeoffs—tradeoffs discussed below as we introduce the theorems. TCFTs, meanwhile, combine the strengths of both and so are adaptable to a variety of systems and experimental conditions. Here we present a general TCFT valid for NESS systems. It simultaneously extends the previously known ESS FT and reveals experimental difficulties unique to NESS systems, ultimately suggesting a need for new experimental tools.

Jarzynski's equality [**36**,**37**], an IFT and the progenitor of the FTs we consider, links equilibrium free energies to the averaged exponential work distribution. It applies specifically to ESS systems that begin in their equilibrium distribution and are connected to a single heat bath. Under these conditions:

$$\left\langle \mathrm{e}^{-\beta W} \right\rangle = \mathrm{e}^{-\beta \Delta F^{\text{eq}}} \ , \tag{3.10}$$

where the angle brackets again refer to a weighted average over all possible trajectories. That is, Jarzynski's equality ties an arbitrarily nonequilibrium quantity—the averaged exponential work

$\left\langle \mathrm{e}^{-\beta W} \right\rangle$—to the equilibrium free energy difference ΔF^{eq}—a state function. Practically, this enables free energy estimation from nonequilibrium work measurements [**54**].

It comes with disadvantages, however. In particular, extremely rare paths often dominate the exponential work distribution [**55**], leading to poor statistical accuracy when estimating with finitely many experimental realizations. Nonetheless, Jarzynski's equality has been confirmed for a wide variety of systems [**56**, **57**, **58**]. In addition, while Eq. (3.10) only applies to ESS systems, a variety of generalizations have been derived and tested for NESS systems [**30**, **40**, **41**, **59**].

In contrast to Jarzynski's IFT, the *detailed* FTs (DFTs), express a symmetry relation between a particular trajectory-protocol pair and its appropriate time reversal. Perhaps the most well-known of these is due to Crooks [**38**, **39**], which is complementary to Jarzynski's IFT in several ways. For one, it makes the same assumptions: an ESS system connected to a bath, beginning in equilibrium and driven away from it. For another, Jarzynski's IFT results directly from trajectory-averaging both sides of Crooks' DFT. Before presenting the DFT, though, we pause to precisely define and set notation for what we mean by an "appropriate reversal".

Consider a system that begins in state distribution $\boldsymbol{\mu}_{\mathrm{F}}$, is driven by the protocol $\lambda_{1:N}$, and realizes a trajectory in the measurable subset $C \subseteq \mathcal{X}^{N+1}$. We call C a *trajectory class*. Then, we define the *forward process probability* as:

$$\mathcal{P}_{\boldsymbol{\mu}_{\mathrm{F}}}(C) \doteq \Pr\left(C \mid X_0 \sim \boldsymbol{\mu}_{\mathrm{F}}; \lambda_{1:N}\right) \ . \tag{3.11}$$

(Here, $\sim$ means "is distributed as".) Now, consider the same system beginning in the distribution $\boldsymbol{\mu}_{\mathrm{R}}$ and driven by the *reverse protocol* $\widetilde{\lambda}_{N:1}$, where the tilde indicates negation of time-odd variables (such as magnetic field). In turn, we define the *reverse process probability*:

$$\mathcal{R}_{\boldsymbol{\mu}_{\mathrm{R}}}(C) \doteq \Pr\left(C \mid X_0 \sim \boldsymbol{\mu}_{\mathrm{R}}; \widetilde{\lambda}_{N:1}\right). \tag{3.12}$$

For finite state spaces, Eqs. (3.11) and (3.12) define distinct probability measures on the trajectory space. In a continuous state space, we use the same notation to indicate probability *densities*.

Let $\boldsymbol{\pi}_{\mathrm{F}} \doteq \boldsymbol{\pi}_{\lambda_0}$ and $\boldsymbol{\pi}_{\mathrm{R}} \doteq \boldsymbol{\pi}_{\widetilde{\lambda}_N}$. In these terms, Crooks' DFT reads:

$$\frac{\mathcal{P}_{\boldsymbol{\pi}_{\mathrm{F}}}(x_{0:N})}{\mathcal{R}_{\boldsymbol{\pi}_{\mathrm{R}}}(\widetilde{x}_{N:0})} = \mathrm{e}^{\beta(W - \Delta F^{\mathrm{eq}})}. \tag{3.13}$$

As with Jarzynski's IFT, the Crooks DFT has withstood experimental test [**60**] and seen use in empirically estimating free energy differences [**54**]. Also, paralleling Jarzynski's IFT, Crooks' DFT has been generalized to a variety of NESS systems [**42**, **61**, **62**].

We highlight Ref. [**42**]'s generalization of these. We recall, in particular, its Eq. (25), since it is the DFT upon which we base our TCFT.

Here and in the remainder, we assume even state and protocol spaces (keeping in mind Sec. 3.2.3's notes on relaxing this assumption), so there is never negation under time reversal. However, in further contrast to Crooks' DFT, we do *not* assume equilibrium steady states (or detailed balance), any particular starting distribution for the forward and reverse processes, nor a single heat bath system (or any specific bath structure). Instead, we require only the functionals $\mathcal{W}_{\text{ex}}$ and $\mathcal{Q}_{\text{hk}}$ as defined in Eqs. (3.8) and (3.9), along with an additional one—the (unitless) *nonsteady-state free energy*:

$$\mathcal{F}_{\lambda}^{\text{nss}}(\boldsymbol{\mu}, x) \doteq \ln \frac{\mu(x)}{\pi_{\lambda}(x)} \ . \tag{3.14}$$

Its name derives from its indicating how far a given distribution is from the associated steady-state distribution. Indeed, on state averaging we have $\langle \mu | \mathcal{F}_{\lambda}^{\text{nss}} \rangle = D_{\text{KL}}[\boldsymbol{\mu} \,\|\, \boldsymbol{\pi}_{\lambda}]$, where $D_{\text{KL}}[\boldsymbol{p} \,\|\, \boldsymbol{q}]$ is the Kullback-Leibler divergence between distributions p and q [**63**]. As with the other functionals generalized to the stochastic process picture, it carries meaning—departure from steady-state conditions—outside of energetic or thermal assumptions.

Given this, Ref. [**42**]'s DFT is:

$$\frac{\mathcal{R}_{\boldsymbol{\mu}_{\text{R}}}\left(x_{N:0}\right)}{\mathcal{P}_{\boldsymbol{\mu}_{\text{F}}}\left(x_{0:N}\right)} = \mathrm{e}^{-(\mathcal{W}_{\text{ex}} + \mathcal{Q}_{\text{hk}} - \Delta\mathcal{F}^{\text{nss}})} \ , \tag{3.15}$$

where $\Delta\mathcal{F}^{\text{nss}} = \mathcal{F}_{\lambda_N}^{\text{nss}}(\boldsymbol{\mu}_{\text{R}}, x_N) - \mathcal{F}_{\lambda_0}^{\text{nss}}(\boldsymbol{\mu}_{\text{F}}, x_0)$ is a correction due to starting the forward and reverse processes out of steady state. If we began the forward and reverse processes in their associated steady-state distributions, by definition we would have $\Delta\mathcal{F}^{\text{nss}} = 0$.

As a mathematical statement involving a stochastic process' trajectories, their probabilities, and the functionals $\mathcal{W}_{\text{ex}}$, $\mathcal{Q}_{\text{hk}}$, and $\Delta\mathcal{F}^{\text{nss}}$ we have so far defined, Eq. (3.15) holds *independent* of any thermodynamic assumptions. Yet, as before, reducing it to thermodynamically meaningful cases is

straightforward and illuminates several important considerations when moving from ESS into NESS regimes.

3.3.2. Multiple NESS IFTs and second laws.

In the ESS case, Eq. (3.15) reduces neatly to Crooks' DFT of Eq. (3.13), which under integration directly yields Jarzynski's equality. However, the NESS setting introduces more freedom under this type of integration.

Consider rearranging Eq. (3.15) like so:

$$\mathcal{R}_{\boldsymbol{\mu}_{\mathrm{R}}}\left(x_{N:0}\right) = \mathcal{P}_{\boldsymbol{\mu}_{\mathrm{F}}}\left(x_{0:N}\right) \mathrm{e}^{-\left(\mathcal{W}_{\mathrm{ex}} + \mathcal{Q}_{\mathrm{hk}} - \Delta\mathcal{F}^{\mathrm{nss}}\right)}. \tag{3.16}$$

Then integrate both sides over *all* trajectories $x_{0:N}$. The righthand side directly yields the forward trajectory average of the exponential, while the lefthand side yields 1 by probability conservation (and since the sets of all measurable forward and reverse trajectories are the same set). This gives a generalized IFT:

$$1 = \left\langle \mathrm{e}^{-\left(\mathcal{W}_{\mathrm{ex}} + \mathcal{Q}_{\mathrm{hk}} - \Delta\mathcal{F}^{\mathrm{nss}}\right)} \right\rangle \tag{3.17}$$

and—via Jensen's inequality—a generalized second law:

$$\langle \mathcal{W}_{\mathrm{ex}} \rangle + \langle \mathcal{Q}_{\mathrm{hk}} \rangle - \Delta D_{\mathrm{KL}}[\boldsymbol{\mu} \,\|\, \boldsymbol{\pi}] \geq 0, \tag{3.18}$$

where $\Delta D_{\mathrm{KL}}[\boldsymbol{\mu} \,\|\, \boldsymbol{\pi}] = D_{\mathrm{KL}}[\boldsymbol{\mu}_{\mathrm{R}} \,\|\, \boldsymbol{\pi}_{\mathrm{R}}] - D_{\mathrm{KL}}[\boldsymbol{\mu}_{\mathrm{F}} \,\|\, \boldsymbol{\pi}_{\mathrm{F}}]$. This latter term is a classical analogue to the "initial-state dependence" of Ref. [**64**]; it quantifies additional entropic dissipation when beginning and ending out of the steady-state distribution.

Yet Eq. (3.17) is not unique. To take one example, by direct calculation as in Ref. [**40**], we find for Markov dynamics:

$$1 = \left\langle \mathrm{e}^{-\mathcal{F}^{\mathrm{nss}}_{\lambda_0}(\boldsymbol{\mu}_{\mathrm{F}}, x_0) - \mathcal{W}_{\mathrm{ex}}} \right\rangle \tag{3.19}$$

$$\implies \langle \mathcal{W}_{\mathrm{ex}} \rangle + D_{\mathrm{KL}}[\boldsymbol{\mu}_{\mathrm{F}} \,\|\, \boldsymbol{\pi}_{\mathrm{F}}] \geq 0, \tag{3.20}$$

a slight generalization of their result with the inclusion of initial-state dependence (thereby relaxing the requirement of steady-state initial conditions). In other words, a second law holds for $\langle \mathcal{W}_{\mathrm{ex}} \rangle$

itself, not just for its sum with $\langle \mathcal{Q}_{\mathrm{hk}} \rangle$. This is not only a meaningfully different bound, but this IFT also does not result naturally from the underlying DFT.

To take another example, directly substituting $\langle \mathcal{Q}_{\mathrm{hk}} \rangle = \langle \mathcal{Q} \rangle - \langle \mathcal{Q}_{\mathrm{ex}} \rangle$ into Eq. (3.17) generalizes to a different IFT:

$$1 = \left\langle \mathrm{e}^{-(\mathcal{Q} + \Delta\phi - \Delta\mathcal{F}^{\mathrm{nss}})} \right\rangle , \tag{3.21}$$

first proven in Ref. [**41**].

Finally, taken on their own, Eqs. (3.18) and (3.20) do not imply a third NESS IFT—and so Second Law—for housekeeping heat alone; cf. again Ref. [**41**]. However, as we now show, it is implied rather directly by the combination of Eqs. (3.15) and (3.19).

First rearrange Eq. (3.15) as:

$$\mathcal{R}_{\boldsymbol{\mu}_{\mathrm{R}}}(x_{N:0})\, \mathrm{e}^{\mathcal{W}_{\mathrm{ex}} - \mathcal{F}^{\mathrm{nss}}_{\lambda_N}(\boldsymbol{\mu}_{\mathrm{R}}, x_N)} = \mathcal{P}_{\boldsymbol{\mu}_{\mathrm{F}}}(x_{0:N})\, \mathrm{e}^{-\mathcal{Q}_{\mathrm{hk}} - \mathcal{F}^{\mathrm{nss}}_{\lambda_0}(\boldsymbol{\mu}_{\mathrm{F}}, x_0)} \ .$$

Again, we wish to integrate both sides over all $x_{0:N}$, but at first glance the lefthand side (first line) poses an issue: $\mathcal{W}_{\mathrm{ex}}$ refers to the excess work over a trajectory driven by the forward protocol, while $\mathcal{R}$ is the probability of a trajectory as driven by the reverse protocol.

Fortunately, $\mathcal{W}_{\mathrm{ex}}$ is odd under full time reversal: $\mathcal{W}_{\mathrm{ex}} = -\mathcal{W}^{\mathrm{R}}_{\mathrm{ex}}$, where $\mathcal{W}^{\mathrm{R}}_{\mathrm{ex}}$ is the excess work generated by the time-reversed trajectory driven by the time-reversed protocol. This matches up driving protocols and "initial" conditions on the lefthand side. Since Eq. (3.19) holds regardless of the chosen protocol or starting distribution, under integration the lefthand side is unity. The righthand side, meanwhile, becomes simply the forward trajectory average of its argument. And, we obtain the generalized IFT for housekeeping heat:

$$1 = \left\langle \mathrm{e}^{-\mathcal{F}^{\mathrm{nss}}_{\lambda_0}(\mu_{\mathrm{F}}, x_0) - \mathcal{Q}_{\mathrm{hk}}} \right\rangle \tag{3.22}$$

$$\implies \langle \mathcal{Q}_{\mathrm{hk}} \rangle + D_{\mathrm{KL}}[\boldsymbol{\mu}_{\mathrm{F}} \,\|\, \boldsymbol{\pi}_{\mathrm{F}}] \geq 0, \tag{3.23}$$

once again extended to include the effects of initial-state dependence.

To adopt Ref. [**41**]'s language, these IFTs—for $\mathcal{W}_{\mathrm{ex}} + \mathcal{Q}_{\mathrm{hk}}$, $\mathcal{W}_{\mathrm{ex}}$-only, $\mathcal{Q} + \Delta\phi$, and $\mathcal{Q}_{\mathrm{hk}}$-only—are "genuinely different" but no longer require especially "different derivation[s]" nor restrictive physical

assumptions. To emphasize the former point, though: just as with the equilibrium second law $\langle \Delta S \rangle \geq 0$, these hold only under full trajectory averaging. That is, individual rare trajectories (or sets thereof) may well produce negative excess works, negative housekeeping heats, or both. Sec. 3.5.1 explores these consequences for our example ion channels.

Notably absent is the notion of steady-state free energy, analogous to the equilibrium free energy from Jarzynski's equality. Defining one for general NESS systems remains problematic, in part since the steady-state distributions may no longer be Boltzmann. Instead, the excess work subsumes what would have been a steady-state free energy difference, and we work directly with it. The downside, however, is the inability to extract such a free energy as a "steady state" quantity separate from the path-dependent nonequilibrium dynamical ones. Indeed, this was an extremely important consequence of Eq. (3.10).

The suite of IFTs given by our Eqs. (3.18), (3.20), and (3.23), however, do include strong connections between path-independent and path-dependent quantities in the form of initial-state dependence and changes in steady-state surprisal. Unlike for ESS systems, however, even in well-controlled NESS thermal examples applying the IFTs presents a rather serious experimental challenge: direct measurement of *heat* (most notably housekeeping heat). Even when testing FTs phrased in terms of heat, often *work* (excess or not) is experimentally tracked [**30**]. And so, we expect direct measurement to be a key, requisite step in leveraging the resulting FTs to analyze experimental NESS systems.

3.3.3. NESS trajectory class fluctuation theorem.

With appropriate DFTs and IFTs for NESS systems now in hand, we are confronted with yet another challenging experimental tradeoff. Just as the IFTs suffer from extremely rare-but-large contributions, the DFTs require precise control and measurement of *individual* realizations, as well as accurate estimations of individual realization probabilities (or their ratios). This is often intractable even in principle. For example, as experimental systems, the ion channels considered shortly do not permit measurement of the conformational states themselves. Instead, ionic current is the only observable. Moreover, the state space topology varies with each individual rate model [**65**]. This is all to say that thermodynamic analysis requires a more flexible intermediary between the DFT's trajectory-level information and the IFT's ensemble-level information.

Ref. [**53**] recently provided just such an intermediary for ESS systems—the *trajectory class* FT (TCFT). At root, it relates the forward and reverse probabilities of an arbitrary subset of trajectories—the *trajectory class* C as introduced earlier—to the average exponential work *within* that trajectory class. In this way, the TCFT is maximally adaptable to experimental conditions: It need neither suffer rare-event errors nor require individual-trajectory-level control. Instead, whatever the unique experimental conditions at hand, it provides a framework for laying out an associated FT. As a practical matter, the TCFT has already provided a diagnostic tool for monitoring the thermodynamics of successful and failed microscopic information processing in superconducting flux logic [**53**, **66**].

The following extends Ref. [**53**]'s ESS TCFT (Eq. (3) there) in two ways. First, we allow for NESS systems. Second, we allow starting the forward and reverse processes in arbitrary distributions $\boldsymbol{\mu}_{\mathrm{F}}$ and $\boldsymbol{\mu}_{\mathrm{R}}$, respectively. This results in our exponential NESS TCFT, derived in App. A:

$$\frac{\mathcal{R}_{\boldsymbol{\mu}_{\mathrm{R}}}\left(C_{\mathrm{R}}\right)}{\mathcal{P}_{\boldsymbol{\mu}_{\mathrm{F}}}\left(C\right)} = \left\langle \mathrm{e}^{-(\mathcal{W}_{\mathrm{ex}}+\mathcal{Q}_{\mathrm{hk}}-\Delta\mathcal{F}^{\mathrm{nss}})} \right\rangle_{C} \ , \tag{3.24}$$

where $\langle\cdot\rangle_{C}$ denotes the conditionally-weighted average over only those trajectories in the class C and the *reverse trajectory class* $C_{\mathrm{R}} \doteq \{x_{N:0} \,|\, x_{0:N} \in C\}$.

Eq. (3.24) imports to the NESS setting all the benefits of the TCFT. Most notably, it adapts readily to a variety of experimental conditions while maintaining robust statistics. The associated DFT and IFT emerge simply by setting the class C to be a single trajectory or the set of all trajectories, respectively. Equation (3.24), as with its ESS counterpart, allows selecting trajectory classes most accessible in a particular experimental configuration and *then* proposes the appropriate theory against which to test.

Once again, in this form Eq. (3.24) makes only two assumptions about a stochastic process, as outlined previously: a unique stationary distribution for each λ and an even state space. It reproduces Ref. [**53**]'s TCFT given ESS assumptions. Similarly, it reproduces Ref. [**42**]'s Eq. (52) when the class is chosen to start and end in a particular desired subset of states. However, our main result holds independently of any energetic, Markovian, or particular class assumption.

Generalization to NESS systems is not without caveat, however. $\mathcal{Q}_{\mathrm{hk}}$ plays a central role and we do not have our state- and path-independent equilibrium free energy to extract from the

average and estimate. This suggests experimentally tracking the housekeeping heat itself is key to understanding nondetailed balanced, NESS systems. (Alternatively, one could monitor the total heat per Eq. (3.18).) This is not surprising, considering $Q_{\rm hk}$ is *the* defining difference between an ESS and NESS system.

3.4. Na^+ and K^+ Ion Channels

Armed with this toolkit, we are now ready to probe the thermodynamic functionality of two example biophysical systems: Ref. [**44**]'s delayed-rectifier potassium (K^+) and fast sodium (Na^+) voltage-gated ion channels. (See its Figs. 5.12 and 5.13, reproduced in our Figs. 3.3 and 3.4, respectively. The conformational states of these models from Ref. [**44**] are defined in our Figs. 3.3 and 3.4, while the forms for their transition rates we reproduce in Eqs. (3.27)–(3.31).) These single-channel models are based on relatively more macroscopic descriptions of channel ensembles due to Hodgkin and Huxley [**67**]. However, they better represent the interdependencies between molecular-conformational transformations and more accurately reproduce experimentally-observed currents, especially for the Na^+ channel [**44**].

The models are both continuous-time Markov chains (CTMCs), whose dynamics are described by the stochastic master equation:

$$\frac{\mathrm{d}}{\mathrm{d}t} \langle \mu(t)| = \langle \mu(t)| \, \mathbf{G}_\lambda . \tag{3.25}$$

The row vector $\langle \mu(t)|$ specifies the *state distribution* or *mixed state* at time t; its elements are $\mu(x,t) \doteq \Pr\left(X(t) = x\right)$. The transition rate matrix $\mathbf{G}_\lambda$ is controlled by the protocol and, thus, varies with time. The would-be steady-state distributions for each λ are given by:

$$\langle \pi_\lambda | \, \mathbf{G}_\lambda = \langle 0| \ , \tag{3.26}$$

with $\langle 0|$ the all-0 vector.

The transition-rate matrices corresponding to the two channels are:

$$\mathbf{G}_{\lambda}^{\mathrm{K}^+} = \begin{bmatrix} -4a_n & 4a_n & 0 & 0 & 0 \\ b_n & -(3a_n + b_n) & 3a_n & 0 & 0 \\ 0 & 2b_n & -(2a_n + 2b_n) & 2a_n & 0 \\ 0 & 0 & 3b_n & -(3b_n + a_n) & a_n \\ 0 & 0 & 0 & 4b_n & -4b_n \end{bmatrix} \tag{3.27}$$

and

$$\mathbf{G}_{\lambda}^{\mathrm{Na}^+} = \begin{bmatrix} -3a_m & 3a_m & 0 & 0 & 0 \\ b_m & -(2a_m + b_m + k_1) & 2a_m & 0 & k_1 \\ 0 & 2b_m & -(a_m + 2b_m + k_2) & a_m & k_2 \\ 0 & 0 & 3b_m & -(3b_m + k_3) & k_3 \\ 0 & 0 & a_h & 0 & -a_h \end{bmatrix}. \tag{3.28}$$

Letting λ denote the transmembrane voltage, the associated transition rates are:

$$a_m(\lambda) = \frac{(\lambda + 40\ \mathrm{mV})/10\ \mathrm{mV}}{1 - \exp\left(-(\lambda + 40\ \mathrm{mV})/10\ \mathrm{mV}\right)}, \qquad b_m(\lambda) = 4\exp\left(-(\lambda + 65\ \mathrm{mV})/18\ \mathrm{mV}\right), \tag{3.29}$$

$$a_h(\lambda) = \frac{7}{100}\exp\left(-(\lambda + 65\ \mathrm{mV})/20\ \mathrm{mV}\right), \qquad k_1 = \frac{6}{25}\ \mathrm{ms}^{-1},\ k_2 = \frac{2}{5}\ \mathrm{ms}^{-1},\ k_3 = \frac{3}{2}\ \mathrm{ms}^{-1}, \tag{3.30}$$

$$a_n(\lambda) = \frac{(\lambda + 55\ \mathrm{mV})/100\ \mathrm{mV}}{1 - \exp\left(-(\lambda + 55\ \mathrm{mV})/10\ \mathrm{mV}\right)},\ \text{and}\ b_n(\lambda) = \frac{1}{8}\exp\left(-(\lambda + 65\ \mathrm{mV})/80\ \mathrm{mV}\right). \tag{3.31}$$

We map these CTMC systems to discrete-time stochastic processes by taking $\lambda(t)$ fixed for sufficiently small time intervals Δt, generating the transition matrices:

$$\mathbf{T}_{\lambda}^{\Delta t} \doteq \mathrm{e}^{\Delta t\, \mathbf{G}_{\lambda}} \tag{3.32}$$

for each such time interval. Having discretized time in this way, they are examples of the thermodynamic scheme in Fig. 3.2, being surrounded by a single thermal environment at body temperature.

A voltage-gated ion channel's basic function is to selectively allow ions to permeate a cell membrane. The selection is based on the transmembrane voltage—the voltage difference between the membrane's inside and outside. In our models, this difference is specified by the parameter λ, and so a neuronal action potential spike is a specific protocol. Ref. [**44**]'s K^+ and Na^+ models correspond to channels that play crucial roles in generating and propagating such spikes in mammalian neuronal axons. Both Markov chain models are estimated from single-channel experiments.

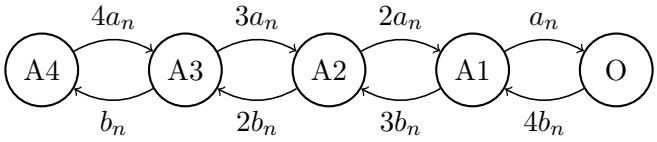

FIGURE 3.3. Continuous-time Markov chain model of the K^+ channel adapted from Fig. 5.12 of Ref. [**44**]. Self-transitions are implied. In the states labeled An, a number $n \in \{1, 2, 3, 4\}$ activation gates close the channel. O labels the open channel state, the only one in which K^+ current can flow through the channel. The rate parameters a_n and b_n are voltage-dependent; their functional forms are given in Eqs. (3.31). This channel model is fully detailed-balanced, in the sense that Eq. (3.33) vanishes for every allowed transition pair.

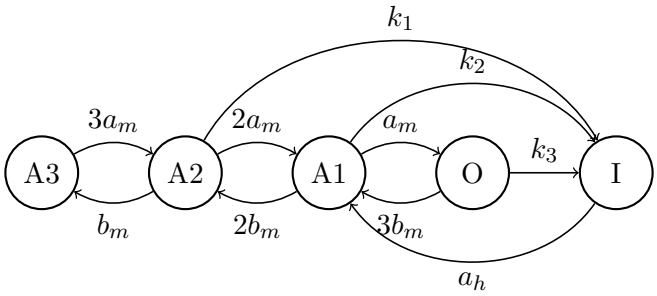

FIGURE 3.4. Continuous-time Markov chain model of the Na^+ channel adapted from Fig. 5.13 of Ref. [**44**]. Self-transitions are implied. In the states labeled An, a number $n \in \{1, 2, 3\}$ activation gates close the channel. O labels the open channel state, in which Na^+ current flows through the channel. Finally, I labels the channel's inactivation by its inactivation gate—its so-called ball and chain. The rate parameters a_m, b_m, and a_h are all voltage-dependent; their functional forms are given in Eqs. (3.29)–(3.30). The rate constants k_1, k_2, and k_3 are given by Eqs. (3.30). Unlike the K^+ channel, this model of the Na^+ channel features one-way transitions in the rate dynamic—states O to I and A2 to I. These transitions are maximally irreversible. These imply divergent infinitesimal housekeeping heat in the sense of Eq. (3.33). In addition to these, many of the other transition pairs do *not* satisfy detailed balance—Eq. (3.33) evaluates finite but nonzero.

We selected these two channel models for several reasons.

First, in terms of their biological function, they are comparable: They accomplish similar tasks, are connected to the same environmental parameters, and are suitably mesoscopic. That is, despite being more detailed than the Hodgkin-Huxley ensemble models, neither model accounts for the many additional molecular degrees of freedom involved in the channel dynamics, be it steady-state or transient functions. The small effective state spaces in the Markov chain models reflect this.

One consequence of this implied coarse graining is that any total entropy production is a lower bound [**31**]. Still, we are able to make headway analyzing their nonequilibrium dynamics *without* knowledge of the underlying coarse-graining methods—knowledge missing for the vast majority of mesoscopic complex systems.

Second, the Na^+ channel's transition rates do not, in general, satisfy detailed balance, while the K^+ channel's do. Indeed, the Na^+ channel model includes both finitely nondetailed-balanced transition pairs and one-way-only transition rates, which imply divergent infinitesimal-time housekeeping heat. These violations of ideality are typical and widely encountered in molecular biophysical systems, as well as in real-world thermodynamic processes.

To appreciate these nonidealities, note that under our time discretization and the Markov property, the *infinitesimal*(-time) *housekeeping heat*—a single "step" of Eq. (3.9)—for a transition between states indexed by i to j:

$$\begin{aligned}
[\mathrm{d}\mathcal{Q}_{\mathrm{hk}}]_{ij} &= \lim_{\Delta t \to 0} \ln \frac{\pi_\lambda(x_i)\left[\mathbf{T}_\lambda^{\Delta t}\right]_{ij}}{\pi_\lambda(x_j)\left[\mathbf{T}_\lambda^{\Delta t}\right]_{ji}} \\
&= \lim_{\Delta t \to 0} \ln \frac{\pi_\lambda(x_i)[\mathbf{G}_\lambda + \mathcal{O}(\Delta t)]_{ij}}{\pi_\lambda(x_j)[\mathbf{G}_\lambda + \mathcal{O}(\Delta t)]_{ji}}.
\end{aligned} \tag{3.33}$$

For systems with one-way-only transition rates, such as from the second to the fifth state of the Na^+ channel (indexing the states left to right, Fig. 3.4), infinitesimal housekeeping heat diverges. This contrast between the two channels—wherein one of them exhibits equilibrium steady states and the other nonequilbrium steady states—allows showcasing several features of the NESS TCFT and of the NESS framework more broadly. These, in turn, reveal the dynamical interplay of different modes of thermodynamic transformation.

As a test case, the K^+ channel should satisfy the ESS TCFT (where $\mathcal{Q}_{\text{hk}} = 0$) while the Na^+ channel should violate it. Both, however, should satisfy our NESS TCFT of Eq. (3.24).

One benefit of the TCFT's averaging over arbitrary trajectory classes comes from avoiding the divergences implied by one-way transition rates: We select only those trajectories that do *not* include one-way transitions in the Na^+ channel, but still satisfy the appropriate DFT (and therefore TCFT) with those trajectories.

In this way, the NESS TCFT allows monitoring nonequilibrium fluctuations in systems with drastically different steady-state characteristics: detailed balance on the one hand and spurious divergences on the other.

Yet separating heat into excess and housekeeping components *also* enables direct comparison of the channels' *adaptive* energetics. Given the same environmental drive, which components of their dissipations are due solely to their internal adaptation to that drive? The excess heat, $\mathcal{Q}_{\text{ex}}$. This remains true without regard for the divergence implied by one model's steady states. In essence, we cleave the housekeeping infinity to directly compare adaptive energetics.

Finally, both models are simple and illustrative. There are many more-detailed candidate state-space models for the Na^+ channel: take those found in Refs. [**65**] and [**68**], for instance, whose variations have important implications for understanding responses to drug treatments [**69**]. While we do not analyze them directly, our techniques generalize to *any* such candidate models straightforwardly and provide an alternative formulation to that of Ref. [**68**]. Indeed, our ability to carry out these thermodynamic analyses provides new grounds for model selection, contingent on measurement techniques to experimentally extract the appropriate quantities.

3.5. Methods and Results

Our goal, ultimately, is to describe the nonequilibrium thermodynamics of driven mesoscopic NESS systems—how they respond thermodynamically to environmental stimuli and attempt to maintain stability. We take up the challenges here in two ways.

First, Sec. 3.5.1 samples individual trajectories from both channels under the neurobiologically-plausible action potential *spike protocol*. Trajectories in hand, it compares Q_{hk} versus W_{ex} for each, showcasing the need for corrected NESS D/TCFTs, revealing various modes of second-law-type

violations allowed of each channel, and discussing the surprising biophysical functionality these violations imply.

We derived the spike protocol by solving the reduced ODEs (8.5) and (8.6) of Izhikevich [**70**] (also presented earlier [**71**]), integrating via the explicit forward Euler method. We adopt the "regular spiking" parameters of their Fig. 8.12, except that we set the membrane capacitance to 1 pF and input DC pulse to 80 pA to change the time scale of a single pulse to 2 ms, more accurately reflecting measurements in Ref. [**44**]. For transmembrane voltage λ in units of [mV] and "recovery variable" (as introduced in Ref. [**70**]) u in [pA], we thus directly evolve

$$\frac{\mathrm{d}\lambda}{\mathrm{d}t} = 0.7(\lambda + 60)(\lambda + 40) - u + 80 \tag{3.34}$$

$$\frac{\mathrm{d}u}{\mathrm{d}t} = -0.03(2(\lambda + 60) + u) \tag{3.35}$$

$$\text{if } \lambda > 35 : \lambda \leftarrow -50, u \leftarrow u + 100. \tag{3.36}$$

The protocol begins with the transmembrane voltage at its resting potential (of -60 mV in this parameter set) and u at 0 pA. We took 200 001 equidistant time steps, resulting in 10 ns increments. A visual of this protocol appears as the dashed line in Figs. 3.8 and 3.10.

Second, Sec. 3.5.2 calculates the full trajectory-averaged excess heat and work—$\langle Q_{\mathrm{ex}} \rangle$ and $\langle W_{\mathrm{ex}} \rangle$, respectively—of the two channel models under both our spike protocol and the 12 ms *pulse protocol* matching Ref. [**42**]'s and providing for direct comparison with their results. We took the same number of equidistant time steps, resulting in 60 ns increments. In the spike case, we directly compare for the first time the detailed *adaptive* energetics of the two channel types under a neurobiologically-plausible protocol. Our analysis both reveals functionality not visible under a pulse drive and highlights the preceding theoretical framework's ability to directly compare the channels' adaptive energetic response to the same drive, despite their dramatically different steady-state behaviors.

The ion channels are examples of Fig. 3.2's scheme with a single heat bath, so we have that $\mathcal{Q}_{\mathrm{hk}} \to \beta Q_{\mathrm{hk}}$, $\mathcal{W}_{\mathrm{ex}} \to \beta W_{\mathrm{ex}}$, and $\mathcal{Q}_{\mathrm{ex}} \to \beta Q_{\mathrm{ex}}$. For convenience, then, we label all thermodynamic axes in units of $[k_{\mathrm{B}}T]$. In more general settings, however, these functionals are purely dynamical quantities, to be understood and interpreted as indicated in Secs. 3.2.2 and 3.2.3.

Admittedly, the selected ion channel models are not realistic in the sense that they do not incorporate feedback between the transmembrane potential and the ion channel states themselves. (Or, put another way, they ignore correlation between channels.) This feedback is crucial to *in vivo* generation of the spike patterns. In one sense, this simplification is actually an *advantage* of our approach, since we ask: *Given* a particular transmembrane protocol—regardless of how it got there—how do these individual channels respond? How do they absorb and dissipate energy in response to this environment?

3.5.1. NESS TCFT reveals thermal response.

This section compares the detailed-balanced dynamics of the K^+ channel with the nondetailed-balanced dynamics of the Na^+ channel. It demonstrates agreement between ESS and NESS FTs in the former, but violation in the latter. This exposes the channels' different dynamical responses—how thermodynamic fluxes of energy and entropy support their distinct biophysical functioning. The TCFT's flexibility allows us to select only trajectories-of-interest and take partial sums on either side of the underlying DFT. This helps not only to gather experimental statistics—improving statistical efficiency—but also to generate statistics from models, as the following does.

While Eq. (3.33)'s first-order approximation is valid in the infinitesimal time limit, *any* finite time step—no matter how small—maps every zero in the transition-rate matrices to nonzero values in the discrete-time transition matrices. As long as any state can transition to any other *eventually* in the rate dynamic, we observe a *direct* transition from any state to any other state after *any finite time*. Mathematically, this results from higher-order terms in the matrix exponentials.

Since we wish to explicitly highlight the differences between the channels—the ESS in the K^+ case and the divergent transitions in the Na^+—we take the first-order approximation of Eq. (3.33). Formally, it defines a distinct discrete-time dynamic compared to taking the full matrix exponentials, but the fluctuation theorems apply just as well to this approximated dynamic. In sampling trajectories, we avoid the divergent Na^+ transitions altogether by selecting only paths that do not include them, yet another advantage the TCFT affords. This does not alter the TCFT's validity as long as we accurately collect the probabilities of the selected trajectories.

We *can* collect those probabilities, having the full transition dynamic in hand. However, simulating 200 001-step trajectories, the resulting probabilities are extraordinarily small. To ameliorate

numerical precision issues, we instead directly collect the natural logarithms of trajectory probabilities. Finally, since we wish to isolate the differences between the channels due to NESSs (or, equivalently in our case, to nondetailed-balanced dynamics), we make one last simplifying assumption before numerical simulation: We begin all forward and reverse processes in their local stationary distributions, setting $\Delta\mathcal{F}^{\text{nss}} = 0$.

This simplifies Eq. (3.24)'s DFT kernel to:

$$\ln \frac{\mathcal{P}_{\boldsymbol{\pi}_{\text{F}}}\left(x_{0:N}\right)}{\mathcal{R}_{\boldsymbol{\pi}_{\text{R}}}\left(x_{N:0}\right)} = \mathcal{W}_{\text{ex}}[x_{0:N}] + \mathcal{Q}_{\text{hk}}[x_{0:N}] \ . \tag{3.37}$$

Comparing this to Crooks' DFT as in Eq. (3.13) reveals the presence of $\mathcal{Q}_{\text{hk}}$ as the only difference. For an ESS system, this should vanish for all inputs; otherwise, it represents a violation of Crooks' DFT by a factor of $\mathrm{e}^{\mathcal{Q}_{\text{hk}}}$. To probe the violation, Fig. 3.5 directly plots Q_{hk}—via Eq. (3.9)—on the vertical axis, where each point represents an individual trajectory. We plot these values against W_{ex}—with the discrete form of Eq. (3.8)—on the horizontal axis to aid interpretation: Via Eqs. (3.20) and (3.23), there are individual second laws for both the generalized housekeeping heat and the generalized excess work. As with the familiar equilibrium second law, however, these are strictly true only on full trajectory averaging.

To arrive at Fig. 3.5, we sampled trajectories according to their distributions as given by each channel's first-order dynamics under spike driving. For the Na^+ channel, as previously mentioned, this excludes the one-way-only transitions. For the K^+ channel, we obtained 9 626 individual trajectories; for the Na^+ channel, we obtained 23 834.

Plotting housekeeping heat against excess work in Fig. 3.5 directly visualizes the independent kinds of negative entropy trajectories: where $\mathcal{W}_{\text{ex}} < 0$, we have single-shot violations of the familiar second law. In the isothermal environment of the ion channel models, trajectories for which this is the case imply channels that, under the spike protocol's drive, funnel energy to the work reservoir. Where $\mathcal{Q}_{\text{hk}} < 0$, however, we have a new kind of second law violation unique to the NESS setting: in context, these are channels which have taken energy *from* the heat bath to maintain NESS conditions, rather than dissipated to it. For this reason, in Fig. 3.5, we label these quadrants "housekeeping (HK) engines." Where total heat is also negative, the channel as a whole functions as a "total heat engine," but—notably—these possibilities are independent of one and other. To take

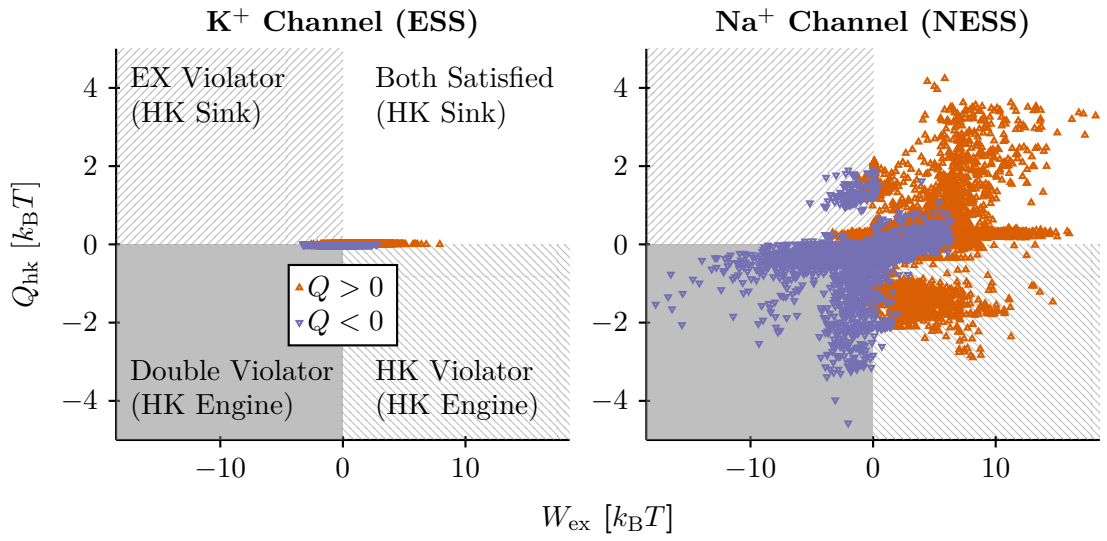

FIGURE 3.5. The two channel models compared via their nonequilibrium excess work and housekeeping heat distributions, respectively, in response to the spike protocol drive. Nonzero values of $Q_{\rm hk}$ indicate violations of the Crooks DFT Eq. (3.13), where the corrected NESS DFT Eq. (3.15) is needed. In addition, the axes here are of quantities with associated independent second laws; see Eqs. (3.20) and (3.23). And so, the labeled quadrants carry thermodynamic (and, in this case, biophysical) meaning as single-shot violations of each statistical second law.

but one example: the negative total heat trajectories of the first quadrant act as heat engines, yet the *housekeeping* part of their total heat remains dissipative.

As Fig. 3.5 shows, the additional dimension of thermodynamic behavior afforded by nonzero housekeeping heat and its associated second law gives rise to a number of otherwise inaccessible combinations. Driving the channels according to the biologically-plausible spike protocol also reveals a greater range of possible Crooks DFT violations than did the more artificial pulse-driven result of Ref. [**42**], with only several violations. Taken together, Fig. 3.5 reveals a rich taxonomy of thermodynamic behaviors for the Na^+ channel—behaviors that are not reflected (indeed, flattened) in the K^+ channel or, indeed, in *any* ESS system, where Crooks' DFT is satisfied and $\mathcal{Q}_{\rm hk} = 0$. In particular, there are four functionally-distinct thermodynamic quadrants, corresponding to the positive and negative values of $\mathcal{Q}_{\rm hk}$ and $\mathcal{W}_{\rm ex}$, and labeled by their excess and housekeeping functionality on the K^+ channel plot for clarity.

To be clear, each point on these plots corresponds to a *single* trajectory-reverse pair that itself is a valid trajectory class. Yet, (i) the samples themselves may be taken from a special class—for

the Na^+ channel we explicitly exclude resource-divergent trajectories—and (ii) any subsample on the plot corresponds to its own valid trajectory class as well.

The clustering of realized Q_{hk} in Fig. 3.5 reveals additional *structure* in Crooks DFT violations not previously observed. Apparently, there are distinct thermodynamic mechanisms that generate the violations. These result directly from the relative frequencies of transitions as functions of the driving protocol.

To lend additional insight into this structure, Fig. 3.6 plots the one-step rates of Q_{hk} production for each allowed transition in our modified Na^+ channel. The A3 $\leftrightarrow$ A2 transition pair is the only one of this dynamic that is fully detailed-balanced for all inputs; the A2 $\leftrightarrow$ A1 pair is *nearly* detailed-balanced, with very small housekeeping heat production. By comparison, both of the remaining transitions *strongly* violate detailed balance, and so contribute the bulk of the nonzero Q_{hk}.

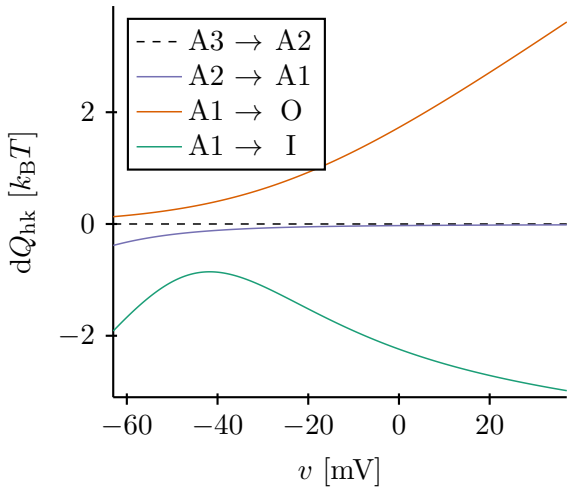

FIGURE 3.6. Na^+ channel transition rates of housekeeping heat Q_{hk} as functions of transmembrane potential. The total housekeeping heat along any stochastic trajectory, driven by any protocol, is the sum of these values with the associated state transition-protocol parameter pairs here.

Physically, the latter two correspond to transitions directly to/from the open and inactivated channel states. Interestingly, their violations run in opposite directions. On the one hand, the A1 $\rightarrow$ O transition dissipates housekeeping heat to the thermal environment—indeed, more as the membrane voltage rises. On the other hand, the A1 $\rightarrow$ I transition describes the "ball and chain" that plugs the channel without opening, leaving no opportunity for Na^+ current to flow.

Thermodynamically, this transition actually *absorbs* housekeeping heat from the thermal environment. Since the housekeeping heat production rate is odd under transition reversal, these roles are reversed for the reverse transitions. Thus, the results shown in Fig. 3.5 arise directly from integrating those in Fig. 3.6 according to each trajectory-protocol pair.

As a final consideration, we note that the ESS TCFT (and so the Crooks DFT) *do not claim* to be valid for NESS systems. That said, our results visually verify the facts that the NESS generalization both extends the range of validity of the TCFT and reduces in the correct way for ESS systems.

That we have a trajectory class form for the NESS TCFT, captured in our Eq. (3.24), imports its ESS progenitor's flexibility. That is, we need capture neither individual trajectory-level information to verify the DFT nor accurately sample the full trajectory space for an IFT.

That said, experimental verification remains a significant challenge. Generalizing to NESS systems requires not only the excess work distribution but housekeeping heats as well. Indeed, these results suggest that carefully considering how to measure housekeeping dissipation is crucial to characterizing fluctuations in NESS systems. As Fig. 3.5 demonstrates, improper accounting leads in general to TCFT *violations* and, if the ESS FTs are used to estimate free energy differences, to potentially drastically mischaracterizing the system of interest.

3.5.2. Average excess energetics.

Despite the channels' distinct thermodynamic functioning as revealed by the TCFT, we can compare the channels' adaptive energetics via the excess works and heats. We begin by directly calculating the full trajectory averages, obtaining for discrete time:

$$-\langle \mathcal{Q}_{\text{ex}} \rangle = \sum_{n=0}^{N-1} \langle \mu(t_{n+1}) - \mu(t_n) | \phi_{\lambda_n} \rangle \text{ and} \tag{3.38}$$

$$\langle \mathcal{W}_{\text{ex}} \rangle = \sum_{n=0}^{N-1} \langle \mu(t_n) | \phi_{\lambda_{n+1}} - \phi_{\lambda_n} \rangle \,, \tag{3.39}$$

in agreement with Ref. [**42**]. As above, we set the initial distributions to the local stationary distribution for convenience. Armed with the discrete protocols, time steps, and starting distributions, we directly evaluate the mixed states (Eq. (3.25)) and steady-state distributions (Eq. (3.26)) for each time step. These are all that is needed to calculate $\langle \mathcal{Q}_{\text{ex}} \rangle$ and $\langle \mathcal{W}_{\text{ex}} \rangle$ via Eqs. (3.38) and (3.39).

Figs. 3.7 and 3.8 give the simulation results for excess heat.

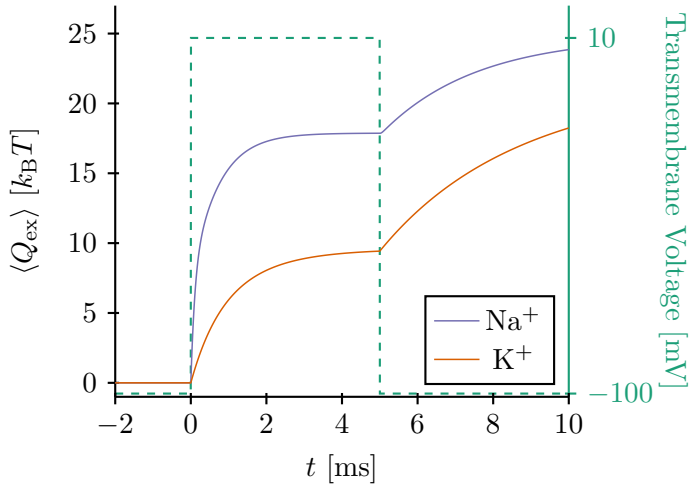

FIGURE 3.7. Excess heats (solid lines) for both channels under the pulse protocol (dashed line). The K^+ channel is less dissipative. Both expend energy as they relax to environmentally-induced steady states.

First, driven by the pulse, the K^+ channel dissipates less excess heat over the course of this protocol. Its rate of relaxation to steady state—corresponding to constant epochs in the protocol—appears slower than the Na^+ channel's on the jump from -100 to 10 mV, but faster on the subsequent drop back down to -100 mV.

The spike protocol paints a very different picture. Here, while the Na^+ channel still dissipates (in this case, significantly) more over the course of the protocol, it *also* appears to respond much more quickly to changes in the protocol than does the K^+ channel. A tradeoff appears: the cost of the Na^+ channel adapting more *quickly* to its environment is that it dissipates more in the process. This did not arise when driven by the pulse protocol. (Likely, this is due to that protocol operating outside of the "normal" voltage range for these channels—by dropping as low as -100 mV.)

Besides showcasing a detailed energetic comparison between different channels, the discrepancy between the pulse- and spike-driven behaviors demonstrate that *in vivo* thermodynamic response can qualitatively differ from that elicited by voltage-clamp experiments.

The corresponding results for excess work are given in Figs. 3.9 and 3.10. Unlike excess heat, the excess work is not sensitive to the timescales of each channel's relaxation to steady state. Instead, it tracks *environmental* entropy produced by the external drive. Yet it is still sensitive to the dynamics

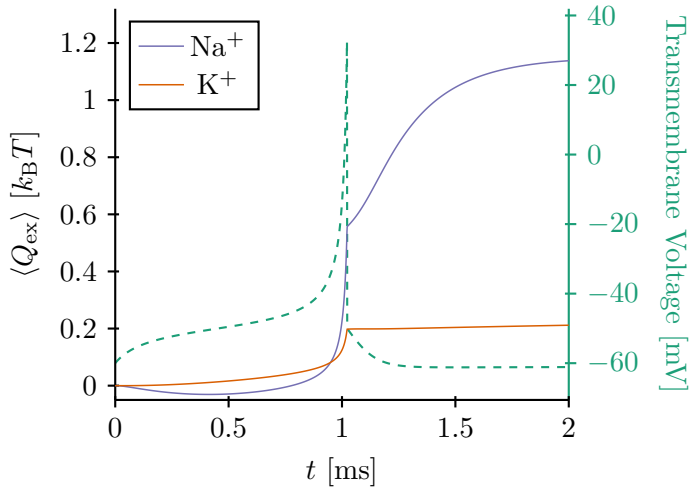

FIGURE 3.8. Excess heats (solid lines) for both channels driven by the spike protocol (dashed line). Under this more biologically realistic protocol, the Na^+ dissipates significantly more than the K^+ channel and does so responding much more rapidly to changes in membrane voltage. This suggests a tradeoff between the speed of the channel's response and its dissipation, one not necessarily present in the more artificial pulse protocol.

of the individual channel (per Eq. (3.8)), and this sensitivity is reflected in the thermodynamic responses.

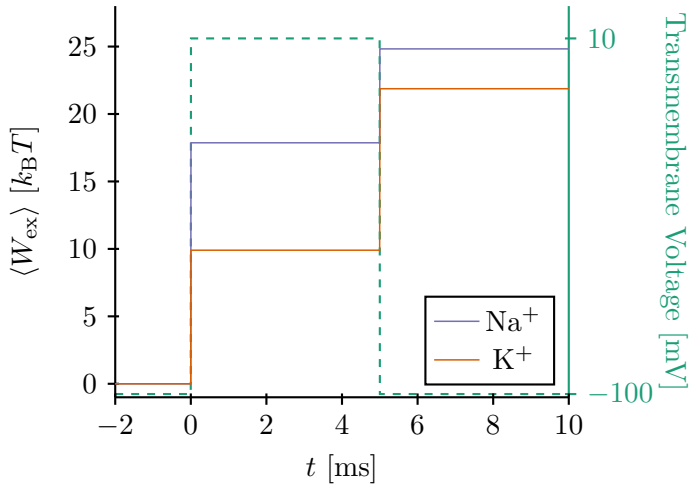

FIGURE 3.9. Excess works (solid lines) for both channels under the pulse protocol (dashed line). Unlike excess heat, excess work is only done upon *change* in the driving parameter. Thus, we see changes only at the pulse's rise and fall. Much more excess work is done on the Na^+ channel than on the K^+ during the rise of the pulse, but these roles are reversed on its fall. Over the entire protocol, the Na^+ channel produces (slightly) more excess environmental entropy.

As in the excess heat calculations, there is a difference in behavior between the pulse and spike protocols. In the former, the K^+ channel and Na^+ channels trade off under the rise and fall of the pulse. Driven by the spike protocol, however, the Na^+ channel induces more environmental entropy production across the board, though they track extremely closely until the peak and reset phases of the action potential spike. This reflects not only the larger potential for dissipation in the Na^+ channel under the spike protocol, but highlights where during the protocol most of the difference arises.

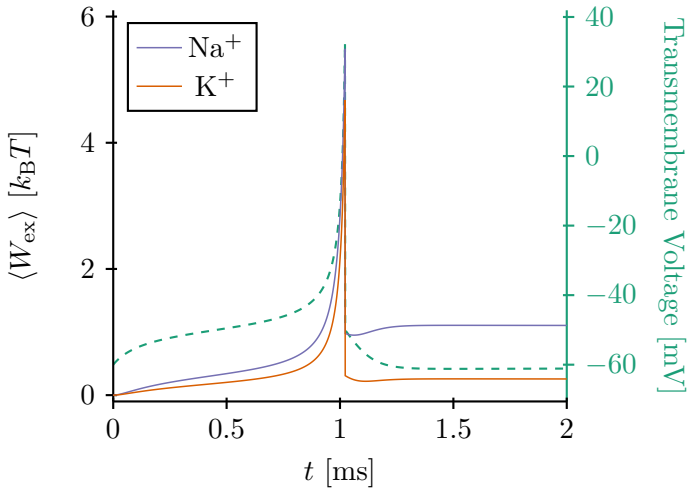

FIGURE 3.10. Excess works (solid lines) for both channels under the spike protocol (dashed line). We see that more excess work is done on the Na^+ channel across the board. This corresponds to larger environmental entropy production and indicates a greater potential for dissipated work in the Na^+ channel. However, we also see that up until the peak and reset phases of the action potential spike, they track very closely before diverging.

To reiterate, while excess heat is an energetic signature of relaxation to steady state, these calculations do not assume the system ever *reaches* such a steady state. While the pulse protocol allows each of the channels to do so, over the course of the spike we see a *dynamic* dissipation in the two channels—this is energy expended while attempting to reach an ever-evolving steady-state target. We close by highlighting that our theoretical developments enabled quantitatively comparing the channels' adaptive energetics under realistic environmental stimuli—captured in our Fig. 3.8—*despite* the departures in their underlying steady-state dissipation.

3.6. Conclusion

We reviewed and extended the techniques of stochastic thermodynamics, culminating in a TCFT for NESSs, and even, for nonthermal stochastic processes. Using these, we analyzed the adaptive and homeostatic energetic signatures of two neurobiological systems—systems key to propagating action potentials in mammalian neurons. Along the way, we developed a toolkit for probing the nonequilbrium thermodynamics in a broad range of mesoscopic complex systems that requires little in the way of restrictive assumptions.

Our results exposed a new quantitative structure in how systems appear to violate equilibrium steady-state assumptions, both warning against and elucidating the consequences of inappropriately assuming detailed-balanced dynamics. In this, they suggest a need for both our corrected and generalized TCFT *and* new experimental tools.

Specifically, for nonequilibrium steady-state systems, tracking housekeeping entropy production is *crucial* to extracting functionally relevant thermodynamics and observing an additional *kind* of second law dynamics. While experimental tests have verified Eq. (3.20), they do not allow for observing the housekeeping thermodynamics, which play an important and independent part in assessing a system's functionality.

Our simulations of the averaged excess energetics, in contrast, show how to compare specific aspects of a system's functionality—the adaptive energetics—despite what are in this case infinite differences in the steady state behavior. In essence, these tools allow us to "cleave off" the divergence in the Na^+ model's housekeeping heat and still compare the channels' adaptation to environmental drive on equal footing. Finally, the spike protocol simulations also identified what *would not* be observed in traditional patch-clamp experiments on ion channels, namely detailed differences in response to each segment of an action potential.

Taken together, our development and associated numerical experiments revealed a rich—and, indeed, necessary—set of tools with which to probe the nonequilibrium dynamics of mesoscopic complex systems.

CHAPTER 4

First and Second Laws of Information Processing by Nonequilibrium Dynamical States

Preprint posted as Ref. [**72**]: M. T. Semaan and J. P. Crutchfield, *arXiv:2211.05849*.

The averaged steady-state surprisal links a driven stochastic system's information processing to its nonequilibrium thermodynamic response. By explicitly accounting for the effects of nonequilibrium steady states, a decomposition of the surprisal results in an information processing first law that extends and tightens—to strict equalities—various information processing second laws. Applying stochastic thermodynamics' integral fluctuation theorems then shows that the decomposition reduces to the second laws under appropriate limits. In unifying them, the first law paves the way to identifying the mechanisms by which nonequilibrium steady-state systems extract work from information-bearing degrees of freedom. To illustrate, we analyze an autonomous Maxwellian information ratchet that tunably violates detailed balance in its effective dynamics. This demonstrates how the presence of nonequilibrium steady states qualitatively alters an information engine's allowed functionality.

4.1. Introduction

In 1861, Maxwell introduced a thought experiment in which a "very neat-fingered being" leveraged observations to control a system that violated the second law of thermodynamics [**9**]. A century later, attempting to resolve the paradox, Landauer quantitatively bounded the requisite thermodynamic resources for erasing a single bit of information in a physical information-bearing degree of freedom [**10**]. These results have since stimulated many explorations of the fundamental physics tying a system's thermodynamic behavior to its functioning as an information processor [**73**].

One particular line of inquiry focused on "autonomous Maxwellian ratchets". In this, a ratchet embedded in a thermal environment moves along an "information tape", interacting with a single bit

of information at a time. The information on the tape modifies the ratchet's statistical properties while the ratchet absorbs and dissipates energy [**74**]. Recent results introduced an "information processing second law" (IPSL) for such systems that bounds the asymptotic rate $\langle \dot{W} \rangle$ of extracted work [**46**]:

$$\beta \langle \dot{W} \rangle \leq \Delta h_\mu, \tag{4.1}$$

where $\Delta h_\mu = h'_\mu - h_\mu$, h'_μ is the Shannon entropy rate of the statistical process generating the output tape, h_μ is the same for the input tape, and β is the inverse temperature of the thermal environment. Notably, this bound corrected previous "single-symbol" relations by accounting explicitly for arbitrary-order temporal correlations in the input and output bit strings, and found directly that removing such correlation increases the capacity of the system to produce work—even though the ratchet interacts with only a single bit at once.

More recently, Ref. [**75**] developed a similar IPSL not for the asymptotic rate of extracted work from an infinite tape, but for the finite-time ensemble-averaged work extracted when operating on a finite tape:

$$\beta \langle W \rangle \leq \Delta H[Z], \tag{4.2}$$

where Z is the random variable associated with the joint space of the ratchet and tape, and $H[Z]$ is its Shannon entropy.

The following first derives a simple information-thermodynamic equality by considering the averaged steady-state surprisal of a general driven stochastic process:

$$\Delta H[Z] = \langle \mathcal{W}_{\text{ex}} \rangle - \langle \mathcal{Q}_{\text{ex}} \rangle - \Delta D_{\text{KL}}[Z \,\|\, \Lambda]. \tag{4.3}$$

Here Z is the random variable associated with a system's state, Λ that associated with an environmentally-induced steady state, and $\langle \mathcal{W}_{\text{ex}} \rangle$ and $\langle \mathcal{Q}_{\text{ex}} \rangle$ are the (unitless) average excess work and heat of nonequilibrium steady state thermodynamics, respectively [**28**,**42**]. The Kullback-Leibler divergence [**11**] $D_{\text{KL}}[Z \,\|\, \Lambda]$ monitors the difference in information between the system's state and the would-be steady state.

We refer to Eq. (4.3) as the *information processing First Law* (IPFL) since, beyond the obvious change in the joint system's information content, the lefthand side acts as a kernel that describes a ratchet's "information processing" (discussed in detail in Sec. 4.4). Additionally, the righthand side expresses a generalized first law used to define excess heat and work (discussed further in Secs. 4.2.1 and 4.2.3). In essence, Eq. (4.3) expresses a first law for the system's information content in the same way the original equilibrium first law does for a system's energy (non)conservation.

Subsequently we show how the IPFL together with stochastic thermodynamics' integral fluctuation theorems—particularly those presented in Refs. [**28**, **40**]—both generalize and tighten the two preceding asymptotic and finite-tape IPSLs. In particular, identifying the role of the average dissipated housekeeping heat $\langle \mathcal{Q}_{\text{hk}} \rangle$ and the divergence from final steady-state conditions $D_{\text{KL}}[Z_N \,\|\, \Lambda_N]$, it shows that for finite-tape systems:

$$\beta \langle W \rangle \leq \Delta H[Z] + D_{\text{KL}}[Z_N \,\|\, \Lambda_N] - \langle \mathcal{Q}_{\text{hk}} \rangle \,. \tag{4.4}$$

Compared to Eq. (4.2), this explicit accounting for the effects of nonequilibrium steady states (NESSs) and nonequilibrium *dynamical* (nonsteady) start and end configurations gives a strictly tighter bound for finite- and even-state systems.

In short: for NESS systems an increase in randomness—as measured by $\Delta H[Z]$—must additionally compensate for persistent housekeeping costs—as measured by $\langle \mathcal{Q}_{\text{hk}} \rangle$—to produce work.

Finally, we demonstrate that for infinite-tape, finite-ratchet systems the asymptotic bound is similarly tightened:

$$\beta \langle \dot{W} \rangle \leq h'_\mu - h_\mu - \langle \dot{\mathcal{Q}}_{\text{hk}} \rangle, \tag{4.5}$$

where $\langle \dot{\mathcal{Q}}_{\text{hk}} \rangle$ is the asymptotic rate of housekeeping dissipation.

To summarize, fluctuation theorems take the IPFL directly to a suite of simultaneously-true second laws for information. This, once again, mirrors informational generalization of the familiar equilibrium second law.

Besides clarifying and unifying derivations of these IPSLs, extending their domains to explicitly include the effects of initial- and final-state dependence as well as nonequilibrium steady states opens the door to considering detailed information-energy tradeoffs for systems that arbitrarily

violate detailed balance in their effective dynamics. We demonstrate this via an example ratchet designed to tunably violate detailed balance while remaining tractable for analysis. This uncovers qualitative corrections to a ratchet's ability to extract work.

The development proceeds as follows. First, Sec. 4.2 sets out the preliminary notation and introduces the relevant stochastic dynamical functionals. Section 4.2.3 maps the general stochastic dynamical picture to an explicitly thermodynamic one, immersed in a single-temperature environment. This then points toward concrete example realizations of the general stochastic theory. Section 4.2.4 reviews autonomous Maxwellian information ratchets, which comprise our example system class.

With the preliminaries in hand, Sec. 4.3 derives Eq. (4.3)'s IPFL. It applies the IPFL to the information ratchet picture, revealing strict equalities relating a ratchet's thermodynamic dissipation with its information processing in transforming an input tape to an output.

Section 4.4 then specializes the IPFL in two ways. First, Sec. 4.4.1 introduces and uses integral fluctuation theorems to take the equality to an inequality. This arrives at the kernel of previous IPSLs, explicitly generalizing and tightening that of Ref. [**75**]. Then, Section 4.4.2 considers the asymptotic rate limit of an infinite tape, similarly generalizing the previous asymptotic IPSL to include the effects of nonequilibrium dynamical state dependence and potentially infinite-state ratchets. The restriction to finite ratchets in Sec. 4.4.3 rounds out our derivations, revealing a simple correction tightening Ref. [**46**]'s asymptotic IPSL.

Finally, Sec. 4.5 applies the developed theory to a finite-state information ratchet that arbitrarily violates detailed balance and so exhibits NESSs. We find that even for simple cases, NESSs have dramatic effects on a ratchet's ability to extract work, qualitatively changing its landscape of allowed behaviors.

Taken together, our results (i) unify both previously-reported IPSLs for ratchets by deriving them explicitly from the underlying IPFL and integral fluctuation theorems, (ii) place the specific application of autonomous ratchet function in the broader context of the exchange between energy and information in complex systems, including generally nondetailed-balanced ratchets, and (iii) demonstrate severe restrictions NEDSs place on allowed ratchet functionality—restrictions critical to understanding the thermodynamics of information processing in complex systems.

4.2. Preliminaries

We consider a system under study (SUS) that stochastically realizes states z in a countable space $\mathcal{Z}$. It is driven in discrete time by a *protocol* written as a sequence of parameter values $\lambda \in \Lambda$, denoted by $\lambda_{0:N} \doteq \lambda_0 \lambda_1 \ldots \lambda_N$ for a positive integer N. The resulting driven stochastic process $Z_{0:N}$ is not stationary. However, we assume it is *conditionally stationary*: for any protocol indefinitely fixed at λ, there is a unique corresponding stationary state distribution $\boldsymbol{\pi}_\lambda$.

We place no further restrictions on our system. In particular, it need not have a particular dynamical structure—Markov, master equation, Langevin, detailed balanced, and so on. And, we make no particular claim about the scale of its state space. The protocol itself may be a realization of a separate stochastic process, and the state space may be a joint one with meaningfully decomposable parts. In point of fact, we treat both of these cases as examples later. First, though, we derive our main result in greater generality, requiring only the conditional stationarity assumption and involving only the functionals of trajectory-protocol pairs we now define.

4.2.1. Dynamical Functionals.

With a would-be stationary distribution $\boldsymbol{\pi}_\lambda$ associated to each driving parameter λ, denote its elements by $\pi_\lambda(z)$. Without loss of generality we define the *steady-state surprisal* [**28**, **40**, **42**, **42**, **76**]:

$$\phi_\lambda(z) \doteq -\ln \pi_\lambda(z), \tag{4.6}$$

so called as it is the Shannon self-information [**11**] of seeing state z under the distribution $\boldsymbol{\pi}_\lambda$.

For notational uniformity, we cast the sequence of stationary distributions during a protocol as a stochastic process over random variables $\Lambda_t \sim \boldsymbol{\pi}_{\lambda_t}$. Upon averaging we have:

$$\begin{aligned} \langle \boldsymbol{\pi}_\lambda | \boldsymbol{\phi}_\lambda \rangle &\doteq \sum_{z \in \mathcal{Z}} \pi_\lambda(z)\, \phi_\lambda(z) \\ &= H[\Lambda], \end{aligned} \tag{4.7}$$

the Shannon entropy of the distribution $\boldsymbol{\pi}_\lambda$.

In suitable settings ϕ_λ carries a thermodynamic meaning as well—the *nonequilibrium potential* [**28**, **42**]. For example, consider a canonical statistical mechanical system in contact with a single heat reservoir at inverse temperature $\beta \doteq 1/k_\mathrm{B}T$, with k_B Boltzmann's constant. Then $\boldsymbol{\pi}_\lambda$—an

equilibrium steady state, in this case—is Boltzmann distributed in the energies [**49**]:

$$\phi_\lambda(z) = -\beta(E_\lambda(z) - F_\lambda^{\text{eq}}), \tag{4.8}$$

with F_λ^{eq} the equilibrium free energy—the familiar logarithm of the canonical partition function. While under these circumstances ϕ_λ acts as a kind of generalized energy, we stress that it retains dynamical and information-theoretic meaning—via Eqs. (4.6) and (4.7)—even when a thermodynamic interpretation is inappropriate.

We carry this theme to our following three "special" functionals—stochastic dynamical generalizations of the excess heat, excess work, and housekeeping heat of nonequilibrium steady state thermodynamics [**28**, **33**, **40**, **42**, **52**]. That is, while their definitions are drawn from thermodynamic counterparts and while we ultimately focus on a thermodynamic setting, we stress that these functionals retain tractable, interpretable meaning well outside of those domains.

Excess work $\mathcal{W}_{\text{ex}}$ and *excess heat* $\mathcal{Q}_{\text{ex}}$ are distinct contributions to a system's change in steady-state surprisal:

$$\begin{aligned}\Delta\phi &= \underbrace{\Delta_\lambda\phi}_{} + \underbrace{\Delta_z\phi}_{} \\ &\doteq \mathcal{W}_{\text{ex}} - \mathcal{Q}_{\text{ex}},\end{aligned} \tag{4.9}$$

where, for N time steps:

$$\Delta_\lambda\phi \doteq \sum_{n=0}^{N-1} \phi_{\lambda_{n+1}}(z_n) - \phi_{\lambda_n}(z_n) \qquad \text{and} \tag{4.10}$$

$$\Delta_z\phi \doteq \sum_{n=0}^{N-1} \phi_{\lambda_{n+1}}(z_{n+1}) - \phi_{\lambda_{n+1}}(z_n). \tag{4.11}$$

That is, by excess work $\mathcal{W}_{\text{ex}}$ we refer to the change in steady-state surprisal owing to a changing environmental drive. And, by excess heat $\mathcal{Q}_{\text{ex}}$ we identify the change in steady-state surprisal owing to the system's state change—its response or *adaptation* to environmental conditions. This formulation agrees with standard ones in steady-state thermodynamics [**28**, **40**, **42**].

We generalize *total* heat $\mathcal{Q}$ in a manner consistent with microscopic reversibility [**38**]:

$$\mathcal{Q} \doteq \ln \frac{\Pr\left(Z_{1:N} = z_{1:N} \quad \mid Z_0 = z_0 \;\; ; \lambda_{0:N}\right)}{\Pr\left(Z_{1:N} = \tilde{z}_{N-1:0} \mid Z_0 = \tilde{z}_N \;\; ; \tilde{\lambda}_{N:0}\right)}, \tag{4.12}$$

where the tilde indicates negation of odd-parity variables, such as momentum and magnetic field.

Finally, we define the generalized *housekeeping heat*—heat dissipated to maintain nonequilibrium steady states—in terms of Eqs. (4.9), (4.12), and Ref. [**33**]'s phenomenological breakdown:

$$\mathcal{Q}_{\text{hk}} \doteq \mathcal{Q} - \mathcal{Q}_{\text{ex}}. \tag{4.13}$$

Stated as-is, this definition differs slightly from others in Refs. [**40**, **42**, **50**, **51**, **52**, **77**] in that it is not restricted to Markovian dynamics. It differs from that of Ref. [**28**] in that it allows for odd-parity state or protocol variables.

Since we allow odd-parity variables, it further decomposes in the manner of Refs. [**50**, **51**, **77**]. While we leave detailed discussion along these lines to a sequel, we reiterate here the delicacy in interpreting $\mathcal{Q}_{\text{hk}}$ directly as a measure of detailed balance violation—both in the presence of odd-parity variables as discussed by Refs. [**50**, **51**, **77**] and under non-Markovian dynamics as discussed by Ref. [**28**].

Let us be explicit. With Markov dynamics *part* of $\mathcal{Q}_{\text{hk}}$ corresponds directly to detailed-balance violation. With both Markov dynamics *and* strictly even-parity variables, $\mathcal{Q}_{\text{hk}}$ in its entirety measures detailed-balance violation. In the general case, the sources of housekeeping heat are more complex than these. However, its interpretation remains simple. It is that (generalized, unitless) heat dissipated atop the excess, as defined by Eq. (4.13).

4.2.2. Nonequilibrium Dynamical States.

In the special case where the system is Markovian (order 1) and subject to an indefinitely fixed drive—yielding a stationary Markov process—the rate of housekeeping heat is the asymptotic entropy production rate familiar to stochastic thermodynamics [**78**], sometimes taken on average to be a measure of the system's fundamental time-reversal asymmetry [**79**, **80**].

We wish to consider the more general case of systems that have not yet reached their (equilibrium or nonequilibrium) steady state distributions—processes that are not stationary. While

Eqs. (4.9)–(4.13) leverage a suite of "would-be" stationary distributions, they are defined for arbitrary paths, including when the system is nowhere near such a steady state at any stage of the observed interval. We call such transient state distributions $\boldsymbol{\mu}_t \nsim \boldsymbol{\pi}_{\lambda_t}$ "nonequilibrium dynamical states" (NEDSs). In treating system trajectories that begin and/or end in NEDSs, a final term appearing in our derivations and related results remains: the *nonsteady state free energy*:

$$\mathcal{F}^{\text{nss}}_{\boldsymbol{\mu}\|\lambda}(z) \doteq \ln \frac{\mu(z)}{\pi_\lambda(z)}. \tag{4.14}$$

Its name derives from two places. First, "nonsteady state" since it indicates how "far" a given distribution is from the associated steady-state distribution: on state averaging, we have $\left\langle \boldsymbol{\mu} \middle| \mathcal{F}^{\text{nss}}_{\boldsymbol{\mu}\|\lambda} \right\rangle = D_{\text{KL}}[\boldsymbol{\mu} \,\|\, \boldsymbol{\pi}_\lambda]$. Second, "free energy" since with an equilibrium steady state system in contact with a single ideal heat reservoir, it is the nonequilibrium part of the nonequilibrium free energy: $\mathcal{F}^{\text{neq}} = \beta F^{\text{eq}}_\lambda + \mathcal{F}^{\text{nss}}$.

Various naming schemes have appeared over nonequilibrium thermodynamics' history. For example, $\dot{\mathcal{Q}}_{\text{hk}}$ and its average are often referred to as "adiabatic" entropy production while $\dot{\mathcal{W}}_{\text{ex}}$ is the "nonadiabatic" component [**78**]. Excess work is alternatively (and equivalently) called "excess environmental entropy production" [**42**]. We restrict our vocabulary to that we explicitly introduced for clarity: housekeeping heat and excess work, respectively. We make a final note that on its own, excess heat is *not* an entropy production in the sense of stochastic thermodynamics. Yet it plays a central (and interpretable) role as a measure of "adaptive" dissipation both in that domain and in the general dynamical case.

4.2.3. Excess and Thermodynamic First Laws.

We call Eq. (4.9) our "excess first law" due to its structural similarity to the first law of thermodynamics. For an equilibrium steady-state system affixed to an ideal work reservoir parameterized by λ and an ideal heat bath at inverse temperature β, we assign to each state z an energy $E_\lambda(z)$ and obtain [**37**]:

$$\begin{aligned} \Delta E &= \underbrace{\Delta_\lambda E} + \underbrace{\Delta_z E} \\ &\doteq \; W \; - \; Q. \end{aligned} \tag{4.15}$$

Equation (4.15) defines work and heat in this restricted case as distinct contributions to the system's change in energy. Superficially, Eq. (4.9) is then a first law for steady-state surprisal in exactly the same way that Eq. (4.15) is a first law for energy.

The change of viewpoint from E to ϕ as the central object represents a subtle but useful generalization. It is one without additional risk. The same restrictions that give Eq. (4.15) also imply (i) Boltzmann-distributed steady states as in Eq. (4.8), (ii) $\mathcal{W}_{\rm ex} \to \beta(W - \Delta F^{\rm eq})$, and (iii) $\mathcal{Q}_{\rm ex} \to \beta Q$. Thus, in this case:

$$\begin{aligned} k_{\rm B}T\Delta\phi &= \Delta E - \Delta F^{\rm eq} \\ &= (W - \Delta F^{eq}) - Q. \end{aligned} \tag{4.16}$$

And, the mapping from energetic to surprisal-based first laws involves only the switch in viewpoint from total work to *excess* work as the more direct quantity. Here, it is that work done atop the change in equilibrium free energy. Since fluctuation theorems are phrased quite naturally in terms of functionals of ϕ and realized path probabilities [**28**, **40**, **42**], generalizing the first law of Eq. (4.9) is particularly amenable to interaction with them.

Treating ϕ as more fundamental than E carries utility beyond this convenience, however. There are many more general settings than the canonical ensemble. These include, for example, biological, active matter, and other NESS systems not Boltzmann-distributed in the energies at stationarity [**30**,**81**,**82**]. In these cases, in defining $\mathcal{W}_{\rm ex}$ stochastic-dynamically, Eq. (4.9) circumvents issues with appropriately defining nonequilibrium steady-state free energies [**28**, **42**]. Finally, in any situation where a relationship between E and ϕ can be derived, one can map Eq. (4.9) to Eq. (4.15) directly. Moreover, the former retains its meaning—and, as we shall show, utility—even when the latter is ill-defined.

Such is often the case in highly coarse-grained, effective state-space models of mesoscopic complex phenomena where, at best, one estimates bounds on "true" entropy production [**31**, **83**, **84**, **85**, **86**]. The coarse-grained dynamics themselves, however, may be directly observed. And in these cases, Eq. (4.9) holds exactly and remains interpretable *at the level of the observed phenomena*. This is reminiscent of several similarly-phrased fluctuation theorems; e.g., Ref. [**28**]'s NESS trajectory class fluctuation theorem.

4.2.4. Information Ratchets.

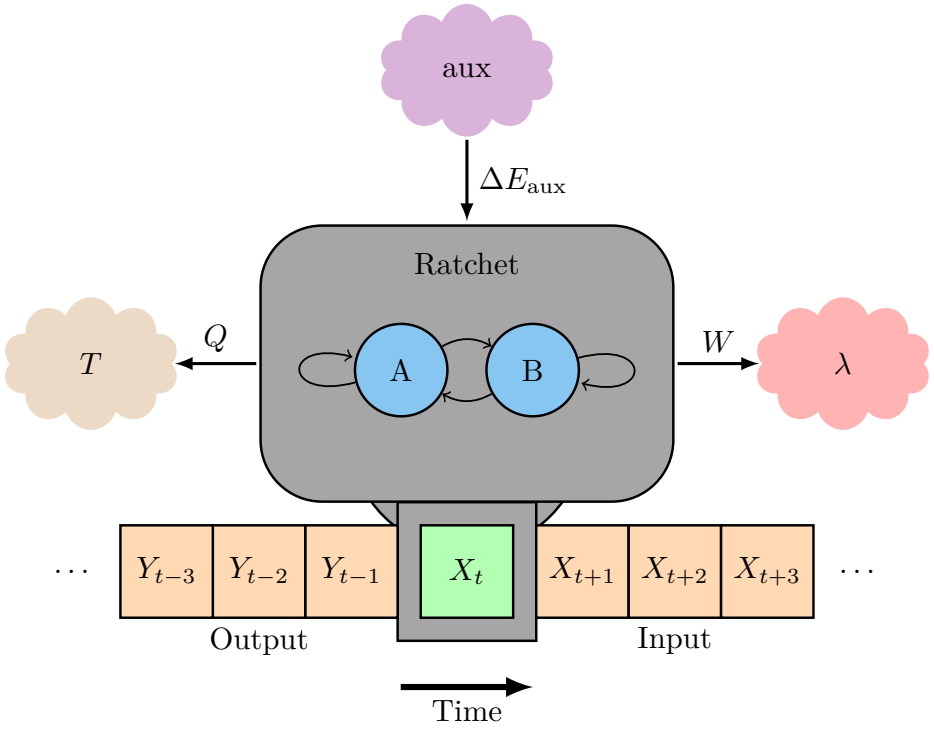

FIGURE 4.1. The information ratchet system. At each time step, the ratchet moves along the tape, interacting with one bit at a time, and exchanging energy with the coupled reservoirs in the process. New here is the auxiliary reservoir, which allows for nonequilibrium steady states and another mode of energy exchange with the ratchet-tape subsystem.

We are especially interested in a particular decomposition of $\mathcal{Z}$ into distinct subspaces—a *ratchet* and a semi-infinite *information tape*—embedded in an isothermal environment at inverse temperature $\beta = 1/k_{\text{B}}T$, with k_{B} Boltzmann's constant, and in contact with a work reservoir parameterized by λ and an auxiliary reservoir. The latter is introduced to capture unaccounted-for degrees of freedom and allows the ratchet to exhibit NESSs [**28**]. Figure 4.1 illustrates the setting [1].

The ratchet interacts directly with only a single cell at a time from the information tape. (Hereafter, we refer to a *bit* since we consider a tape with a binary alphabet. Generalizing to other alphabet sizes is straightforward.) We assume that the joint dynamics of the ratchet, interacting bit, and reservoirs is Markovian and independent of λ [2]. At each time step, the ratchet:

[1]Figure 4.1 was created in part by modifying Ref. [**87**]'s Fig. 1, with permission from the authors.

[2]This is not to say that λ has no effect on the system, but to say that the system's *dynamics* do not depend on it. That is, the Markov transition matrix is independent of λ.

(1) Moves one cell along the tape, putting it in contact with the next interaction bit; and

(2) Thermally interacts with the coupled reservoirs for a time τ.

Note that previous information ratchet schemes allowed the joint ratchet-tape system to fully thermalize in each interaction step [**46**, **48**, **74**, **75**, **88**]. This meant that the joint dynamics reached their stationary distribution before the next step. The following relaxes this timescale separation requirement, unless specified.

We do not assign an energy to each state in the joint dynamics, since detailed balance is not required. (In the previously-cited results detailed balance was used to calculate state energy differences.) Rather, we leverage the fact that in this isothermal setting $\mathcal{Q}, \mathcal{Q}_{\text{ex}}, \mathcal{Q}_{\text{hk}} \to \beta Q, \beta Q_{\text{ex}}, \beta Q_{\text{hk}}$. And, we require the joint ratchet-bit-work system possesses fixed average energy over the sum of both steps above. This is analogous to the canonical ensemble, but allows for the auxiliary reservoir. Thus, any exchange of heat is rectified by an exchange of work with the work reservoir (on average).

In effect, rather than directly calculating works via state-to-state relative energies—as determined by the detailed balance requirement—we can directly calculate heats via Eqs. (4.9)–(4.13) and the isothermal setting.

To describe and decompose the information-bearing degrees of freedom, we split the random variable Z_n in to three parts: the random variable R_n (with alphabet $\mathcal{R}_n$) corresponds to the ratchet subsystem's state at time n; the joint random variable $X_{n:\infty}$ to the *input tape* at time n—that portion of the information tape to which the ratchet has not yet written—and the joint random variable $Y_{0:n-1}$, for the *output tape*, to which the ratchet has written. Thus, at each n, $Z_n = R_n, X_{n:\infty}, Y_{0:n-1}$. This mirrors the decomposition of Refs. [**46**, **48**].

4.3. Information Processing First Law

Let $\boldsymbol{\mu}_n$ be the NEDS at time step n, such that $Z_n \sim \boldsymbol{\mu}_n$. Suppose we have a system that begins in $\boldsymbol{\mu}_0$ and ends in $\boldsymbol{\mu}_N$, as driven by the protocol $\lambda_{0:N}$. We begin by equating the trajectory (over all possible state-space trajectories $z_{0:N}$) and state averages (justified in App. B) of the steady-state

surprisal:

$$\begin{aligned}\langle \Delta\phi_\lambda \rangle &= \Delta \langle \boldsymbol{\mu} | \boldsymbol{\phi}_\lambda \rangle \\ &\doteq \langle \boldsymbol{\mu}_N | \boldsymbol{\phi}_{\lambda_N} \rangle - \langle \boldsymbol{\mu}_0 | \boldsymbol{\phi}_{\lambda_0} \rangle \,. \end{aligned} \tag{4.17}$$

The lefthand side is, by definition, $\langle \mathcal{W}_{\text{ex}} \rangle - \langle \mathcal{Q}_{\text{ex}} \rangle$. This is the averaged first law for ϕ as in Eq. (4.9).

For the righthand side, notice that:

$$\begin{aligned}\langle \boldsymbol{\mu} | \boldsymbol{\phi}_\lambda \rangle &= \sum_{z \in \mathcal{Z}} \mu(z) \phi_\lambda(z) \\ &= -\sum_{z \in \mathcal{Z}} \mu(z) \ln \pi_\lambda(z) \\ &= H[Z] + D_{\text{KL}}[Z \,\|\, \Lambda]. \end{aligned} \tag{4.18}$$

Hence, we have the *information processing first law* (IPFL):

$$\Delta H[Z] + \Delta D_{\text{KL}}[Z \,\|\, \Lambda] = \langle \mathcal{W}_{\text{ex}} \rangle - \langle \mathcal{Q}_{\text{ex}} \rangle \,. \tag{4.19}$$

The lefthand side accounts for the "information processing": the Shannon entropy change of the system plus the change in its divergence from the local stationary distribution. (Alternatively, the change in cross entropy between the system's state distribution and the local steady-state distribution.)

Moving $\Delta D_{\text{KL}}[Z \,\|\, \Lambda]$—the averaged nonsteady-state free energy from Eq. (4.14)—to the righthand side recovers Eq. (4.3). This is, quite directly, a first law for information that expresses its changes in terms of thermodynamic variables. This IPFL holds for transitions between NEDSs, implying validity for NESS and even nonthermal systems, since the generalized excess quantities are still well-defined by Eq. (4.9).

Stated in the form of Eq. (4.19), the IPFL makes no reference to the "conjugate" or "reversed" dynamics involved in the definitions of $\mathcal{Q}$ and by extension $\mathcal{Q}_{\text{hk}}$. Rather, it is concerned strictly with averages weighted by forward trajectories. However, substituting Eq. (4.13) does involve these

conjugated dynamics. This leads us to express the IPFL equivalently as:

$$\Delta H[Z] + \langle \mathcal{Q} \rangle = \langle \mathcal{W}_{\text{ex}} \rangle + \langle \mathcal{Q}_{\text{hk}} \rangle - \Delta D_{\text{KL}}[Z \,\|\, \Lambda]. \tag{4.20}$$

Here, the lefthand side is stochastic thermodynamics' familiar *total entropy production*. Now, though, it is broken into system ($\Delta H[Z]$) and environment ($\mathcal{Q}$) pieces [**42**,**43**,**89**]. The righthand side thus represents an alternative breakdown of the total entropy production into excess environmental ($\mathcal{W}_{\text{ex}}$) and housekeeping ($\mathcal{Q}_{\text{hk}}$) pieces, as well as one due to initial (and final) state dependence ($\Delta D_{\text{KL}}[Z \,\|\, \Lambda]$) [**28**, **42**, **64**, **90**]. Explicitly, the IPFL expresses a particular breakdown of the average total entropy production. This decomposition clearly links information content with thermodynamic processes that do *not* invoke conjugated dynamics.

4.3.1. Application to Information Ratchets.

Arriving at Eq. (4.19) required very minimal assumptions about the underlying SUS. Now, we wish to specialize it to the information ratchet system implied by Sec. 4.2.4 and Fig. 4.1. In particular, the isothermal environment takes our stochastic dynamical functionals to thermodynamic energies. And, the distinct ratchet and tape subspaces allow for meaningful decomposition of the information-bearing degrees of freedom.

First, we expand $\Delta H[Z]$ from Eq. (4.19). Splitting the joint Shannon entropies in terms of mutual informations:

$$\begin{aligned} \Delta H[Z] &= \Delta H[R] + \Delta H[X, Y] - \Delta I[R : X, Y] \\ &= \Delta H[R] + \Delta H[X] + \Delta H[Y] - \Delta I[R : X, Y] - \Delta I[X : Y]. \end{aligned} \tag{4.21}$$

This further decomposes the IPFL of Eq. (4.19):

$$\begin{aligned} &\langle \mathcal{W}_{\text{ex}} \rangle - \langle \mathcal{Q}_{\text{ex}} \rangle = \\ &\quad \Delta H[R] + \Delta H[X] + H[Y_{0:n-1}] - \Delta I[R : X, Y] - I[X_{n:\infty} : Y_{0:n-1}] + \Delta D_{\text{KL}}[Z \,\|\, \Lambda]. \end{aligned} \tag{4.22}$$

Equivalently, the breakdown of average total entropy production in Eq. (4.20) becomes:

$$\langle \mathcal{W}_{\text{ex}} \rangle + \langle \mathcal{Q}_{\text{hk}} \rangle - \Delta D_{\text{KL}}[Z \,\|\, \Lambda] =$$
$$\mathcal{Q} + \Delta H[R] + \Delta H[X] + H[Y_{0:n-1}] - \Delta I[R : X, Y] - I[X_{n:\infty} : Y_{0:n-1}]. \tag{4.23}$$

Equation (4.21)'s decomposition took Eqs. (4.19) and (4.20) to Eqs. (4.22) and (4.23), respectively. The decomposition is identical to that in Ref. [**46**]. However, there the goal was to take asymptotic rates. That, together with the finite-state ratchet requirement, removed several terms. Here, we pause to interpret each term in its finite-time context and comment on its contribution to the averaged total entropy production.

The first term $\Delta H[R]$ monitors the change in information content of the ratchet's states—a change in the ratchet's internal memory. If, as the ratchet interacts with the tape, it "gains memory" in this sense, the joint system as a whole must become "more randomized". And, equivalently, the total averaged entropy production increases.

The second term $\Delta H[X] = H[X_{N:\infty}] - H[X_{0:\infty}]$ quantifies a change in the information content of the input tape. Or, more specifically for finite alphabets, this is strictly nonpositive—the opposite of the information "added" by the random variables $X_{0:N-1}$. And so, the more random the input tape, the more negative this term can be. We expect memoryless inputs to reduce the potential to extract work compared to memoryful ones. There is, colloquially, less pattern to scramble [**48**]. We shall see later that this is indeed the case for IPSLs. For the IPFL, in the meantime, negativity of this term acts to reduce averaged total entropy production. Intuitively, removing randomness in the input tape reduces overall entropy production.

The third term $\Delta H[Y]$ is simply the information content of the output tape. Its impact with countable spaces—that we assumed—is straightforward. Due to Shannon information's nonnegativity, the more random the ratchet has made the output tape, the greater the positive contribution to average total entropy production.

The fourth term $\Delta I[R : X, Y]$ tracks the change in *shared* information between the ratchet and tape. As the ratchet interacts with bits from the input tape and writes to the output tape, it induces correlation between it and the tape. While at first glance this recalls Ref. [**48**]'s (de)randomizer axes, it is an altogether different term: they tracked induced correlations *internal* to the information

tape; this term tracks induced correlation *between the ratchet and the (entire input-output) tape*. Mutual information's positivity makes the contribution to Eq. (4.21) strictly nonpositive, lowering the average total entropy production. In other words, inducing correlation between the ratchet and tape reduces the joint system's entropy production and vice-versa.

Finally, the fifth term $I[X:Y] = I[X_{N:\infty} : Y_{0:N-1}]$ is the mutual information between the input and output tapes. And so, again due to mutual information's nonnegativity, it has the effect of increasing the averaged total entropy production.

Taken altogether, Eqs. (4.22) and (4.23) delineate exact links between finite-time ratchet-tape information processing and the joint system's thermodynamic behavior. It is a specialization of the IPFL to the case of a system constructed as in Fig. 4.1. Shortly, we use it as a starting point to derive and generalize the previously-reported IPSLs for ratchet-tape systems [**46**, **75**]. However, the inequalities in IPSLs are replaced by equalities of the IPFL in the same way that fluctuation theorems of stochastic thermodynamics replace inequalities of thermodynamic second laws with strict equalities. In point of fact, as we now show, fluctuation theorems directly take the IPFL to IPSLs.

4.4. Ratchet First to Second Laws

The following derives several specializations to the information ratchet system class of Sec. 4.2.4 and Fig. 4.1, starting from Eqs. (4.19) and (4.20). To do this, it first leverages an integral fluctuation relation to take the equality to an inequality. It then splits the effective state space as in Sec. 4.2.4, along the way generalizing a recently-reported finite-time IPSL [**75**]. Finally, it takes asymptotic limits to similarly modify the previous asymptotic IPSL [**46**, **48**, **88**] for these regimes. Since we adopt the same assumptions as Sec. 4.2.4, hereafter the underlying dynamics of the joint ratchet-tape space are Markovian. However, the statistical process that produces input and output tape symbol sequences need not be Markovian. In the infinite tape case, they may even possess infinite-order temporal correlations.

4.4.1. Fluctuations and Second Laws.

A crowning achievement of stochastic thermodynamics over the last several decades was to tame arbitrarily far-from-equilibrium fluctuations. The results were collected as "fluctuation relations" or

"fluctuation theorems". (See Ref. [**43**] for a recent review.) They come in three main flavors: (i) *integral*, relating to exponential averages over all possible trajectories [**36**, **37**, **40**, **41**]; (ii) *detailed*, exposing a time reversal (a)symmetry between forward and reverse paths [**38**, **39**, **42**, **61**, **62**]; and *trajectory class*, intermediating between the two [**28**, **89**, **91**].

We will not attempt a comprehensive review here. Besides noting those just listed, see also Refs. [**52**, **92**]. Nonetheless, we will use, in particular, two integral fluctuation theorems:

$$1 = \left\langle \mathrm{e}^{-\left(\mathcal{W}_{\mathrm{ex}} + \mathcal{Q}_{\mathrm{hk}} - \Delta \mathcal{F}^{\mathrm{nss}}_{\boldsymbol{\mu}\|\lambda}\right)} \right\rangle \qquad \text{and} \tag{4.24}$$

$$1 = \left\langle \mathrm{e}^{-\left(\mathcal{W}_{\mathrm{ex}} + \mathcal{F}^{\mathrm{nss}}_{\boldsymbol{\mu}_0\|\lambda_0}\right)} \right\rangle . \tag{4.25}$$

Asserting the convexity of the exponential, we apply Jensen's inequality to derive the *generalized second laws*:

$$\Delta H[Z] + \langle \mathcal{Q} \rangle \geq 0 \qquad \text{and} \tag{4.26}$$

$$\langle \mathcal{W}_{\mathrm{ex}} \rangle + D_{\mathrm{KL}}[Z_0 \,\|\, \Lambda_0] \geq 0. \tag{4.27}$$

Equation (4.26) is thus a consequence of the "total entropy production" integral fluctuation theorem. Equation (4.25), first introduced in Ref. [**28**], generalizes the integral fluctuation theorem of Hatano and Sasa [**40**] to include initial-state dependence. The resulting inequality in Eq. (4.27) shows that initially-nonsteady states *lower* the bound on $\mathcal{W}_{\mathrm{ex}}$.

As we now demonstrate, various IPSLs result directly from applying these integral fluctuation theorems and Jensen's inequality to the IPFL.

First, substitute Eq. (4.26) into Eq. (4.20). This gives, directly:

$$\langle \mathcal{Q} \rangle \geq -\Delta H[Z]. \tag{4.28}$$

In an appropriate thermal environment, such as that of Fig. 4.1, this lower bounds the finite-time dissipated heat and it is Ref. [**46**]'s Eq. (A7). In the ratchet setting, we may also rephrase this

bound in terms of the averaged work $\beta \langle W \rangle$ done *by* the system [3]:

$$\beta \langle W \rangle \leq \Delta H[Z]. \tag{4.29}$$

This is exactly Ref. [**75**]'s finite-tape IPSL, where Z denotes the joint random variable of their ratchet and tape subspaces.

However, we can also substitute Eq. (4.25) into Eq. (4.19), yielding:

$$\langle \mathcal{Q}_{\text{ex}} \rangle \geq -\Delta H[Z] - D_{\text{KL}}[Z_N \,\|\, \Lambda_N] \tag{4.30}$$

or, equivalently via Eq. (4.13):

$$\langle \mathcal{Q} \rangle \geq \langle \mathcal{Q}_{\text{hk}} \rangle - \Delta H[Z] - D_{\text{KL}}[Z_N \,\|\, \Lambda_N]. \tag{4.31}$$

This adjusts the finite-time dissipated heat to account for NEDSs. First, for finite spaces, nonsteady final states lower the bound on dissipated heat. That is, we need not dissipate to full relaxation. The presence of NESSs instead *raises* the bound by the amount of the total housekeeping heat. As long as $\mathcal{Q}_{\text{hk}} \geq D_{\text{KL}}[Z_N \,\|\, \Lambda_N]$, the result of an established IFT in even state spaces [**28**], this is always a tighter bound than Eq. (4.28).

Rephrasing in terms of work:

$$\beta \langle W \rangle \leq \Delta H[Z] + D_{\text{KL}}[Z_N \,\|\, \Lambda_N] - \langle \mathcal{Q}_{\text{hk}} \rangle . \tag{4.32}$$

This explicates the role of NEDSs on Ref. [**75**]'s Eq. (19) and establishes it under very general conditions. We see two additional effects. The ensemble-averaged housekeeping heat $\mathcal{Q}_{\text{hk}}$ *lowers* the bound as an additional source of dissipation, while the final-state dependence $D_{\text{KL}}[Z_N \,\|\, \Lambda_N]$ raises it. That is, we need not account for what *would be* dissipation if the system fully relaxed to its steady states [**90**]. As before, this bound is always tighter than Eq. (4.29) in the case of even state spaces.

Finally, applying the preceding decomposition of $\Delta H[Z]$ to Eq. (4.32) gives the analogue to Ref. [**75**]'s finite-tape IPSL further decomposed to both account for NEDSs and delineate ratchet-tape

[3] That is, $\beta \langle \Delta E \rangle = 0 = -\beta \langle W \rangle - \langle \mathcal{Q} \rangle$. The signs here indicate that heat flowing from the bath into the ratchet is done by the ratchet on its nonthermal surroundings.

information dynamics:

$$\beta \langle W \rangle \leq D_{\mathrm{KL}}[Z_N \,\|\, \Lambda_N] - \langle \mathcal{Q}_{\mathrm{hk}} \rangle + \Delta H[R] + \Delta H[X] + H[Y_{0:n-1}] - \Delta I[R : X, Y] - I[X_{N:\infty} : Y_{0:N-1}]. \quad (4.33)$$

With this, we can translate what each term of $\Delta H[Z]$ did to the averaged total entropy production to its effect on the maximum extracted work. In short, information processing that reduces the averaged total entropy production identically reduces the upper bound on ensemble-averaged work production: we must produce entropy to generate work.

4.4.2. General Asymptotics.

The preceding results apply for all finite times or, equivalently, for finite tapes. Now, we address asymptotics. Our procedure is to take $N \to \infty$ and divide the quantities of interest by N, giving an asymptotic rate per time step. For notational simplicity, we use the dot notation for the thermodynamic quantities:

$$\langle \dot{\mathcal{Q}} \rangle \doteq \lim_{N \to \infty} \frac{1}{N} \langle \mathcal{Q} \rangle . \quad (4.34)$$

And so on, for $\langle \dot{\mathcal{W}}_{\mathrm{ex}} \rangle$, $\langle \dot{\mathcal{Q}}_{\mathrm{ex}} \rangle$, $\langle \dot{\mathcal{Q}}_{\mathrm{hk}} \rangle$, and $\langle \dot{W} \rangle$. We now take the asymptotic limit of Eq. (4.22). In particular, as in Ref. [**46**] we have (i) $\lim_{N \to \infty} \Delta H[X]/N = -h_\mu$, (minus) the Shannon entropy rate of the process generating the input tape; (ii) $\lim_{N \to \infty} H[Y_{0:N-1}]/N = h'_\mu$, the Shannon entropy rate of the process generating the output tape; and (iii) $\lim_{N \to \infty} I[X_{N:\infty} : Y_{0:N-1}]/N = 0$.

The remaining two pieces of $\Delta H[Z]$, however, vanish only under restricting to finite-state ratchets. Without that assumption we are left with an asymptotic IPFL:

$$\langle \dot{\mathcal{W}}_{\mathrm{ex}} \rangle - \langle \dot{\mathcal{Q}}_{\mathrm{ex}} \rangle = \Delta h_\mu + \lim_{N \to \infty} \frac{1}{N} (\Delta H[R] - \Delta I[R : X, Y]) + \lim_{N \to \infty} \frac{1}{N} \Delta D_{\mathrm{KL}}[Z_N \,\|\, \Lambda_N], \quad (4.35)$$

and similarly write the asymptotic IPSL:

$$\begin{aligned}\beta\langle\dot{W}\rangle \leq{}& \Delta h_\mu - \langle\dot{\mathcal{Q}}_{\mathrm{hk}}\rangle \\ &+ \lim_{N\to\infty}\frac{1}{N}(\Delta H[R] - \Delta I[R:X,Y]) \\ &+ \lim_{N\to\infty}\frac{1}{N}D_{\mathrm{KL}}[Z_N \,\|\, \Lambda_N].\end{aligned} \tag{4.36}$$

This generalizes the previous bound by accounting explicitly for final-state dependence, nonequilibrium steady states, and potentially infinite-state ratchets.

We leave detailed analytical consideration of the limits for infinite-state ratchets and their con/divergence to a sequel. However, we will interpret the contextual meaning of the remaining limits for countably infinite ratchets.

First, $\lim_{N\to\infty}\Delta H[R]/N$ is the rate of change of the ratchet's *statistical complexity* $C_\mu[R]$ per time step, lower bounded by the statistical complexity of its ϵ-machine representation from computational mechanics [**4**, **19**]. In essence, this limit measures the rate of increase in ratchet memory as it reads an infinite stream of incoming bits. It is only nonzero for a ratchet with an infinite memory capacity. The resulting device is able to violate the finite-state asymptotic IPSL by leveraging its infinite internal memory to produce work in excess of that bound [**88**]. For any finite-state ratchet, since in that case $H[R_N]$ is bounded from above, $\lim_{N\to\infty}\Delta H[R]/N$ vanishes. It vanishes also for any infinite-state ratchet that does not asymptotically *gain* memory from an infinite stream of inputs. More precisely, this holds for a ratchet whose internal state distribution approaches a fixed value unaffected by the incoming bit stream.

Second, $\lim_{N\to\infty}\Delta I[R:X,Y]/N$ is the rate of change of correlation between the ratchet and the total information tape. This limit evaluates nonzero only if (i) the ratchet continually gains memory as above and (ii) the ratchet continually induces correlation between itself and the total input-output tape.

Finally, $\lim_{N\to\infty}\Delta D_{\mathrm{KL}}[Z \,\|\, \Lambda]/N$ is a kind of asymptotic movement away from steady-state conditions. Specifically, it can be nonzero only when (i) $D_{\mathrm{KL}}[Z_N \,\|\, \Lambda_N]$ is infinite for any N—e.g., the system is "infinitely far" from stationarity—and (ii) after each time step the joint ratchet-tape system's distribution gets progressively farther from the underlying stationary distribution.

Colloquially, the interaction timescale is so short that one can asymptotically take the ratchet-tape system further away from thermalization by constantly changing the interaction bit. The presence of this term at all implies the existence of a stationary distribution for the ratchet-tape system. This is a fact not guaranteed for infinite ratchets [**88**, **93**], but assumed by our stochastic (thermo)dynamical formalism.

4.4.3. Finite Ratchet Asymptotics.

The assumption of a finite-state ratchet—in line with potential physical implementation—simplifies the asymptotic analysis. (As it did in Refs. [**46**, **88**].) This results in an asymptotic IPFL for fnite-state systems:

$$\langle \dot{\mathcal{W}}_{\mathrm{ex}} \rangle - \langle \dot{\mathcal{Q}}_{\mathrm{ex}} \rangle = \Delta h_{\mu} . \tag{4.37}$$

And, finally, we have our correction to the previously-reported IPSL for finite ratchets interacting with an infinite tape:

$$\beta \langle \dot{W} \rangle \leq \Delta h_{\mu} - \langle \dot{\mathcal{Q}}_{\mathrm{hk}} \rangle . \tag{4.38}$$

The correction is simply $\langle \dot{\mathcal{Q}}_{\mathrm{hk}} \rangle$. For for even state spaces this is nonnegative and so tightens the previous bound. Said simply, housekeeping dissipation takes away from the maximum extracted work: One cannot harness what must go towards maintaining NESSs.

4.5. Asymmetric Stochastic 4-Cycle

The presence of housekeeping dissipation in Eq. (4.38) suggests meaningful change in ratchet functional thermodynamics [**94**], depending on the degree to which the joint ratchet-bit system violates detailed balance. To demonstrate this dependence we introduce a two-state ratchet coupled to the information tape called the "asymmetric stochastic 4-cycle" (AS4C). The states are labeled A and B. With joint ratchet-bit Markov chain states ordered by $(\mathrm{A} \otimes 0, \mathrm{A} \otimes 1, \mathrm{B} \otimes 1, \mathrm{B} \otimes 0)$, the

dynamics obey the row-stochastic transition matrix:

$$\mathbf{T}(p,q) \doteq \begin{bmatrix} 0 & p & 0 & 1-p \\ 1-p & 0 & p & 0 \\ 0 & 1-qp & 0 & qp \\ p & 0 & 1-p & 0 \end{bmatrix}. \tag{4.39}$$

This two-parameter ratchet family, pictured in Fig. 4.2, generically violates detailed balance and allows $0 \to 1$ and $1 \to 0$ transitions to be unequally favored in terms of transition probabilities. The latter fact manifests as a "rotational" asymmetry in the cycle—given by the scaling parameter $q \in (0, 1]$. When $q = 1$, the cycle is symmetric: $0 \to 1$ and $1 \to 0$ transitions are equally favored, but the system exhibits stationary directionality in its joint state space. In the symmetric case, detailed balance is achieved only when $p = \frac{1}{2}$.

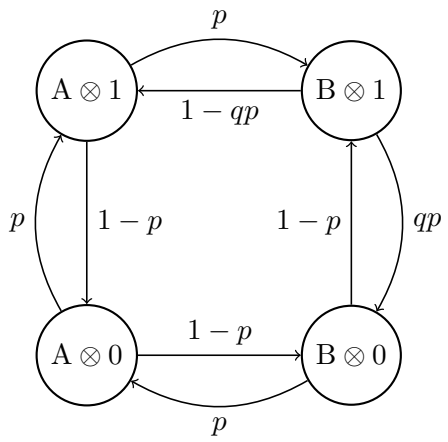

FIGURE 4.2. Markov chain describing the joint ratchet-bit dynamics of the asymmetric stochastic 4-cycle (AS4C) ratchet family. The behavior is parameterized by $p \in (0, 1)$ and $q \in (0, 1]$. The ratchet generically violates detailed balance. When $q \neq 1$ it controls rotational asymmetry and can probabilistically favor either $0 \to 1$ or $1 \to 0$ transitions. The symmetric value $q = 1$ equally favors these transitions, but the dynamics of the joint space still exhibits directionality. Detailed balance is satisfied in this case only for $p = 1/2$.

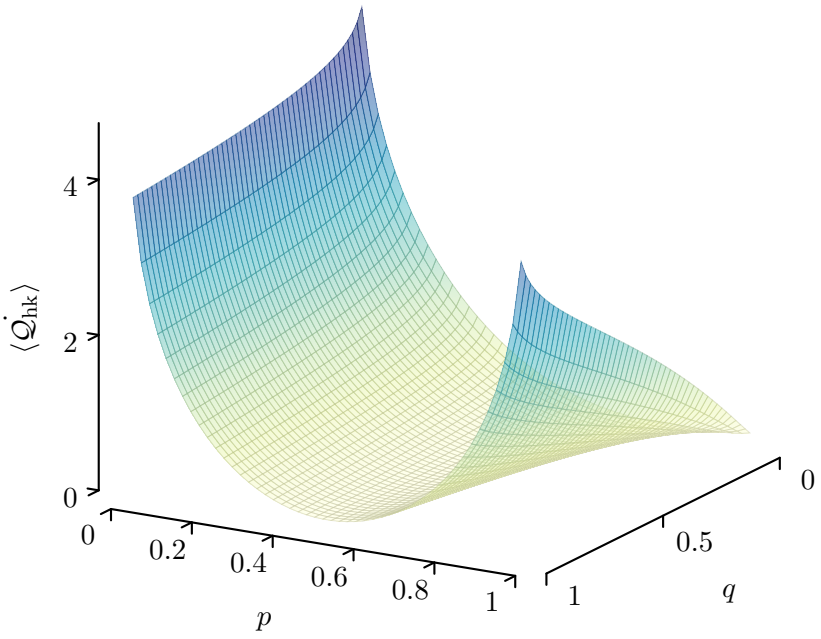

FIGURE 4.3. Averaged rate of housekeeping entropy production $\langle \dot{\mathcal{Q}}_{\rm hk} \rangle$ for the asymmetric stochastic 4-cycle, as a function of parameters p and q. This is the exact amount by which Eq. (4.38) tightens Eq. (4.1).

The extent to which a discrete- and even-state Markov chain system violates detailed balance on average is given by $\langle \dot{\mathcal{Q}}_{\rm hk} \rangle$. Via a single-step average of Eq. (4.13) we thus obtain for discrete time:

$$\langle \dot{\mathcal{Q}}_{\rm hk} \rangle = \sum_{i \neq j} \pi(i)[\mathbf{T}]_{ij} \log \frac{\pi(i)[\mathbf{T}]_{ij}}{\pi(j)[\mathbf{T}]_{ji}}, \tag{4.40}$$

where i and j index the states.

This is also the exact amount by which the previous asymptotic ESS IPSL Eq. (4.1) was tightened by our NESS IPSL in Eq. (4.38). To visualize this difference—the degree of tightening—Fig. 4.3 plots $\langle \dot{\mathcal{Q}}_{\rm hk} \rangle$ while sweeping parameters p and q.

4.5.1. Input-Output Transducer.

As one sees, $\langle \dot{\mathcal{Q}}_{\rm hk} \rangle$ is far from zero over a wide range of the parameter space. These are energies that must be dissipated to maintain the system's NESS character. And so, one expects, they have significant impacts on the system's ability to extract work from an information reservoir. This is to say, with our correction to the ESS IPSL in hand, we can analyze bounds on the functional thermodynamics of this ratchet family.

To do so, we must calculate the remaining terms in Eq. (4.38), namely the Shannon entropy rates h_μ and h'_μ of the processes generating the input and output state sequences. Following Refs. [**46**, **48**], we achieve this by first translating our 4-state joint ratchet-bit Markov chain into a 2-state ratchet *transducer* that accepts as input the process generating the input symbol statistics—in the form of a hidden Markov chain (HMC)—and produces as output the HMC generating the output symbol statistics.

A transducer is specified by its input-output-labeled matrices $\mathbf{M}^{(\text{out}|\text{in})}$:

$$\mathbf{M}^{(\text{out}|\text{in})} \doteq \mathbb{P}_{\text{in}}^{\mathsf{T}} \, \mathbf{T} \, \mathbb{P}_{\text{out}}. \tag{4.41}$$

The AS4C ratchet has two projection matrices $\mathbb{P}_0$ and $\mathbb{P}_1$ given by:

$$\mathbb{P}_0 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad \mathbb{P}_1 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}. \tag{4.42}$$

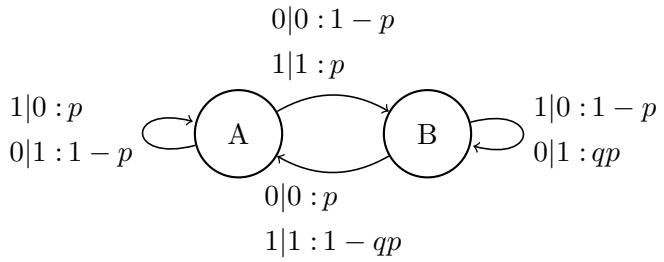

FIGURE 4.4. The transducer corresponding to the 4-state joint ratchet-bit Markov chain of the asymmetric stochastic 4-cycle (AS4C). The transducer has two internal states, A and B, and accepts 0 or 1 as an input bit symbol. The edges are labeled output bit|input bit : probability.

This defines the AS4C's transducer, whose state-transition diagram is visualized in Fig. 4.4. Now, we compose it with any input HMC—specified by its symbol-labeled transition matrices $\mathbf{U}^{(x)}$—to give the output HMC producing the symbol statistics on the output tape, specified by $\mathbf{V}^{(y)}$ [**13**, **48**]. The output HMC state space is the Cartesian product of the state spaces of the input HMC and the transducer. Let i, j index the states of the input HMC and i', j' index those of the

transducer. Then:

$$V^{(y)}_{i\times i' \to j \times j'} = \sum_x M^{(y|x)}_{i'j'} U^{(x)}_{ij}. \tag{4.43}$$

4.5.2. All-1s Driving.

To simplify determining h'_μ, we drive the AS4C transducer with the *all-ones process*: an input tape of all 1s, exhibiting no randomness whatsoever. Note that generically the output HMC of a memoryless ratchet driven by a memoryless input process results in a highly nonunifilar output HMC, for which determining the entropy rate is very challenging [**48**]. However, for all-1s driving, the AS4C produces the unifilar output HMC shown in Fig. 4.5.

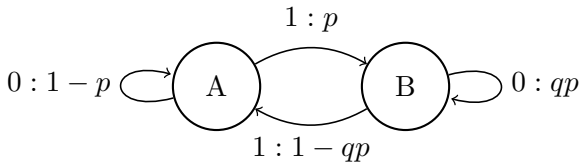

FIGURE 4.5. Output HMC given by the AS4C transducer acting on the all-1s input tape. The ratchet in this case scrambles an informationless input, thereby introducing the capacity to do work.

Since this HMC is unifilar—an internal state and an output symbol completely determine the next internal state—and since its two states make probabilistically distinct future predictions, it is a finite-state ϵ-machine of computational mechanics [**4**]. That the output tape's process can be described this way enables direct calculation of the output entropy rate [**12**]. Letting i and j index the output HMC's internal states, $\boldsymbol{\pi}$ be its stationary distribution, and y an output symbol, one has:

$$h'_\mu = -\sum_{y,i,j} \pi(i) V^{(y)}_{ij} \log V^{(y)}_{ij}. \tag{4.44}$$

Figure 4.6 plots this over the parameter space.

Setting the input process to have zero randomness also sets $h_\mu = 0$ for it: $\Delta h_\mu = h'_\mu$, all intrinsic randomness in the output tape is induced by the ratchet and, therefore, is available as a thermodynamic resource for work extraction. For the case of totally ordered input, Eq. (4.38) reads:

$$\beta \langle \dot{W} \rangle \leq h'_\mu - \langle \dot{\mathcal{Q}}_{\text{hk}} \rangle. \tag{4.45}$$

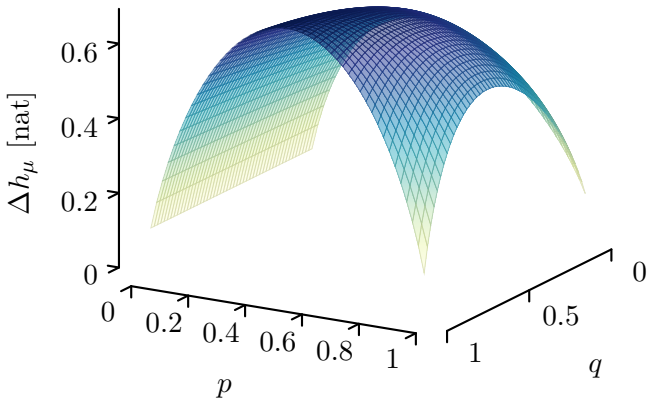

FIGURE 4.6. Change in Shannon entropy rate Δh_μ of the information tape, as generated by the AS4C ratchet driven by the all-1s process. Since the input process is entirely ordered with $h_\mu = 0$, this represents the ESS IPSL's maximum upper bound $\beta\langle\dot{W}\rangle_{\max}$. And this, in turn, precludes eraser functionality. That is, one cannot erase information that was never there.

Summarizing the requirements for work extraction: the ratchet must, at minimum, induce randomness in the output tape faster than it dissipates entropy to maintain its NESS.

Since with this particular driving $\Delta h_\mu > 0$ for all parameter combinations, the information eraser functionality of Refs. [**46**, **48**] is precluded. Instead, we have either an engine ($W > 0$) or a dud ($W \leq 0$). Most importantly, the presence of $\langle\dot{\mathcal{Q}}_{\text{hk}}\rangle$ here restricts the regions of parameter space where the ratchet *can* function as an engine. Or, alternatively, for some p and q the housekeeping costs are higher than the ratchet's ability to compensate by scrambling the information tape. This forces the previous "potential engine" regions into dud ones.

This is indeed the case, as Fig. 4.7 shows. In fact, only a small part of the "engine" functionality remains within bounds. Figure 4.8 shows this directly, where only "potential engine" regions of parameter space are colored. Since an entirely ordered input drives the ratchet, without accounting for the NESS correction one would expect all parameter space to allow potential work extraction. In this way, explicitly accounting for a system's NESS nature enables qualitative (and quantitative) correction to its allowed behaviors.

4.6. Conclusion

We began by deriving, under very general circumstances, an IPFL that connects ensemble-averaged thermodynamic behavior to a system's information processing via a strict equality. We

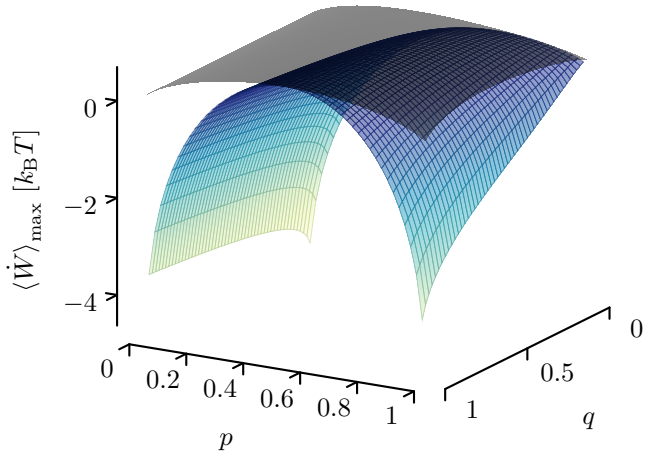

FIGURE 4.7. NESS-tightened upper bound on work extraction $\beta\langle \dot{W} \rangle_{\max} = \Delta h_\mu - \langle \dot{\mathcal{Q}}_{\text{hk}} \rangle$ for the AS4C driven by the all-1s process. To the extent that this differs from Fig. 4.6, superimposed here in grey, it represents a change of maximum possible work extraction.

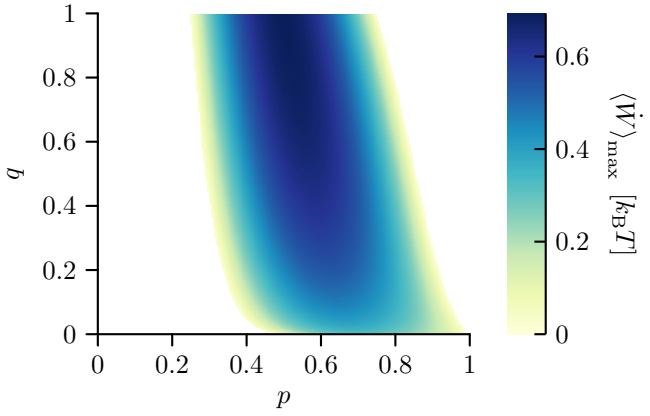

FIGURE 4.8. Parameter space regions that permit the all-1s-driven AS4C ratchet to function as an information engine. Notably, this includes only a band centered around detailed-balanced dynamics. The remainder of the uncolored parameter space forces $\beta\langle \dot{W} \rangle_{\max} < 0$, where the ratchet must function as a dud.

showed that this equality is, equivalently, a decomposition of stochastic thermodynamics' average total entropy production. To get there, we placed very few restrictions on the underlying system's dynamics, considering transitions between nonequilibrium dynamical states.

From this first law, we then applied integral fluctuation theorems to take the equalities to inequalities, reproducing and then tightening established bounds on average work production. By splitting the system into ratchet and tape subspaces and considering both finite and infinite-time

cases, we similarly reproduced and then tightened previous IPSLs of autonomous Maxwellian ratchets [**46**, **48**, **75**, **88**] to explicate the effects of nonequilibrium dynamical states. Finally, we illustrated these results with an example ratchet-tape system—the AS4C, driven by the ordered all-1s process. This demonstrated that, even under extreme simplification, the presence of NESSs introduced qualitative corrections to a ratchet's allowed behavior. In short, the presence of housekeeping entropy costs, induced by NESSs, directly counteracts a ratchet's ability to glean useful work from the creation of information.

Much room for further development remains, particularly in light of the role of fluctuation theorems in deriving these IPSLs. While our derivation concerned full ensemble averages, recent development of various trajectory-class fluctuation theorems [**28**, **89**] highlight opportunity to derive "trajectory class" IPSLs that would be more amenable to experimental verification via their freedom from rare-event statistical errors [**55**].

In addition, that odd-parity variables allow for meaningful decomposition of the housekeeping heat suggests further explication of their effects on the derived IPFLs and IPSLs, including bounds on asymptotic work extraction. Indeed, recent results in stochastic thermodynamics show that where a known, constrained splitting of the joint state space is available, it may be used to tighten the corresponding second laws [**95**]. Finally, considering infinite-state ratchets revealed new terms in the underlying IPSL whose convergence or divergence in general cases warrants detailed analytical investigation.

Taken altogether, the results demonstrate how a straightforward combination of tools—average change in steady-state surprisal and a single integral fluctuation theorem—simplify and generalize the derivation of various IPSLs. In turn, these place bounds on the extent to which systems can leverage information-bearing degrees of freedom to perform useful work. Furthermore, we showed explicitly how such inequalities arise from underlying equalities. This is in much the same fashion as stochastic thermodynamics' fluctuation theorems simplify to the original statement of the second law [**36**, **37**].

CHAPTER 5

Conclusion

After reviewing some of the basic notions of information theory and computational mechanics, this dissertation began with a detailed look at the foundations of *fluctuation theorems* in stochastic thermodynamics, a body of theory binding fluctuations to steady-state quantities in arbitrarily nonequilibrium systems. We found that, suggestively, many of the core functionals of stochastic thermodynamics generalize to dynamical functionals of individual path probabilities and surprisals, suggesting use binding fluctuating to steady-state quantities in generally *non*stationary stochastic processes—in particular, ones which violate detailed balance in their effective dynamics, and so exhibit nonequilibrium steady states.

With an eye toward just such an application—models of voltage-gated ion channels in mammalian neural membranes—we generalized and extended the theoretical reach of fluctuation theorems, deriving a trajectory class theorem for nonequilibrium steady-state, mesoscopic complex systems. This, in turn, provided a theoretical bedrock for probing highly nonequilibrium systems at whichever scales are most appropriate for experiment. In simulating such a potential experiment *in silico*, we found that the existence of nonequilibrium steady states in one of the two considered channels—channels which otherwise perform similar biological functionality—led to drastically different observed thermodynamic behavior. Indeed, the presence of nonequilibrium steady states brings with it—via an additional integral fluctuation theorem—an additional "dimension" of thermodynamic second law and additional opportunities for its single-shot violation. We observed both of these phenomena in simulations of the channels as driven by a biologically-plausible action potential signal.

Having extended stochastic thermodynamics in this vein, we then turned to explicit models of thermodynamic information processing: autonomous Maxwellian ratchets, which probe Maxwell's demon-like functionality and its requisite thermodynamic cost. We found that the viewpoint of generalized stochastic thermodynamics in the preceding study, one in which a fundamental shift is

made from viewing energy to viewing surprisal as the main "state function" of interest, lent itself naturally to deriving a first law for information processing. This described a system's change in Shannon information in exactly the same way that the familiar first law of thermodynamics does for a system's change in energy. Then, a splitting of the effective state space as appropriate to analyze a Maxwellian ratchet led directly to an equality in place of the "second law-type" inequalities of previous studies.

That equality, when decomposed to explicate the effects of nonequilibrium steady states, was then shown to couple explicitly with stochastic thermodynamics' fluctuation theorems to induce generalized and tightened forms of two information processing second laws for Maxwellian ratchets—one for finite information reservoirs and one for the infinite, asymptotic case, made explicitly tractable by the techniques of computational mechanics. Indeed, the information processing first law itself turned out to be a particular breakdown of the entropy production in stochastic thermodynamics—in a sense, it turned out that fluctuation theorems themselves lay at the heart of producing the links between information processing and thermodynamic second laws in information-thermodynamic systems.

Application of this revised second law to explicitly nonequilibrium steady-state information ratchets revealed the dramatic potential changes in allowed thermodynamic functionality: while a system may still, in principle, leverage the creation of information to do work, it must now compensate additionally for the entropic housekeeping costs associated with nonequilibrium steady states. Since an enormous majority of mesoscopic systems appear to violate detailed balance at observed scales—take, for example, the sodium ion channel model of Ch. 3—this suggests detailed consideration of such housekeeping costs and their corrections to related information processing.

In summary, this dissertation extened, leveraged, and combined various techniques of stochastic thermodynamics, information theory, and computational mechanics towards a principled understanding of the relationships between thermodynamics and information processing in mesoscopic complex systems.

5.1. Future Outlook

The results of the foregoing leave exciting space for future exploration.

First, the ion channel models considered as part of the example class in Ch. 3 were *Markov chains*, explicit effective-state models derived from single-channel measurements of current distributions. We did not question their genesis in using them to demonstrate the thermodynamic techniques we developed, but noted that they are not *unique*: indeed, many different effective models of those ion channels exist and are under development in the literature. At root, they share a common feature: output current—primarily, is it flowing or not?—is the observed quantity; the internal states themselves are not observed.

This suggests considering *hidden* Markov models for these systems, and leveraging the techniques of computational mechanics to describe their information-bearing structure and function. However, such a model is inherently nonunifilar: only one state (or transition) is associated with current production, while multiple may be associated with its absence. This calls forth use of recently-developed computational mechanical tools for probing nonunifilar hidden Markov models [**18**, **19**, **96**], and additionally prompts reevaluation of a common way to determine the energetics of a complex system, namely by treating an explicit Markov model over its state space.

Ch. 4 bumps into this same conundrum under a different context: for all but an entirely-ordered drive, even the simply-constructed hidden Markov model for its information ratchet system becomes highly nonunifilar. A continuing analysis of the effects of nonequilibrium steady states on an information-processing complex system might leverage recently-developed tools in the theory of computational mechanics.

The central theoretical result of Ch. 4, meanwhile, rests upon a full trajectory average of the steady-state surprisal. The result, an information processing first law, proceeded to unify and generalize several previously independent results in information and second law thermodynamics. What of the trajectory class theorem of Ch. 3 and those of Refs. [**89**, **91**]? Application to the theory in Ch. 4 suggests the develoment of a trajectory class information processing first law, capable of tuning the averages of the theory to adapt to appropriate experimental conditions.

Additionally, recent breakdowns of the housekeeping entropy production to include the effects of odd-parity state and control variables suggest yet another layer of decomposition to both the first

law of Ch. 4 and to the trajectory class theorem of Ch. 3: can we isolate entropic and information-processing costs associated directly with odd-parity spaces? Indeed, the recently-developed momentum computing paradigm [**97**] promises a rich testbed for such theoretical development, particularly in light of computing with nonequilibrium steady-state systems.

Finally, broad classes of complex systems—to the author's particular interest, landscape-ecological ones, for example—are probed via (nearly) continuous-state, continuous-time, noisy measurements. In addition, the continuous measurement alphabet—say, the interval $[0, 1]$ indicating a particular chemical's concentration in soil—often carries genuine scientific meaning in the ordering and metric over its alphabet space. While the theoretical bedrock of computational mechanics—its predictive equivalence relation—still applies, computational cost as well as interpretive power suffer using its traditional techniques, and the equivalence relation as-stated is agnostic to ordering or metric structure of the alphabet space. Research considering the construction of predictive states in more general settings—including continuous spaces—is ongoing [**20**, **21**, **22**, **23**]. Yet ecosystems, like ion channels, operate on physical substrates; they and their constituents consume, transorm, and produce energy as they do the same for information; they as well often operate far from equilibrium and far from steady state—for example, as they move through successional stages. The results of Chs. 3 and 4, along with developments of continuous-state extensions of computational mechanics, lead to an exciting program for future application-based inquiry: to what extent does the emergent complexity of ecological systems reflect their nonequilibrium thermodynamics? Are there computation-theoretic and thermodynamic signatures of biotic life throughout successional stages—if so, what form do they take?

Altogether, this dissertation has shown the remarkable reach of generalizing the paradigm and purview of stochastic thermodynamics. However, the generalizations in many cases bring added subtleties and considerations—most notably, subtleties induced by infinite predictive states, odd-parity variables, an appropriate thermodynamics of hidden Markov models, and the delicate questions of computational mechanics for continuous-alphabet, continuous-time systems. It is the author's greatest hope that these ideas, already deeply intertwined, someday coalesce into a theory combining the computational and stochastic-thermal mechanics of the complex systems so ubiquitous in the natural world.

APPENDIX A

NESS TCFT Derivation

In addition to requiring a unique stationary distribution for each protocol value, we assume that for any $x_{0:N} \in \mathcal{X}^{N+1}$:

(1) $x_{N:0} \in \mathcal{X}^{N+1}$,
(2) $\mathcal{P}_{\boldsymbol{\mu}_\mathrm{F}}(x_{0:N}) \neq 0 \implies \mathcal{P}_{\boldsymbol{\mu}_\mathrm{F}}(x_{N:0}) \neq 0$, and
(3) $\mathcal{R}_{\boldsymbol{\mu}_\mathrm{R}}(x_{N:0}) \neq 0 \implies \mathcal{R}_{\boldsymbol{\mu}_\mathrm{R}}(x_{0:N}) \neq 0$.

The second and third requirements, in particular, forbid one-way-only transitions in the discrete-time dynamic. Once we derive the TCFT, we will discuss the edge cases of completely irreversible trajectories.

Given the preceding constraints, a slightly rearranged form of Ref. [**42**]'s NESS DFT reads:

$$\mathcal{R}_{\boldsymbol{\mu}_\mathrm{R}}(x_{N:0}) = \mathcal{P}_{\boldsymbol{\mu}_\mathrm{F}}(x_{0:N})\, \mathrm{e}^{-(\mathcal{W}_\mathrm{ex} + \mathcal{Q}_\mathrm{hk} - \Delta\mathcal{F}^\mathrm{nss})} \ .$$

We wish to integrate both sides over a *trajectory class*—the measurable subset $C \subseteq \mathcal{X}^{N+1}$ of trajectories. We also define the *reverse trajectory class* $C_\mathrm{R} \doteq \{x_{N:0} \,|\, x_{0:N} \in C\}$. The following derivation mimics that of Ref. [**53**] after their Eq. (F3).

Integrating the lefthand side gives:

$$\begin{aligned}
&\int [x_{0:N} \in C]\, \mathcal{R}_{\boldsymbol{\mu}_\mathrm{R}}(x_{N:0})\, \mathrm{d}x_{0:N} \\
&\quad = \int [x_{0:N} \in C]\, \mathcal{R}_{\boldsymbol{\mu}_\mathrm{R}}(x_{N:0})\, \mathrm{d}x_{N:0} \\
&\quad = \int [x_{N:0} \in C_\mathrm{R}]\, \mathcal{R}_{\boldsymbol{\mu}_\mathrm{R}}(x_{N:0})\, \mathrm{d}x_{N:0} \\
&\quad = \mathcal{R}_{\boldsymbol{\mu}_\mathrm{R}}(C_\mathrm{R}) \ ,
\end{aligned}$$

where $[\cdot]$ is the Iverson bracket.

Integrating the righthand side gives:

$$\begin{aligned}
&\int [x_{0:N} \in C]\, \mathcal{P}_{\boldsymbol{\mu}_{\mathrm{F}}}(x_{0:N})\, \mathrm{e}^{-(\mathcal{W}_{\mathrm{ex}}+\mathcal{Q}_{\mathrm{hk}}-\Delta\mathcal{F}^{\mathrm{nss}})}\, \mathrm{d}x_{0:N} \\
&\quad = \int \mathcal{P}_{\boldsymbol{\mu}_{\mathrm{F}}}(x_{0:N} \cap C)\, \mathrm{e}^{-(\mathcal{W}_{\mathrm{ex}}+\mathcal{Q}_{\mathrm{hk}}-\Delta\mathcal{F}^{\mathrm{nss}})}\, \mathrm{d}x_{0:N} \\
&\quad = \mathcal{P}_{\boldsymbol{\mu}_{\mathrm{F}}}(C) \int \mathcal{P}_{\boldsymbol{\mu}_{\mathrm{F}}}\left(x_{0:N} \,|\, C\right)\, \mathrm{e}^{-(\mathcal{W}_{\mathrm{ex}}+\mathcal{Q}_{\mathrm{hk}}-\Delta\mathcal{F}^{\mathrm{nss}})}\, \mathrm{d}x_{0:N} \\
&\quad = \mathcal{P}_{\boldsymbol{\mu}_{\mathrm{F}}}(C) \left\langle \mathrm{e}^{-(\mathcal{W}_{\mathrm{ex}}+\mathcal{Q}_{\mathrm{hk}}-\Delta\mathcal{F}^{\mathrm{nss}})} \right\rangle_C ,
\end{aligned}$$

where $\langle \cdot \rangle_C$ is the average over the trajectory class C. Thus, we have Eq. (3.24)—a TCFT for NESS systems, whose forward and reverse processes may start in arbitrary distributions:

$$\frac{\mathcal{R}_{\boldsymbol{\mu}_{\mathrm{R}}}(C_{\mathrm{R}})}{\mathcal{P}_{\boldsymbol{\mu}_{\mathrm{F}}}(C)} = \left\langle \mathrm{e}^{-(\mathcal{W}_{\mathrm{ex}}+\mathcal{Q}_{\mathrm{hk}}-\Delta\mathcal{F}^{\mathrm{nss}})} \right\rangle_C . \tag{A.1}$$

Now, it remains to investigate the edge cases. Suppose that either (i) $\mathcal{P}_{\boldsymbol{\mu}_{\mathrm{F}}}(C) = 0$ or (ii) $\mathcal{R}_{\boldsymbol{\mu}_{\mathrm{R}}}(C_{\mathrm{R}}) = 0$, but not both. (The latter would amount to analyzing fluctuations for a *pair* of trajectories that never occur.) Since our probabilities are strictly nonnegative, the possible behaviors of the lefthand side are either $+\infty$ or 0, respectively, by considering the limit of a vanishing probability. In case (i), by definition either $\mathrm{e}^{-\mathcal{Q}_{\mathrm{hk}}} \to +\infty$ or $\mathrm{e}^{\Delta\mathcal{F}^{\mathrm{nss}}} \to +\infty$ (or both) for each forward trajectory, yielding agreement with the lefthand side. In case (ii), similarly either $\mathrm{e}^{-\mathcal{Q}_{\mathrm{hk}}} \to 0$ or $\mathrm{e}^{\Delta\mathcal{F}^{\mathrm{nss}}} \to 0$ (or both) for each forward trajectory. Since the preceding derivation established the TCFT for all nondiverging cases, this establishes its validity even in the divergent limiting cases.

APPENDIX B

Trajectory versus State Averaging

The main result of Ch. 4 relies on the equivalence between $\langle \Delta\phi_\lambda \rangle$ and $\Delta \langle \boldsymbol{\mu} | \boldsymbol{\phi}_\lambda \rangle = \langle \boldsymbol{\mu}_N | \phi_{\lambda_N} \rangle - \langle \boldsymbol{\mu}_0 | \phi_{\lambda_0} \rangle$. The former refers to $\Delta\phi_\lambda$'s average over an ensemble $\{x_{0:N+1}\}$ of repeated trajectories and, thus, means $\langle \phi_\lambda \rangle = \langle \mathcal{W}_{\text{ex}} \rangle - \langle \mathcal{Q}_{\text{ex}} \rangle$. The latter refers to two specific state averages of ϕ_λ. Namely, those at the trajectory's endpoints. And, it is equal to $\Delta H[\boldsymbol{\mu}] + \Delta D_{\text{KL}}[\boldsymbol{\mu} \,\|\, \boldsymbol{\pi}]$. We establish the equivalence here.

Definition B.0.1. The *trajectory average* of a path-dependent quantity $g(x_{0:N+1})$, denoted $\langle g \rangle$, is:

$$\langle g \rangle \doteq \int g(x_{0:N+1}) \Pr\left(x_{0:N+1}\right) \left(\prod_{i=0}^{N} \mathrm{d}x_i \right).$$

Definition B.0.2. The *state average* of a quantity $f(x_i)$, denoted $\langle \boldsymbol{\mu}_i | h \rangle$, is:

$$\langle \boldsymbol{\mu}_i | f \rangle \doteq \int f(x_i) \Pr\left(x_i\right) \mathrm{d}x_i \,.$$

Lemma B.0.1. For any $f(x_n)$ that depends only on one point $0 \leq x_n \leq N$ in the path, $\langle f \rangle = \langle \boldsymbol{\mu}_n | f \rangle$.

Proof. We explicitly evaluate the trajectory average. Consider two cases: (i) $n = N$, and (ii) $0 \leq n < N$.

(i) First, split the path probability into two pieces: $\Pr\left(x_{0:N+1}\right) = \Pr\left(x_{0:N}\right) \Pr\left(x_N \mid x_{0:N}\right)$. Now, evaluate the integrals for $\mathrm{d}x_0$ through $\mathrm{d}x_{N-1}$:

$$\int \Pr\left(x_{0:N}\right) \Pr\left(x_N \mid x_{0:N}\right) \left(\prod_{i=0}^{N-1} \mathrm{d}x_i \right) = \Pr\left(x_N\right),$$

by the law of total probability. The remainder is the $\mathrm{d}x_N$ integral:

$$\int \Pr\left(x_N\right) f(x_N) \, \mathrm{d}x_N = \langle \boldsymbol{\mu}_N | f \rangle \,,$$

by definition.

(ii) Again split the probability, but now as $\Pr(x_{0:N+1}) = \Pr(x_{0:n+1}) \Pr(x_{n+1:N+1} \mid x_{0:n+1})$. Evaluate the integrals for $\mathrm{d}x_{n+1}$ through $\mathrm{d}x_N$:

$$\int \Pr(x_{n+1:N+1} \mid x_{0:n+1}) \left(\prod_{i=n+1}^{N} \mathrm{d}x_i \right) = 1$$

by probability conservation. What remains is exactly case (i).

□

This assumes a truly finite stochastic process, such that no conditioning before x_0 or after x_N is possible or relevant. However, the result is robust in the limit of a bi-infinite stochastic process: evaluating the future integral in (ii) still yields 1 in the $N \to \infty$ limit. And, then, the past integral in (i) still gives $\Pr(x_n)$, even as the lower bound extends to $-\infty$.

Furthermore, we did not require Markovity, ergodicity, or even stationarity for the underlying stochastic process. The result, then, appears quite general. This is not too surprising. A point function's average over paths should not depend on the path. And, indeed, this result would be quite useless if not for the link between a path-independent ($\Delta\phi_\lambda$) and path-dependent ($\mathcal{W}_{\text{ex}}$ and $\mathcal{Q}_{\text{ex}}$) quantities provided by the nonaveraged First Law.

Bibliography

[1] H. Poincaré. *Les Methodes Nouvelles de la Mecanique Celeste*. Gauthier-Villars, Paris, 1892.

[2] P. W. Anderson. More Is Different: Broken symmetry and the nature of the hierarchical structure of science. *Science*, 177(4047):393–396, Aug. 1972.

[3] J. P. Crutchfield. Is Anything Ever New? Considering Emergence. In M. A. Bedau and P. Humphreys, editors, *Emergence: Contemporary Readings in Philosophy and Science*, pages 269–286. The MIT Press, Mar. 2008.

[4] J. P. Crutchfield. Between order and chaos. *Nat. Phys.*, 8(1):17–24, Jan. 2012.

[5] M. S. Williamson, C. W. Thackeray, P. M. Cox, A. Hall, C. Huntingford, and F. J. M. M. Nijsse. Emergent constraints on climate sensitivities. *Rev. Mod. Phys.*, 93(2):025004, May 2021.

[6] A. Chacoma, O. V. Billoni, and M. N. Kuperman. Complexity emerges in measures of the marking dynamics in football games. *Phys. Rev. E*, 106(4):044308, Oct. 2022.

[7] D. P. Varn and J. P. Crutchfield. Chaotic crystallography: How the physics of information reveals structural order in materials. *Current Opinion in Chemical Engineering*, 7:47–56, Feb. 2015.

[8] C. E. Shannon. A Mathematical Theory of Communication. *Bell System Technical Journal*, 27(3):379–423, July 1948.

[9] J. C. Maxwell. *Theory of Heat*. Longmans, Green and Co., London, United Kingdom, ninth edition, 1888.

[10] R. Landauer. Irreversibility and heat generation in the computing process. *IBM J. Res. Develop.*, 5(3):183–191, 1961.

[11] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley-Interscience, second edition, July 2006.

[12] J. P. Crutchfield and D. P. Feldman. Regularities unseen, randomness observed: Levels of entropy convergence. *Chaos*, 13(1):25–54, Mar. 2003.

[13] N. Barnett and J. P. Crutchfield. Computational Mechanics of Input–Output Processes: Structured Transformations and the ϵ-Transducer. *J. Stat. Phys.*, 161(2):404–451, Oct. 2015.

[14] S. Marzen and J. P. Crutchfield. Informational and Causal Architecture of Discrete-Time Renewal Processes. *Entropy*, 17(12):4891–4917, July 2015.

[15] S. E. Marzen and J. P. Crutchfield. Nearly maximally predictive features and their dimensions. *Phys. Rev. E*, 95(5):051301, May 2017.

[16] S. E. Marzen and J. P. Crutchfield. Structure and Randomness of Continuous-Time, Discrete-Event Processes. *J. Stat. Phys.*, 169(2):303–315, Oct. 2017.

[17] A. Rupe and J. P. Crutchfield. Local causal states and discrete coherent structures. *Chaos*, 28(7):075312, July 2018.

[18] A. M. Jurgens and J. P. Crutchfield. Shannon Entropy Rate of Hidden Markov Processes. *J. Stat. Phys.*, 183(2):32, May 2021.

[19] A. M. Jurgens and J. P. Crutchfield. Divergent predictive states: The statistical complexity dimension of stationary, ergodic hidden Markov processes. *Chaos*, 31(8):083114, Aug. 2021.

[20] S. P. Loomis and J. P. Crutchfield. Topology, Convergence, and Reconstruction of Predictive States. *arXiv:2109.09203 [cond-mat, physics:nlin, stat]*, Sept. 2021.

[21] N. Brodu and J. P. Crutchfield. Discovering causal structure with reproducing-kernel Hilbert space ϵ -machines. *Chaos*, 32(2):023103, Feb. 2022.

[22] A. Rupe and J. P. Crutchfield. Algebraic Theory of Patterns as Generalized Symmetries. *Symmetry*, 14(8):1636, Aug. 2022.

[23] S. P. Loomis and J. P. Crutchfield. Exploring predictive states via Cantor embeddings and Wasserstein distance. *Chaos*, 32(12):123115, Dec. 2022.

[24] I. Csiszár. I-Divergence Geometry of Probability Distributions and Minimization Problems. *Ann. Probab.*, 3(1), Feb. 1975.

[25] J. F. Coeurjolly, R. Drouilhet, and J. F. Robineau. Normalized information-based divergences. *Probl Inf. Transm.*, 43(3):167–189, Sept. 2007.

[26] Y. Cai and L.-H. Lim. Distances Between Probability Distributions of Different Dimensions. *IEEE Trans. Inform. Theory*, 68(6):4020–4031, June 2022.

[27] R. G. James, C. J. Ellison, and J. P. Crutchfield. Anatomy of a bit: Information in a time series observation. *Chaos*, 21(3):037109, Sept. 2011.

[28] M. T. Semaan and J. P. Crutchfield. Homeostatic and adaptive energetics: Nonequilibrium fluctuations beyond detailed balance in voltage-gated ion channels. *Phys. Rev. E*, 106(4):044410, Oct. 2022.

[29] M. C. Cross and P. C. Hohenberg. Pattern formation outside of equilibrium. *Rev. Mod. Phys.*, 65(3):851–1112, 1993.

[30] E. H. Trepagnier, C. Jarzynski, F. Ritort, G. E. Crooks, C. J. Bustamante, and J. Liphardt. Experimental test of Hatano and Sasa's nonequilibrium steady-state equality. *Proc. Natl. Acad. Sci. U.S.A.*, 101(42):15038–15041, Oct. 2004.

[31] M. Esposito. Stochastic thermodynamics under coarse graining. *Phys. Rev. E*, 85(4):041125, Apr. 2012.

[32] C. Y. Gao and D. T. Limmer. Principles of low dissipation computing from a stochastic circuit model. *arXiv:2102.13067*, 2021.

[33] Y. Oono and M. Paniconi. Steady State Thermodynamics. *Progress of Theoretical Physics Supplement*, 130:29–44, Jan. 1998.

[34] L. Onsager. Reciprocal relations in irreversible processes, I". *Phys. Rev.*, 37(4):405–426, 1931.

[35] D. Kondepudi. *Introduction to Modern Thermodynamics*. Wiley, Chichester, 2008.

[36] C. Jarzynski. Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach. *Phys. Rev. E*, 56(5):5018–5035, Nov. 1997.

[37] C. Jarzynski. Nonequilibrium Equality for Free Energy Differences. *Phys. Rev. Lett.*, 78(14):2690–2693, Apr. 1997.

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

[39] G. E. Crooks. Nonequilibrium Measurements of Free Energy Differences for Microscopically Reversible Markovian Systems. *Journal of Statistical Physics*, 90(5-6):1481–1487, Mar. 1998.

[40] T. Hatano and S.-i. Sasa. Steady-State Thermodynamics of Langevin Systems. *Phys. Rev. Lett.*, 86(16):3463–3466, Apr. 2001.

[41] T. Speck and U. Seifert. Integral fluctuation theorem for the housekeeping heat. *J. Phys. A: Math. Gen.*, 38(34):L581–L588, Aug. 2005.

[42] P. M. Riechers and J. P. Crutchfield. Fluctuations When Driving Between Nonequilibrium Steady States. *J. Stat. Phys.*, 168(4):873–918, Aug. 2017.

[43] U. Seifert. Stochastic thermodynamics: From principles to the cost of precision. *Physica A*, 504:176–191, Aug. 2018.

[44] P. Dayan and L. F. Abbott. *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. Computational Neuroscience. Massachusetts Institute of Technology Press, Cambridge, Mass, 2001.

[45] J. Schnakenberg. Network theory of microscopic and macroscopic behavior of master equation systems. *Reviews of Modern Physics*, 48(4):571–585, Oct. 1976.

[46] A. B. Boyd, D. Mandal, and J. P. Crutchfield. Identifying Functional Thermodynamics in Autonomous Maxwellian Ratchets. *New J. Phys.*, 18(2):023049, Feb. 2016.

[47] A. B. Boyd, D. Mandal, and J. P. Crutchfield. Thermodynamics of Modularity: Structural Costs Beyond the Landauer Bound. *Phys. Rev. X*, 8(3):031036, Aug. 2018.

[48] A. M. Jurgens and J. P. Crutchfield. Functional thermodynamics of Maxwellian ratchets: Constructing and deconstructing patterns, randomizing and derandomizing behaviors. *Phys. Rev. Research*, 2(3):033334, Aug. 2020.

[49] J. P. Sethna. *Entropy, Order Parameters, and Complexity*. Oxford Master Series in Physics. Oxford University Press, Oxford, second edition, 2021.

[50] R. E. Spinney and I. J. Ford. Nonequilibrium Thermodynamics of Stochastic Systems with Odd and Even Variables. *Phys. Rev. Lett.*, 108(17):170603, Apr. 2012.

[51] J. Yeo, C. Kwon, H. K. Lee, and H. Park. Housekeeping entropy in continuous stochastic dynamics with odd-parity variables. *J. Stat. Mech.*, 2016(9):093205, Sept. 2016.

[52] R. J. Harris and G. M. Schütz. Fluctuation theorems for stochastic dynamics. *J. Stat. Mech.*, 2007(07):P07020–P07020, July 2007.

[53] G. Wimsatt, O.-P. Saira, A. B. Boyd, M. H. Matheny, S. Han, M. L. Roukes, and J. P. Crutchfield. Harnessing fluctuations in thermodynamic computing via time-reversal symmetries. *Phys. Rev. Res.*, 3(3):033115, 2021.

[54] U. Çetiner, O. Raz, S. Sukharev, and C. Jarzynski. Recovery of Equilibrium Free Energy from Nonequilibrium Thermodynamics with Mechanosensitive Ion Channels in *E. coli*. *Phys. Rev. Lett.*, 124(22):228101, June 2020.

[55] C. Jarzynski. Rare events and the convergence of exponentially averaged work values. *Phys. Rev. E*, 73(4):046105, Apr. 2006.

[56] G. Hummer and A. Szabo. Free energy reconstruction from nonequilibrium single-molecule pulling experiments. *Proc. Natl. Acad. Sci. U.S.A.*, 98(7):3658–3661, Mar. 2001.

[57] J. Liphardt, S. Dumont, S. B. Smith, I. Tinoco, and C. Bustamante. Equilibrium Information from Nonequilibrium Measurements in an Experimental Test of Jarzynski's Equality. *Science*, 296(5574):1832–1835, June 2002.

[58] C. Jarzynski. Work Fluctuation Theorems and Single-Molecule Biophysics. *Prog. Theor. Phys. Suppl.*, 165:1–17, 2006.

[59] U. Seifert. Stochastic thermodynamics, fluctuation theorems and molecular machines. *Rep. Prog. Phys.*, 75(12):126001, Dec. 2012.

[60] D. Collin, F. Ritort, C. Jarzynski, S. B. Smith, I. Tinoco, and C. Bustamante. Verification of the Crooks fluctuation theorem and recovery of RNA folding free energies. *Nature*, 437(7056):231–234, Sept. 2005.

[61] S. Lahiri and A. M. Jayannavar. Fluctuation theorems for excess and housekeeping heat for underdamped Langevin systems. *Eur. Phys. J. B*, 87(9):195, Sept. 2014.

[62] D. Mandal, K. Klymko, and M. R. DeWeese. Entropy Production and Fluctuation Theorems for Active Matter. *Phys. Rev. Lett.*, 119(25):258001, Dec. 2017.

[63] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley-Interscience, New York, second edition, 2006.

[64] P. M. Riechers and M. Gu. Initial-state dependence of thermodynamic dissipation for any quantum process. *Phys. Rev. E*, 103(4):042145, Apr. 2021.

[65] C. A. Vandenberg and F. Bezanilla. A sodium channel gating model based on single channel, macroscopic ionic, and gating currents in the squid giant axon. *Biophys. J.*, 60(6):1511–1533, Dec. 1991.

[66] O.-P. Saira, M. H. Matheny, R. Katti, W. Fon, G. Wimsatt, J. P. Crutchfield, S. Han, and M. L. Roukes. Nonequilibrium thermodynamics of erasure with superconducting flux logic. *Phys. Rev. Research*, 2(1):013249, Mar. 2020.

[67] A. L. Hodgkin and A. F. Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.*, 117(4):500–544, Aug. 1952.

[68] K. Pal and G. Gangopadhyay. Dynamical characterization of inactivation path in voltage-gated Na$\hat{}\{+\}$ ion channel by non-equilibrium response spectroscopy. *Channels*, 10(6):478–497, Nov. 2016.

[69] E. Marban, T. Yamagishi, and G. F. Tomaselli. Structure and function of voltage-gated sodium channels. *J. Physiol.*, 508(3):647–657, May 1998.

[70] E. M. Izhikevich. *Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting*. Computational Neuroscience. The MIT Press, first edition, Nov. 2006.

[71] E. M. Izhikevich. Simple model of spiking neurons. *IEEE Trans. Neural Netw.*, 14(6):1569–1572, Nov. 2003.

[72] M. T. Semaan and J. P. Crutchfield. First and Second Laws of Information Processing by Nonequilibrium Dynamical States. *arXiv:2211.05849 [cond-mat, physics:nlin]*, Nov. 2022.

[73] H. Leff and A. F. Rex. *Maxwell's Demon 2: Entropy, Classical and Quantum Information, Computing*. Taylor and Francis, New York, 2002.

[74] D. Mandal and C. Jarzynski. Work and information processing in a solvable model of Maxwell's demon. *Proc. Natl. Acad. Sci. U.S.A.*, 109(29):11641–11645, July 2012.

[75] L. He, A. Pradana, J. W. Cheong, and L. Y. Chew. Information processing second law for an information ratchet with finite tape. *Phys. Rev. E*, 105(5):054131, May 2022.

[76] D. Mandal and C. Jarzynski. Analysis of slow transitions between nonequilibrium steady states. *J. Stat. Mech.*, 2016(6):063204, June 2016.

[77] R. E. Spinney and I. J. Ford. Entropy production in full phase space for continuous stochastic dynamics. *Phys. Rev. E*, 85(5):051113, May 2012.

[78] L. Cocconi, R. Garcia-Millan, Z. Zhen, B. Buturca, and G. Pruessner. Entropy Production in Exactly Solvable Systems. *Entropy*, 22(11):1252, Nov. 2020.

[79] D. J. Skinner and J. Dunkel. Improved bounds on entropy production in living systems. *Proc. Natl. Acad. Sci. U.S.A.*, 118(18):e2024300118, May 2021.

[80] D. J. Skinner and J. Dunkel. Estimating Entropy Production from Waiting Time Distributions. *Phys. Rev. Lett.*, 127(19):198101, Nov. 2021.

[81] C. Bechinger, R. Di Leonardo, H. Löwen, C. Reichhardt, G. Volpe, and G. Volpe. Active Particles in Complex and Crowded Environments. *Rev. Mod. Phys.*, 88(4):045006, Nov. 2016.

[82] A. I. Brown and D. A. Sivak. Theory of Nonequilibrium Free Energy Transduction by Molecular Machines. *Chem. Rev.*, 120(1):434–459, Jan. 2020.

[83] S. Rahav and C. Jarzynski. Fluctuation relations and coarse-graining. *J. Stat. Mech.*, 2007(09):P09012–P09012, Sept. 2007.

[84] A. Puglisi, S. Pigolotti, L. Rondoni, and A. Vulpiani. Entropy production and coarse graining in Markov processes. *J. Stat. Mech.*, 2010(05):P05015, May 2010.

[85] G. Teza and A. L. Stella. Exact Coarse Graining Preserves Entropy Production out of Equilibrium. *Phys. Rev. Lett.*, 125(11):110601, Sept. 2020.

[86] A. Ghosal and G. Bisker. Inferring entropy production rate from partially observed Langevin dynamics under coarse-graining. *Phys. Chem. Chem. Phys.*, 24(39):24021–24031, 2022.

[87] S. P. Loomis and J. P. Crutchfield. Thermal Efficiency of Quantum Memory Compression. *Phys. Rev. Lett.*, 125(2):020601, July 2020.

[88] A. B. Boyd, D. Mandal, and J. P. Crutchfield. Leveraging Environmental Correlations: The Thermodynamics of Requisite Variety. *J. Stat. Phys.*, 167(6):1555–1585, June 2017.

[89] G. Wimsatt, A. B. Boyd, and J. P. Crutchfield. Trajectory Class Fluctuation Theorem, July 2022.

[90] A. Kolchinsky and D. H. Wolpert. Dependence of dissipation on the initial distribution over states. *J. Stat. Mech.*, 2017(8):083202, Aug. 2017.

[91] G. Wimsatt, O.-P. Saira, A. B. Boyd, M. H. Matheny, S. Han, M. L. Roukes, and J. P. Crutchfield. Harnessing fluctuations in thermodynamic computing via time-reversal symmetries. *Phys. Rev. Research*, 3(3):033115, Aug. 2021.

[92] Y.-J. Yang and H. Qian. Unified formalism for entropy production and fluctuation relations. *Phys. Rev. E*, 101(2):022129, Feb. 2020.

[93] J. G. Kemeny and J. L. Snell. *Denumerable Markov Chains*. Graduate Texts in Mathematics. Springer New York, New York, NY, second edition, 1976.

[94] A. B. Boyd, D. Mandal, and J. P. Crutchfield. Identifying functional thermodynamics in autonomous Maxwellian ratchets. *New J. Physics*, 18:023049, 2016.

[95] D. H. Wolpert. Strengthened second law for multi-dimensional systems coupled to multiple thermodynamic reservoirs. *Phil. Trans. R. Soc. A.*, 380(2227):20200428, July 2022.

[96] A. M. Jurgens and J. P. Crutchfield. Ambiguity rate of hidden Markov processes. *Phys. Rev. E*, 104(6):064107, Dec. 2021.

[97] K. J. Ray, A. B. Boyd, G. W. Wimsatt, and J. P. Crutchfield. Non-Markovian momentum computing: Thermodynamically efficient and computation universal. *Phys. Rev. Research*, 3(2):023164, June 2021.