ABSTRACT

Title of dissertation: RECTIFYING THERMAL FLUCTUATIONS: MINIMAL PUMPING AND MAXWELL'S DEMON

Dibyendu Mandal, Doctor of Philosophy, 2013

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

Molecular complexes with movable components form the basis of nanoscale machines. Their inherent stochastic nature makes it a challenge to generate any controllable movement. Rather than fighting these fluctuations, one can utilize them by the periodic modulation of system parameters, or *stochastic pumping*. For the no-pumping theorem (NPT), which establishes minimal conditions for directed pumping, we present a simplified proof using an elementary graph theoretical construction. Motivated by recent experiments, we propose a new class of "hybrid" models combining elements of both the purely discrete and purely continuous descriptions prevalent in the field. We formulate the NPT in this hybrid framework to give a detailed justification of the original experiment observation. We also present an extension of the NPT to open stochastic systems.

Next we consider the paradox of "Maxwell's demon", an imaginary intelligent being that rectifies thermal fluctuations in a manner that seems to violate the second

law of thermodynamics. We present two exactly solvable, autonomous models that can reproduce the actions of the demon. Of necessity, both of these models write information on a memory device as part of their operation. By exposing their explicit, transparent mechanisms, our models offer simple paradigms to investigate the autonomous rectification of thermal fluctuations and the thermodynamics of information processing.

RECTIFYING THERMAL FLUCTUATIONS: MINIMAL PUMPING AND MAXWELL'S DEMON

by

Dibyendu Mandal

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2013

Advisory Committee:
Professor Christopher Jarzynski, Chair/Advisor
Professor Michael E. Fisher
Professor Jay R. Dorfman
Professor Theodore L. Einstein
Professor Garegin A. Papoian

Acknowledgments

It is my great pleasure to acknowledge here all who have helped me through my graduate life and made this thesis possible. First and foremost, I would like to thank my advisor, Prof. Christopher Jarzynski, for his patient support and guidance. By his constant encouragement and interactions he has helped me grow as a researcher. He has taught me the value of clarity in any scientific activity. Thank you Chris; I could not hope for any better experience.

I would then like to thank all the members of our group for their excellent company and the great learning environment they maintain. Thank you Suri for introducing me to the group. Thank you Andy for all our wonderful chats. Thank you Haitao for our collaboration. Thank you Shaon, Sebastian, Zhiyue and Rian for your friendship.

I would then like to thank all my friends in College Park, especially, Basudev Roy, Wrick Sengupta, Anirban Ghosh, Anirban Gangopadhyay, Santanu Debnath, Knightvid Cole, Srimoyee Sen, Ayoti Patra, Subhasis Mukherjee, Kazi Rajibul Islam and Soumya Samanta. Life without you would not have been as joyful and colorful.

I would now like to thank my close friends and family. Thank you Arindam and Aprameyo; I always cherish your friendship. Thank you Swati, for sharing your life with me. Thank you dada, baba and ma, for making all this possible.

I also gratefully acknowledge financial support from the National Science Foundation (USA) (Grant Nos. 0925365 and 0906601), the U.S.-Israel Binational Science Foundation (Grant No. 2010363), and the University of Maryland, College Park.

Table of Contents

List of Figures v

List of Symbols and Abbreviations vi

1 Introduction 1

2 Discrete state master equation 10
2.1 General form and graphical representation . . . 11
2.2 Steady state distribution . . . 13
2.3 Periodic steady state distribution . . . 15
2.4 Detailed balance . . . 18
2.5 Local detailed balance . . . 22

3 No-pumping theorem for discrete pumps 24
3.1 Closed stochastic pumps . . . 26
3.1.1 Illustration . . . 28
3.1.2 General Proof . . . 31
3.2 Open stochastic pumps . . . 35
3.2.1 Kinetic equations . . . 36
3.2.2 Illustration . . . 41
3.2.3 General Proof . . . 43

4 Hybrid pumps 46
4.1 Hybrid Model of a [2]catenane . . . 47
4.2 Constraints Imposed by Detailed Balance . . . 52
4.3 Statement and Proof of NPT . . . 54
4.4 Generalizations . . . 58

5 Szilard's engine 65
5.1 Model . . . 69
5.2 Analysis . . . 77
5.3 Modes of operation . . . 82
5.4 Modified second law of thermodynamics . . . 85

6 Maxwell's refrigerator 89
6.1 Model . . . 90
6.2 Analysis . . . 96
6.3 Phase diagram . . . 100
6.4 Modified Clausius inequality . . . 103

7 Summary and future outlook 108

A Derivation of Eq. 4.21 111

B Derivation of Φ (as in Eq. 5.19) 113

C Derivation of inequality 6.26 116

Bibliography 119

List of Figures

1.1 Artificial molecular machines 2
1.2 Maxwell's demon 7

2.1 Graphical representation of discrete state master equation. 12
2.2 Effective free energies and barriers. 19

3.1 A model 2-catenane. 28
3.2 A simple model 29
3.3 Part of an N state graph with arbitrary topology. 33
3.4 A simple open stochastic system 36
3.5 Particle exchange with a reservoir 38
3.6 Flow from a source to a drain reservoir 44

4.1 Hybrid model of 2-catenane 48
4.2 Elements of hybrid model for 2-catenane 50
4.3 Free energies and barriers 53
4.4 A generic hybrid model 59
4.5 Notations for a generic hybrid network 60

5.1 Szilard's engine 66
5.2 Our model akin to Szilard's engine 68
5.3 Models of our demon and each bit 69
5.4 Detailed balance 71
5.5 Graph of the joint master equation 72
5.6 Illustration of "unphysical" transitions 74
5.7 Phase diagram of the engine model 83

6.1 Setup for Maxwell's refrigerator 90
6.2 Illustration of the dynamics 91
6.3 Phase diagram 101

A.1 Details of source and sink 112

C.1 Illustration of $f(\delta)$ 117

List of Symbols and Abbreviations

The list describes the commonly used symbols and abbreviations in this thesis and the pages where they are mentioned for the first time.

Symbol	**Description**		**Page**
$\mathbf{1}$	Column vector with unit entries	...	13
$\mathcal{A}(c)$	Affinity of cycle c in a graph	...	19
C	Number of cycles in a graph	...	13
D	Diffusion coefficient	...	51
E	Number of edges in a graph	...	13
$\mathbb{I}$ (chap. 5)	3×3 identity matrix	...	77
I	Mutual information	...	105
$J_{ij} = R_{ij}p_j - R_{ji}p_i$	Current from j to i	...	11
J_{ij}^s	J_{ij} in steady state	...	18
J_{ij}^{ps}	J_{ij} in periodic steady state	...	17
k	Boltzmann constant	...	20
$\mathcal{N}$	Normalization factor	...	21
N	Number of nodes in a graph	...	11
$\mathbf{p} = (p_1 \dots p_N)$	Probability distribution	...	11
	among a discrete set of states		
$\dot{\mathbf{p}} = \mathcal{R}\mathbf{p}$	Rate of change of $\mathbf{p}$	...	11
$\mathbf{p}^s$	$\mathbf{p}$ in steady state	...	14
$\mathbf{p}^{ps}$	$\mathbf{p}$ in periodic steady state	...	15
$\mathcal{R}$	Rate matrix	...	11
$\mathrm{Re}(X)$	Real part of X	...	15
$R_{ij}, R_{ij}(t)$	Transition rate from j to i	...	11
$R_{i\leftarrow j}$, $R_{j\rightarrow i}$	Transition rate from j to i	...	91
T	Transposition	...	11
	(when appearing as a superscript)	...	
T	Temperature of the environment	...	20
$\hat{\mathrm{T}}$	Time ordering	...	16

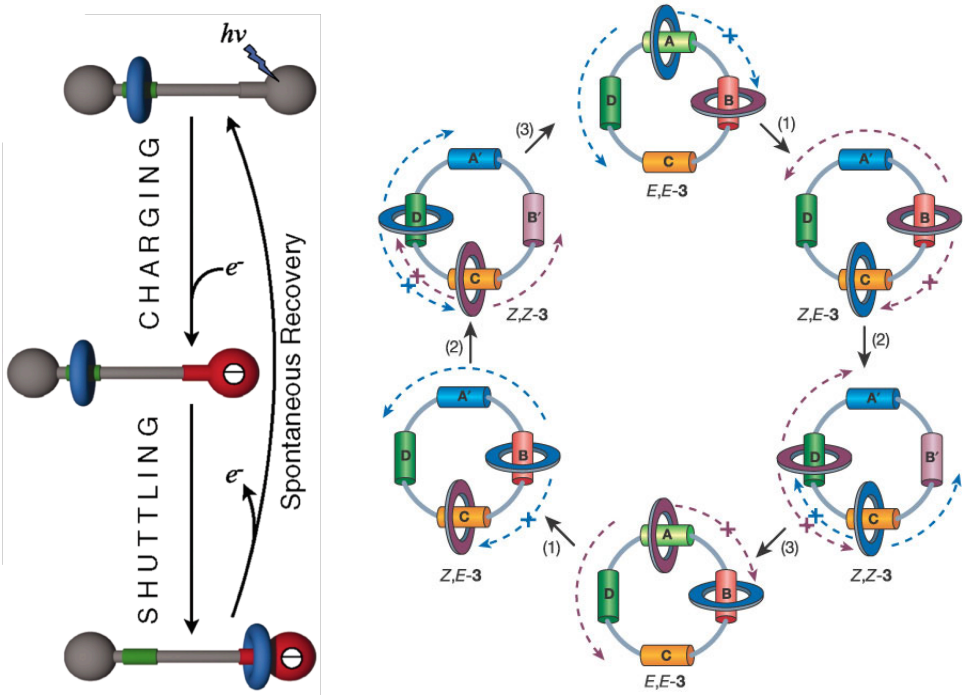
Symbol	Description		Page
$\Delta S_B = S_b' - S_b$	Change in information per bit	...	81
ΔS_R	Change in reservoir-entropy per bit	...	86
$\Delta\chi_n$	CW rotation in n^{th} interaction interval	...	76
$\delta = p_0 - p_1$	Excess of incoming 0's	...	75
$\delta' = p_0' - p_1'$	Excess of outgoing 0's	...	77
$\delta(x)$	Dirac delta function	...	51
δ_{ij}	Kronecker delta	...	14
ϵ (chap. 5)	Rescaled mass	...	71
ϵ (chap. 6)	Rescaled temperature difference	...	92
λ_i	i^{th} eigenvalue of $\mathcal{R}$	...	14
τ	Time period or interaction interval	...	15
$\Phi = p_1' - p_1$	Average production of 1's per τ	...	77
χ, $\chi(t)$	Number of CW rotations	...	75
CW	Clockwise	...	75
CCW	Counter-clockwise	...	75
NPT	No-pumping theorem	...	5
PRT	Pumping restriction theorem	...	5
PQT	Pumping quantization theorem	...	5

Chapter 1

Introduction

In 1959 Richard Feynman gave a visionary lecture describing a world of molecular nanotechnology, where machines would be built with a few hundred atoms and the entire Encyclopedia Britannica could be written on the head of a pin [1]. Technologies like scanning tunneling microscopy (STM), atomic force microscopy (AFM) and electron beam lithography are all examples of this extreme miniaturization. With the works of Eric Drexler [2, 3], among others, an alternative "bottom-up" approach started being explored; the goal was to fabricate molecular components of desired specifications and assemble them to build complex nanomachines. There has been extraordinary growth in this field in recent years [4–11], with achievements including molecular rings that can be shuttled between the ends of a molecular axle [12] or rotated unidirectionally along a large molecular ring [13], nanoscale structures that can perform translational motion along a predefined path on surfaces or tracks [8, 14, 15], a prototypical molecular factory [16], and a single molecule electric motor [17]. Some of these systems are shown in Fig. 1.1.

The small (nanoscale) size of molecular machines has important consequences: thermal fluctuations and viscosity of the environment play dominant roles in their dynamics whereas the gravitational effects are negligible [7]. The thermal fluctua-



(a) (b)

(c)

Figure 1.1: **Artificial molecular machines.** (a) Panman *et al.*'s molecular shuttle, an example of rotaxanes. The macrocycle (ring) can be shuttled back and forth along the axle by controlled application of UV rays. Taken from Ref. [18]. (b) Leigh *et al.*'s [3]catenane. The smaller rings rotate along the bigger ring unidirectionally through a sequence of steps. Taken from Ref. [13]. (c) Pei *et al.*'s quadruped molecular walker diffusing on a prescriptive landscape. Taken from Ref. [19].

tions cause an unavoidable incessant movement, the so-called Brownian movement, at this length scale. The goal in the design of molecular machines is to rectify and harness these movements. Viscous effects dominate over inertial effects. Motion of these molecular machines is solely determined by the instantaneous forces acting on them rather than any previous "push". Because of these differing principles, design and control of molecular machines require fundamentally different frameworks than their macroscopic counterparts.

It is useful to distinguish between two types of strategies to control these molecular systems: the *autonomous* mode, where the machines execute their tasks without any external intervention, and the *non-autonomous* mode, where external interventions are necessary. Let us illustrate each of these modes with an example. Consider the artificial multipedal molecules, or *DNA walkers*, that can walk on a two-dimensional "origami" of folded DNA [8]. In Ref. [15] they are controlled by completely autonomous means. Their intended path is first grafted with longer DNA strands, with longer binding affinity compared to the other stands on the surface. The walkers bind to these longer strands preferentially, cleave them short by enzymatic action, and then unbind to move towards the remaining longer stands. This produces an average forward motion, without the need of any external intervention. The walkers in Ref. [16], in contrast, need periodic injection of new DNA strands for their locomotion. All strands on the surface are equal; depending on the type of the added strand ("anchor" or "fuel"), the walkers may either bind to the path strands or unbind from them. As constant monitoring of the states of the walkers and addition of appropriate strands are necessary, these walkers are non-autonomous

machines.

The theoretical framework of autonomous molecular machines is well-developed from a somewhat different context. Many biomolecular processes in our body such as intracellular cargo transport, muscle contraction or microtubule polymerization are carried out by small biological entities called molecular machines [20–23]. These highly evolved naturally occurring molecular complexes provided much of inspiration for the development of artificial molecular machines [1, 7]. These biological machines are autonomous, utilizing ATP hydrolysis, ion concentration gradient, or transmembrane electrical potential, to carry out their tasks. Because of their similar length scale and environment, the well-developed stochastic description of natural molecular machines also applies to their artificial analogs. Refs. [24–26] and references therein discuss this framework in detail.

The theoretical framework of non-autonomous molecular machines is relatively less explored. Key results in this framework have been reported only recently [27–50]. Two possible ways to control the non-autonomous machines are *stochastic pumping* and *feedback control*. Stochastic pumping involves periodic variation of the external parameters of the system. Feedback control, on the other hand, involves making measurements on the system and then changing the external parameters suitably according to the outcome. In this thesis, we shall be concerned with the former mode of operation, namely stochastic pumping.

Of primary concern in theory of stochastic pumps is their average response characterized by average probability currents. When the variation of external parameters is quasi-static (adiabatic pumping) the average probability currents are

shown to have geometric aspects [27, 28, 32, 33], much like the Berry phase in quantum mechanics [51]. For the more general case of non-adiabatic pumping Rahav, Horowitz, and Jarzynski (2008) have derived an exact (formal) expression for the probability current [29]. With this they have established a no-pumping theorem (NPT) which explicitly states the minimal pumping protocols necessary to drive non-zero average current; see the discussion related to Eq. 3.5. Chernyak and Sinitsyn (2008) have derived a more general pumping restriction theorem (PRT) which relates the number of independent currents to any given pumping protocol [30]. They have also discovered the quantized nature of the average current, the so-called pumping quantization theorem (PQT), in the limit of low temperature [32, 37, 38]. Further developments noted supersymmetry [34], duality [35] and other attributes of the current [28, 31, 36].

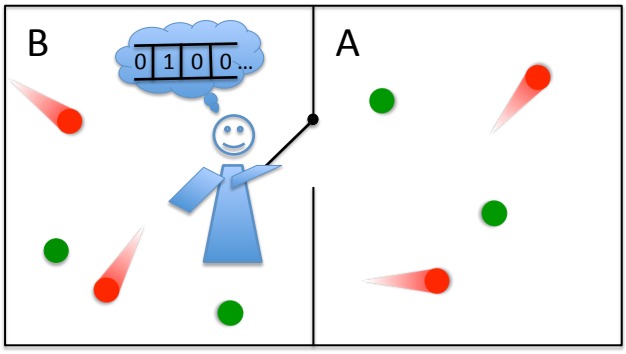
It is desirable that the above results are understood in simple terms, and possibly in multiple ways, to provide us further insight into these results and a plausible framework to unify them. With this goal in mind we devote the first part of the thesis to a simple graphical derivation of the NPT of Rahav, Horowitz, and Jarzysnki (2008) [29] and its extension to more general systems. The central relation underlying the derivation is precursor to the derivation of PQT of Chernyak and Sinitsyn (2009) [32].

In the second part of the thesis, we change to the topic of "Maxwell's demon". This an imaginary intelligent being, introduced by James Clerk Maxwell, that rectifies thermal fluctuations in a manner that seems to violate the second law of thermodynamics. In his book "Theory of heat" [53], Maxwell wrote

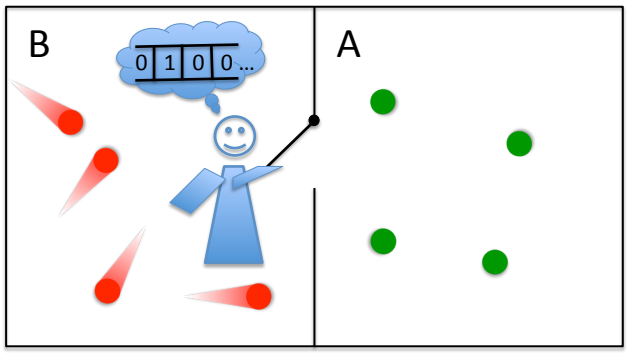
> " ... let us suppose that ... a vessel is divided into two portions, A and B, by a division in which there is a small hole, and that a being, who can see the individual molecules, opens and closes this hole, so as to allow only the swifter molecules to pass from A to B, and only the slower ones to pass from B to A. He will thus, without expenditure of work, raise the temperature of B and lower that of A, in contradiction to the second law of thermodynamics. "

He took it to be an illustration of the statistical basis of the second law of thermodynamics. We have schematically depicted his setup in Fig. 1.2.

William Thomson, among others, emphasized that the intelligent being of Maxwell's, which he named "Maxwell's demon", provided a mechanism to systematically violate the second law of thermodynamics [54]. Subsequent questions and confusions about the status of the second law have generated more than a century of discussions [55–64]: " If such a demon cannot defeat the second law, then why not? And if it *can* defeat the second law, then how does that affect that law's status?" [52]. Several versions of the demon have been invented and several solutions proposed. Consensus in the community now lies with the works of Rolf Landauer [58], Oliver Penrose [59] and Charles H. Bennett [60]. They proposed that the demon accumulates information about the molecules' motion during its operation, and discarding this information has a minimum entropic cost that compensates the demon's violation of the second law. In effect, information is seen as being equivalent to thermodynamic entropy: accumulation of information by the de-



(a)



(b)

Figure 1.2: **Maxwell's demon.** (a) Initially, both sides of the box have the same temperature, therefore they have both fast (with jet) and slow (without jet) particles. (b) The demon sorts the fast particles to the B half and the slow particles to the A half. As a result, temperature goes up in the B side and goes down in the A side.

mon lets it decrease the thermodynamic entropy of the rest of the universe; as soon as the memory is "full" and the demon needs to reset it, the minimum entropic cost of information erasure (*Landauer's principle*) compensates for the previous decrease of thermodynamic entropy.

The past few years have seen an increased interest in the thermodynamics of information processing [65–70]. Discussions of Maxwell's demon, Landauer's principle and related topics arise in contexts such as quantum information theory [71], the synthesis of artificial nanoscale machines [7], feedback control in microscopic systems [39–50], and single-photon cooling of atoms [72]. Moreover the consensus or "favored explanation" [64] described above is widely but not universally accepted, as suspicions persist that it assigns an unwarranted thermodynamic significance to random data [62–64, 73, 74].

In spite of this attention, the field has lacked tangible examples or model devices of the demon. Discussions are often framed around general principles rather than particular instances. Furthermore, the actions of measurement and information accumulation, in themselves, do not require the demon to be intelligent; a computer program, for example, can be made to do these tasks. One may therefore wonder if it is possible to design an *autonomous mechanical* device, without any intelligence or external intervention, which can behave like the demon. The second part of the thesis is devoted to two such models. Specifically, in Ch. 5, we propose a stochastic device that extracts energy from a single thermal reservoir and converts it into work by raising a mass against gravity (in violation of the Kelvin-Planck statement of the second law [75]); the device, however, requires a memory register to which it

can write information. In Ch. 6 we propose a similar device which can generate a flow of energy against a thermal gradient without any external work, just like the original Maxwell's demon (in violation of the Clausius statement of the second law [75]). We solve for the steady state behavior of these models exactly and draw their non-equilibrium phase diagram. To the best of our knowledge, they are the first models of their kinds in the field.

The structure of the thesis is as follows. In Ch. 2 we discuss the mathematical framework of discrete state Markov processes, the mathematical basis of much of the thesis. In Ch. 3 we discuss our graphical proof of the NPT. We also extend the usual case of closed stochastic pumps to open stochastic pumps. In Ch. 4 we extend this work furthermore to a new class of models, the "hybrid models", which combine elements of both diffusive and jump dynamics of Markov processes. We shift the discussion to Maxwell's demon in the ensuing chapters. In Ch. 5 we discuss the engine model, and in Ch. 5 we discuss the refrigerator model. We conclude in Ch. 7 by outlining the possible extensions of the thesis. A few appendices present the details omitted in the main text.

Chapter 2

Discrete state master equation

All model systems in this thesis are stochastic. Furthermore, they all follow Markovian dynamics: their past states do not give any new information about their future states if the present states are known to us [76]. The time evolution of any such system is described by the so-called master equation: a linear differential equation(s), first order in time, involving the (transition) probability distribution of the system [76]. Apart from the systems in Ch. 4, all models of the thesis are also discrete – they can access only a finite number of states. The corresponding master equation is formed by a set of first order, linear, coupled differential equations involving the discrete state probabilities of the system. In this chapter, we discuss the general properties of such discrete-state master equations to facilitate our discussions in the rest of the thesis.

In Sec. 2.1 we discuss a convenient graphical representation of discrete state master equations. In Sec. 2.2 we discuss the algebraic properties of their rate matrices (Eq. 2.2). We describe how all initial conditions relax to a steady state distribution and the conditions under which this steady state is unique. In Sec. 2.3 we consider time-periodic rate matrices i.e. stochastic pumps. We show how all initial conditions now relax to a periodic steady state. Finally, in Sec. 2.4 we consider the

important restriction of detailed balance on the rate matrices.

2.1 General form and graphical representation

Consider a model system which has N discrete states $i \in \{1, \ldots, N\}$. Let $p_i(t)$ denote the probability to find the system in states i at time t. One can construct a probability vector $\mathbf{p}(t) = (p_1(t), \ldots, p_N(t))^T$ for the instantaneous probability distribution of the system. Let R_{ij} denote the conditional transition rate from state j to state i, for any distinct pair of state i and j. $R_{ij} = 0$ if no direct transition is possible from j to i. For the time being we assume that these rates are time-independent. If we define

$$R_{ii} = -\sum_{j \neq i} R_{ji}, \tag{2.1}$$

for all i, we can construct a matrix $\mathcal{R}$. The master equation for the system is then given by

$$\dot{\mathbf{p}}(t) = \mathcal{R}\,\mathbf{p}(t). \tag{2.2}$$

There is a nice physical way to describe these dynamics. Whenever the system jumps into some state i, its subsequently waits for a time τ_i, which is an exponentially distributed random variable with average $1/|R_{ii}|$, and then makes a sudden transition to some other state j with probability $R_{ji}/|R_{ii}|$.

Eq. 2.2 can also be viewed as a continuity equation. Consider the quantities

$$J_{ij}(t) = R_{ij}\,p_j(t) - R_{ji}\,p_i(t), \tag{2.3}$$

for any pair of distinct states i and j. The quantity $J_{ij}(t)$ is the net rate of transition from j to i,i.e. the instantaneous probability current in that direction. The master

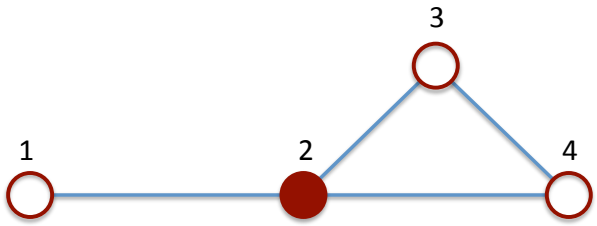


Figure 2.1: **Graphical representation of discrete state master equation 2.5.** The system has $N = 4$ states and $E = 4$ edges. Therefore it has $C = E - N + 1 = 1$ cycle in the graph: {2,3,4}.

equation 2.2 can be rewritten as

$$\dot{p}_i(t) = \sum_{j \neq i} J_{ij}(t), \tag{2.4}$$

which implies that the rate of change of probability of any state i is the total instantaneous probability current into it. This is the usual content of any continuity equation.

In all subsequent discussions we assume that $R_{ij} = 0$ if and only if $R_{ji} = 0$. This implies that the system is *reversible*: if a transition is possible in one direction the reverse transition is also possible. Furthermore, we assume that any state can be reached from any other state either directly or via intermediate states. This implies the system is *ergodic*, the importance of which will become clear in the next section.

The dynamics under master equation 2.2 can be conveniently represented by a graph whose nodes represent the discrete states of the system and the edges represent the allowed transitions. Note that we do not need to add direction to edges because the dynamics have been assumed to be reversible. Ergodicity, on the other hand, implies that the graph is connected – any node can be reached from any

other node through intermediate edges and nodes. We have illustrated the graphical representation in Fig. 2.1 for the following $N = 4$ state master equation

$$\begin{pmatrix} \dot{p}_1 \\ \dot{p}_2 \\ \dot{p}_3 \\ \dot{p}_4 \end{pmatrix} = \begin{pmatrix} -\frac{1}{2} & 1 & 0 & 0 \\ \frac{1}{2} & -3 & 1 & 1 \\ 0 & 1 & -2 & 1 \\ 0 & 1 & 1 & -2 \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \end{pmatrix} \tag{2.5}$$

Note that the rates $\{R_{13}, R_{31}\}$ and $\{R_{14}, R_{41}\}$ are zero in pairs, because of reversibility. Also, the graph is connected because of the imposed condition of ergodicity; for example, even though no direct transition is possible between 1 and 3, one can be reached from the other via 2.

We can characterize the system furthermore from the graphical perspective. Consider a graph with E edges. From Euler's theorem it must have $C = E - N + 1$ independent cycles in the system [24]. By *cycles* we mean closed loops formed by connected nodes. The system shown in Fig. 2.1 has only one cycle, $\{2, 3, 4\}$, because it has $N = 4$ states, $E = 4$ edges, and therefore $C = 4 - 4 + 1 = 1$ cycle. The cycles will play an important role in our analyses.

2.2 Steady state distribution

If we multiply both sides of Eq. 2.2 from the left by the row vector $\mathbf{1}^T = (1, \ldots, 1)$, we must get a zero on the left because of conservation of probability: $\sum_i \dot{p}_i = 0$. This is consistent with the right hand side because

$$\mathbf{1}^T \mathcal{R} = (0, \ldots, 0). \tag{2.6}$$

This follows from the fact that elements of each column of the rate matrix $\mathcal{R}$ add up to zero by Eq. 2.1. Eq. 2.6 implies $\mathcal{R}$ has a left eigenvector with eigenvalue zero. $\mathcal{R}$ should therefore have a right eigenvector with the same eigenvalue. If we denote such a vector by $\mathbf{p}^s$ we have

$$\mathcal{R}\mathbf{p}^s = (0, \ldots, 0)^T \tag{2.7}$$

i.e. $\mathbf{p}^s$ is a stationary distribution. Because $\mathcal{R}$ has been assumed to be reversible and ergodic, it can be shown that $\mathbf{p}^s$ is unique [24, 86].

For simplicity of notation, we assume in the following that all eigenvalues of $\mathcal{R}$ are non-degenerate (all explicit models in this thesis fall in this category). Let λ_i $(i = 1, \ldots N-1)$ be the i^{th} eigenvalue of $\mathcal{R}$ and $\langle i|$ $(|i\rangle)$ the corresponding left (right) eigenvector. If $\lambda_0 = 0$, from our previous discussion we have

$$\langle 0| = \mathbf{1}^T \quad , \quad |0\rangle = \mathbf{p}^s. \tag{2.8}$$

Both the left and right eigenvectors form a complete basis in their respective vector spaces. Note, however, that $\langle i| \neq |i\rangle^\dagger$, i.e. left and right eigenvectors of an eigenvalue λ_i need not be complex conjugate of the other. This is because the rate matrix $\mathcal{R}$ need not be symmetric. The left and eigenvectors form a complete biorthogonal basis

$$\langle i|j\rangle \propto \delta_{ij}. \tag{2.9}$$

As a result we can rewrite any initial probability vector $|\mathbf{p}(0)\rangle$[1] as the following linear combination

$$|\mathbf{p}(0)\rangle = \sum_i c_i |i\rangle \quad , \quad c_i = \frac{\langle i|\mathbf{p}(0)\rangle}{\langle i|i\rangle}. \tag{2.10}$$

[1]Same as $\mathbf{p}(\mathbf{t} = \mathbf{0})$ of Eq. 2.2.

Eq. 2.10 is useful in studying the time evolution of the system. The probability vector at any time t later is given by

$$\mathbf{p}(t) = \sum_i c_i \, e^{\lambda_i t} \, |i\rangle, \tag{2.11}$$

which follows from the formal solution of Eq. 2.2

$$\mathbf{p}(t) = e^{\mathcal{R}t}\mathbf{p}(0). \tag{2.12}$$

Multiplying Eq. 2.10 from the left by $\langle 0| = \mathbf{1^T}$ and then using the biorthogonality relations (Eq. 2.9) and normalization condition ($\langle 0|0\rangle = \sum_i p_i^s = 1$) we also note that $c_0 = 1$. Furthermore, from Perron-Frobenius theorem [86], it can be shown that all the non-zero eigenvalues have negative real parts: $\mathrm{Re}(\lambda_{i>0}) < 0$. It is then easy to see from Eq. 2.11 that all initial probability vectors relax toward the unique steady state vector $|0\rangle = \mathbf{p}^s$.

2.3 Periodic steady state distribution

So far in our discussion we have assumed the rates R_{ij} to be time-independent. In the context of stochastic pumps, because of the periodic variation of external parameters, these rates become time-dependent. In particular, the rates satisfy, for some common period τ,

$$R_{ij}(t+\tau) = R_{ij}(t). \tag{2.13}$$

In this section we show that the stochastic pumps relax to time-dependent *periodic steady state distributions*,

$$\mathbf{p}^{ps}(t+\tau) = \mathbf{p}^{ps}(t), \tag{2.14}$$

as opposed to time-independent steady state distributions, $\mathbf{p}^s$, discussed in the last section. The superscript *ps* is used to denote values in the periodic steady state distribution throughout this thesis.

The probability distribution of a stochastic pump satisfies the master Eq. 2.2 with a time-dependent periodic rate matrix

$$\mathcal{R}(t+\tau) = \mathcal{R}(t). \tag{2.15}$$

If $\mathbf{p}(0)$ is the initial probability distribution of the pump, at any later time t, the distribution is given by the formal solution

$$\mathbf{p}(t) = \hat{\mathrm{T}} e^{\int_0^t \mathrm{d}t\, \mathcal{R}(t)}, \tag{2.16}$$

where $\hat{\mathrm{T}}$ implies time ordering in the expansion of the following exponential. (The time ordering is needed because the rate matrix may not commute with itself at different times.) In particular, the distribution of the pump after a time period τ is given by

$$\mathbf{p}(\tau) = \mathcal{T}\,\mathbf{p}(0) \quad , \quad \mathcal{T} = \hat{\mathrm{T}} e^{\int_0^\tau \mathrm{d}t\, \mathcal{R}(t)}. \tag{2.17}$$

The matrix $\mathcal{T}$ can be interpreted as the transition matrix whose element T_{ij} fixes the transition probability to state i from state j over a time period τ. Because the rate matrix is periodic, Eq. 2.15, the transition matrix $\mathcal{T}$ is the same for all subsequent periods. Hence, the distribution of the system after n periods i.e. at $t = n\,\tau$, for any non-negative integer n, is given by

$$\mathbf{p}(n\,\tau) = \mathcal{T}^n\,\mathbf{p}(0). \tag{2.18}$$

Because of the assumed ergodicity of the underlying dynamics all the elements of $\mathcal{T}$ are all positive: there is a finite probability to reach any state i from any other state j over a time period. The Perron-Frobenius theorem [86] then implies the existence of a unique distribution $\mathbf{q}$ satisfy

$$\mathcal{T}\mathbf{q} = \mathbf{q}, \tag{2.19}$$

which is reached by the system at each moment $t = n\,\tau$ as $n \to \infty$. In the same limit, the distribution of the system at any other time $n\tau < t \leq (n+1)\,\tau$ is given by

$$\mathbf{p}^{ps}(t) = \hat{\mathrm{T}} e^{\int_0^u \mathrm{d}x\, \mathcal{R}(x)} \mathbf{q} \quad , \quad u = t - n\tau, \tag{2.20}$$

where use has been made of the formal solution, Eq. 2.16, with $t_n = n\,\tau$ as the initial time. Note that $\mathbf{p}^{ps}(t)$ is actually independent of n, namely, the system attains the same distribution $\hat{\mathrm{T}} e^{\int_0^u \mathrm{d}x} \mathbf{q}$ after time $0 < u \leq \tau$ from the beginning of *any* time period (in the limit $n \to \infty$). The stochastic pump relaxes to a time-dependent periodic steady state distribution, given by Eqs. 2.19 and 2.20, after sufficiently long time.

It is interesting to consider the net number of transitions between two states over a complete time period τ in the periodic steady state. This quantifies the amount of pumped current per period in the modeled system. For any pair of states i, j this is given by

$$\Phi_{ij}^{ps} = \int_0^\tau \mathrm{d}t\, J_{ij}^{ps}(t) \quad , \quad J_{ij}^{ps}(t) = R_{ij}(t)\, p_j^{ps}(t) - R_{ji}(t)\, p_i^{ps}(t). \tag{2.21}$$

2.4 Detailed balance

Let us consider the time-independent rates for the moment. In all the Markov models we are going to discuss in this thesis, the metastable states correspond to coarse-grained positions (or configurations) of physical systems. With the exception of open systems considered in Sec. 3.2 and the model in Ch. 6, these systems are (a) closed (with respect to mass exchange), (b) devoid of any external time-dependent force, magnetic field or overall rotation, and (c) in contact with a single thermal reservoir. Under these conditions the models (under their realm of validity) can be shown to satisfy the so-called detailed balance condition (Eq. 2.23 in the following) [76].

A system is said to satisfy *detailed balance* [76] if there is no net current in its steady state, that is,

$$J^s_{ij} = 0 \tag{2.22}$$

for all pairs of states i, j. In terms of the transition rates and steady state probability distribution this implies

$$R_{ij}\, p^s_j = R_{ji}\, p^s_i. \tag{2.23}$$

Note that detailed balance is solely an attribute of the rate matrix $\mathcal{R}$ because the steady state distribution is completely determined by the rates R_{ij}.

There is an equivalent definition of detailed balance expressed by the so-called *Kolmogorov cycle conditions* [77]. This is stated in terms of cycle-*affinities*: the

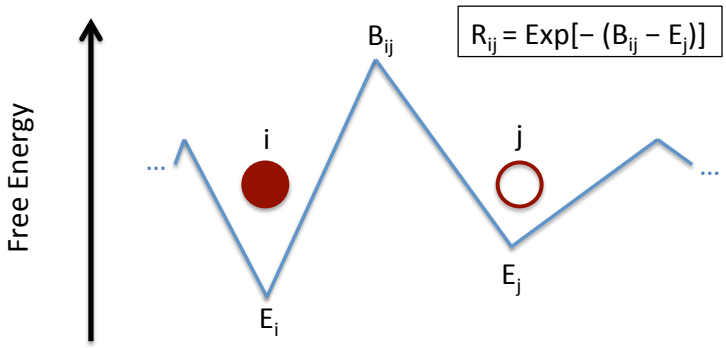


Figure 2.2: **Effective free energies and barriers.** E_i and E_j denote the effective free energies of states i and j, respectively, and B_{ij} denotes the effective free energy barrier between them.

affinity of any cycle $c = \{i, j, k, \ldots m, n\}$ is defined as

$$\mathcal{A}(c) = \ln \frac{R_{ji} R_{kj} \ldots R_{nm} R_{in}}{R_{ij} R_{jk} \ldots R_{mn} R_{ni}} \tag{2.24}$$

i.e., by the natural logarithm of the ratio of two quantities, the product of forward rates along the cycle and that of the reverse rates. E.g., the affinity of the cycle $\{2, 3, 4\}$ in Fig. 2.1 is $\mathcal{A} = \ln\left(R_{32} R_{43} R_{24} / R_{23} R_{34} R_{42}\right)$. The system is said to satisfy detailed balance if the affinity of each cycle in the graph is zero.

The equivalence of the two definitions of detailed balance is not hard to establish. Before we do so, let us introduce the useful notions of state energies (to be denoted by E_i's) and barriers (to be denoted by B_{ij}'s) for a system satisfying detailed balance. If we define the symbols

$$E_i = -\ln p_i^s \tag{2.25}$$

for each state i, the steady state condition, Eq. 2.23, implies

$$R_{ij} = e^{-(B_{ij} - E_j)} \tag{2.26}$$

where (crucially)

$$B_{ij} = B_{ji}. \tag{2.27}$$

It is interesting to note the analogy of Eq. 2.26 to the Arrhenius expression for rate constants per molecule:

$$A\, e^{-E_a/kT}, \tag{2.28}$$

where A is some frequency parameter, E_a is the molecular activation energy, k is Boltzmann constant, and T is the absolute temperature of the environment. Using this analogy we can interpret $(B_{ij} - E_j)$ in Eq. 2.26 to be the activation barrier for transitions from j to i, or equivalently, E_j to be the effective free energy of the state j and B_{ij} the effective free energy barrier between j and i, all in units of kT. (We have chosen the value of the frequency parameter to be unity.) We have illustrated this free energetic description in Fig. 2.2.

We now show that the steady state conditions (Eq. 2.23) lead to the Kolmogorov conditions. If we calculate the affinity of any cycle c and express it in terms of the state energies E_i's and barriers B_{ij}'s introduced above, it is straightforward to see that the affinity satisfies $\mathcal{A}(c) = 0$. Let us illustrate this in terms of the cycle $c = \{2, 3, 4\}$ in Fig. 2.1. We have, from Eq. 2.24,

$$\mathcal{A}(c) = \ln \frac{R_{32}R_{43}R_{24}}{R_{23}R_{34}R_{42}} = \ln \frac{\exp(-B_{32} + E_2)\exp(-B_{43} + E_3)\exp(-B_{24} + E_4)}{\exp(-B_{23} + E_3)\exp(-B_{34} + E_4)\exp(-B_{42} + E_2)}.$$

Because by Eq. 2.27 the B_{ij}'s are symmetric the fraction on the right is equal to unity, and hence $\mathcal{A}(c) = 0$. Thus, the Kolmogorov conditions follow from the steady state conditions.

Let us prove the reverse statement that the Kolmogorov conditions lead to the steady state conditions. Consider a pair of states i, n and a *path* – a sequence of connected states – from i to n, say $l = \{i, j, \ldots, m, n\}$. Then construct the ratio

$$r_{ni}^{l} = \frac{R_{ij} \ldots R_{mn}}{R_{ji} \ldots R_{nm}}. \tag{2.29}$$

This ratio is the same for all paths from i to n. This can be easily seen from the Kolmogorov cycle conditions: two different paths l_1 and l_2 lead to a cycle; the affinity of the cycle is zero from the Kolmogorov conditions; the product of the "forward" rates along the cycle is therefore equal to the product of "reverse" rates; and by a slight rearrangement it then follows that $r_{ni}^{l_1} = r_{ni}^{l_2}$. Because of this condition, we drop the path superscript l in r_{ni}^{l} from now on. We can use the r_{ni}'s to uniquely assign a set of numbers E_i' (up to an additive constant) to each state i: we first assign an arbitrary real number E_1' to state 1, and then use the relations

$$E_i' - E_1' = \ln r_{i1} \tag{2.30}$$

for all other states i. We now claim that the quantities

$$p_i'^s = \frac{e^{-E_i'}}{\mathcal{N}} \quad , \quad \mathcal{N} = \sum_j e^{-E_j'} \tag{2.31}$$

give the steady state probabilities of the system. Clearly they are positive and sum to unity. Furthermore, for any pair of connected states i and j

$$\frac{R_{ij}\, p_j'^s}{R_{ji}\, p_i'^s} = r_{ji} e^{-(E_j' - E_i')} \tag{2.32}$$

where we have used Eqs. 2.29 and 2.31 to arrive at the right hand side. A look at Eq. 2.30 suffices to confirm that the right hand side of Eq. 2.32 is equal to unity

and hence

$$R_{ij}\, p'^{s}_{j} = R_{ji}\, p'^{s}_{i}. \tag{2.33}$$

Hence $\mathcal{R}\mathbf{p}'^{s} = (0, \ldots, 0)^T$, that is, $\mathbf{p}'^{s}$ is the (unique) steady state of the system. The conditions 2.33 are therefore identical to the steady state conditions 2.23, both of which now follow from the Kolmogorov conditions.

Consider now the case of time-dependent transition rates. At any instant of time, the system is said to satisfy *instantaneous detailed balance* if the values of the transition rates at that instant satisfy the conditions of detailed balance. If a stochastic pump, for which the transition rates are periodic functions of time, satisfies instantaneous detailed balance at every instant of a time period, one has

$$R_{ij}(t) = e^{-B_{ij}(t)+E_j(t)} \quad , \quad E_i(t+\tau) = E_i(t) \quad , \quad B_{ij}(t+\tau) = B_{ij}(t). \tag{2.34}$$

In the following chapters, Eq. 2.34 will be considered as the expression of detailed balance for discrete state stochastic pumps.

2.5 Local detailed balance

For systems satisfying detailed balance Eq. 2.26 implies

$$\frac{R_{ij}}{R_{ji}} = e^{-\Delta E} \quad , \quad \Delta E = E_i - E_j. \tag{2.35}$$

i.e., the ratio of the forward and the reverse rates is governed by the corresponding (effective) free energy change. A similar relation holds even when the system does not satisfy detailed balance. In the latter case, the ratio is given by

$$\frac{R'_{ij}}{R'_{ji}} = e^{-\Delta \mu}, \tag{2.36}$$

where $\Delta\mu$ is the change in the total (effective) free energy of the system and the reservoir [24, 25]. We shall refer to Eq. 2.36 as the *local detailed balance* because of its formal resemblance to Eq. 2.35. The individual rates can be parametrized in the following general form:

$$R'_{ij} = \eta\,(1+\epsilon) \quad , \quad R'_{ji} = \eta\,(1-\epsilon), \tag{2.37}$$

where $\eta > 0$ sets the time-scale of the transitions, and

$$-1 < \epsilon < 1 \quad , \quad \frac{1+\epsilon}{1-\epsilon} = e^{\Delta\mu}. \tag{2.38}$$

encodes the effective energetics.

Chapter 3

No-pumping theorem for discrete pumps

In the introductory chapter, we briefly mentioned several different strategies to control artificial molecular machines. In this chapter and the next, we are concerned with stochastic pumping, where external parameters of the system are varied periodically in time. We have already discussed some general behavior of stochastic pumps in Ch. 2. In particular, we have shown that such systems relax to periodic steady states after sufficiently long time time (Sec. 2.3) and an effective free energetic picture can be associated to their dynamics if the conditions of detailed balance are satisfied by them (Sec. 2.4).

Many new theoretical results for stochastic pumps have been reported recently [27–38], some of which were briefly mentioned in the Introduction. We are interested in the particular case of the no-pumping theorem (NPT) which specifies the minimal conditions under which there can be any directed current in the system. This was first derived by Rahav *et al.* [29] by analyzing the algebraic properties of the rate matrices. Shortly thereafter, Chernyak and Sinitsyn [30] showed that the result follows from a quite general "pumping restriction theorem" related to the topology of the stochastic pumps. Horowitz and Jarzynski [78] extended the result to one-dimensional Brownian models. Maes *et al.* [79] obtained and extended the

NPT by considering the embedded Markov chains associated with the stochastic pumps. In the first part of this chapter, we are going to give an alternative and simpler than the original proof of the NPT using an elementary graph theoretic construction based on our discussion in Sec. 2.1.

All the works mentioned in the last paragraph are based on *closed* stochastic pumps where no mass exchange is present between the system and its environment. This is in contrast to *open* stochastic pumps where particles flow among various chemical reservoirs through the system. Interestingly, the NPT for closed stochastic pumps can be extended to the open stochastic pumps; this is the topic of our discussions in the second part of this chapter. There is an important difference, however, between the NPT for closed pumps and its extended form for the open pumps that we are going to present: for the former, the NPT gives the conditions under which there is no net integrated flow of *probability*, whereas for the latter, the extended NPT gives the conditions under which there is no net integrated flow of *particles*. [1]

In the following, we first discuss the case of closed pumps in Sec. 3.1 borrowing heavily from the machinery developed in the last chapter. We illustrate our proof of the NPT with a simple example (Sec. 3.1.1) and then present the general proof

[1]Note that the simple condition that all the chemical reservoirs have the same potential always is neither necessary nor sufficient to guarantee that all integrated currents are zero. Even with different chemical potentials for different reservoirs, one can have zero integrated currents with some simple restrictions, Eq. 3.27. On the other hand, system parameters can be varied to generate a flow of particles (within the system) even when all the chemical potentials are all the same.

(Sec.3.1.2). The case of open pumps is considered in Sec. 3.2. We first develop the kinetic equations which describe the particle dynamics in these systems (Sec. 3.2.1), then illustrate the proof of the extended NPT with a simple example (Sec. 3.2.2), and finally give the corresponding general proof (Sec. 3.2.3).

3.1 Closed stochastic pumps[2]

Consider an N-state system which follows the dynamics described in Sec. 2.1 with the master equation

$$\dot{p}_i(t) = \sum_{j \neq i} J_{ij}(t), \tag{3.1}$$

where

$$J_{ij}(t) = R_{ij}(t)p_j(t) - R_{ji}(t)p_i(t) \quad , \quad J_{ji}(t) = -J_{ij}(t). \tag{3.2}$$

The rates $R_{ij}(t)$ periodic functions of time (Eq. 2.13)

$$R_{ij}(t+\tau) = R_{ij}(t) \tag{3.3}$$

and are assumed to satisfy detailed balance at each moment. As a result, these rates can be expressed as (Eq. 2.34)

$$R_{ij}(t) = e^{-B_{ij}(t)+E_j(t)} \quad , \quad E_i(t+\tau) = E_i(t) \quad , \quad B_{ij}(t+\tau) = B_{ij}(t), \tag{3.4}$$

where the $E_i(t)$'s and $B_{ij}(t)$'s are the time-dependent effective free energies of the states and the barriers among them.

Under these dynamics the system relaxes to a periodic steady state, $\mathbf{p}^{ps}(t+\tau) =$

[2]This section is based on Ref. [90].

$\mathbf{p}^{ps}(t)$ (Sec. 2.3), and we are interested in the integrated currents (Eq. 2.21)

$$\Phi_{ij}^{ps} = \int_{\tau} \mathrm{d}t \, J_{ij}^{ps}(t) \quad , \quad \Phi_{ji}^{ps} = -\Phi_{ij}^{ps}.$$

The integrated currents characterize the amount of directed motion achieved in the modeled systems. In particular, $\Phi_{ij}^{ps} \neq 0$ for some pair of states i and j indicates a net flow of probability, over each period of pumping, between states i and j. Conversely, if $\Phi_{ij}^{ps} = 0$ for all pairs of states, then the probability current may slosh back and forth, so to speak, but there is no net circulation of current.

The no-pumping theorem (NPT) now asserts that *if* either all the state energies $\{E_i\}$ or all the barriers $\{B_{ij}\}$ are kept fixed in time during the pumping, *then* the integrated probability current is zero along all edges, i.e.

$$\Phi_{ij}^{ps} = 0 \ \text{ for all pairs } \ (i,j). \tag{3.5}$$

Consequently one must vary at least one state energy E_i *and* at least one barrier B_{ij} to produce directed probability currents in the periodic steady state.

The case of fixed state energies $\{E_i\}$ and time-dependent barriers $\{B_{ij}(t)\}$ is straightforward: the system relaxes to a *fixed* steady-state distribution $p_i^s = \exp(-E_i)$ [29], which is also the periodic steady state in this case. Eqs. 3.2 and 2.34 then imply

$$\begin{aligned} J_{ij}^{s}(t) &= e^{-[B_{ij}(t)-E_j]}e^{-E_j} - e^{-[B_{ji}(t)-E_i]}e^{-E_i} \\ &= e^{-B_{ij}(t)} - e^{-B_{ji}(t)} = 0 \end{aligned}$$

for all (i,j). Thus, the instantaneous currents vanish, and therefore so do the integrated currents. In the following sections we focus on the less obvious case of

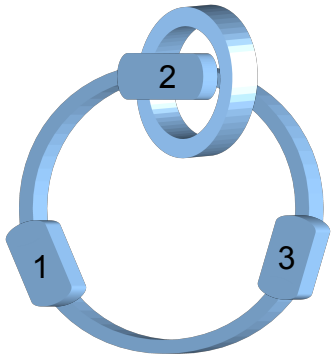


Figure 3.1: **A model [2]catenane.** This depicts the [2]catenane system studied in Ref. [13]. It consists of two unequal rings, the smaller ring having three binding sites on the bigger ring. Because of thermal fluctuations from the environment the smaller ring makes random transitions among the three binding sites.

fixed barrier energies $\{B_{ij}\}$, but periodically pumped state energies, $\{E_i(t)\}$.

3.1.1 Illustration

In this subsection, we illustrate our proof of the NPT with a simple model inspired by the experimental studies in Ref. [13]. Generalization to more complicated models is presented in the next subsection.

The first system used in Ref. [13] was a [2]catenane [3] – a mechanically interlocked complex of two molecular rings (shown schematically in Fig. 3.1). The smaller ring had three binding sites on the bigger ring, where the binding affinity of any site was determined by the number and strength of hydrogen bonds between the two rings in the corresponding configuration. In an effort to rotate the smaller ring

[3] An [n]catenane is a hydrocarbon having n rings "connected in the manner of links of a chain, without a covalent bond." [80]

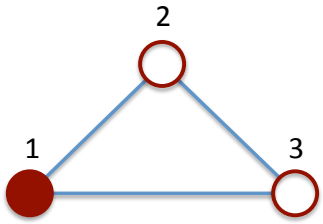


Figure 3.2: **Graphical representation of an N = 3 state system with a single cycle** $\{1, 2, 3\}$**.** It is a simple discrete-state model of the [2]catenane pictured in Fig. 3.1

unidirectionally along the bigger ring, the authors considered a sequential (relative) strengthening of binding affinities, in sequence $1 \to 2 \to 3 \to 1$ corresponding to Fig. 3.1, by photo-chemical, chemical and thermal means. Contrary to intuition, this strategy did not work – even though the smaller ring shifted its position to the maximally binding site each time, there was no directional bias in its motion. This is actually a manifestation of the NPT: cyclic variation of the binding affinities corresponds to a periodic variation of state energies $\{E_i(t)\}$ while keeping the barriers fixed $\{B_{ij}\}$, and the NPT forbids any non-zero integrated current in this situation. We illustrate this in the following with the help of a simple stochastic pump modelling the [2]catenane in Fig. 3.1.

Consider the system in Fig. 3.2, with $N = 3$ states, $E = 3$ edges and a single cycle $c = \{1, 2, 3\}$, and assume that all the B_{ij}'s are fixed in time while one or more of the $E_i(t)$'s are varied periodically. Combining equation 3.1 with the antisymmetry

of $J_{ij}(t)$'s, Eq. 3.2, we have

$$\begin{aligned} \dot{p_1}(t) &= J_{12}(t) - J_{31}(t) \\ \dot{p_2}(t) &= J_{23}(t) - J_{12}(t) \\ \dot{p_3}(t) &= J_{31}(t) - J_{23}(t) \end{aligned} \tag{3.6}$$

From our discussion in Sec. 2.3, the system eventually relaxes to a periodic steady state with no net change in state probabilities over a time period τ, i.e. $\int_\tau \mathrm{d}t\, \dot{p}_i^{ps}(t) = 0$ for all i. As a result

$$\begin{aligned} 0 &= \Phi_{12}^{ps} - \Phi_{31}^{ps} \\ 0 &= \Phi_{23}^{ps} - \Phi_{12}^{ps} \\ 0 &= \Phi_{31}^{ps} - \Phi_{23}^{ps} \end{aligned} \tag{3.7}$$

where we have integrated Eq. 3.6 over one period of the periodic steady state. Since normalization implies $\sum_i \dot{p}_i = 0$, only 2 of the 3 equations in either Eq. 3.6 or Eq. 3.7 are independent. The solution of Eq. 3.7 therefore contains a free parameter:

$$\Phi_{12}^{ps} = \Phi_{23}^{ps} = \Phi_{31}^{ps} = \Phi. \tag{3.8}$$

These results are easy to understand: the currents along all the edges are equal because they all belong to the same cycle, which is the only cycle in the graph. (This intuition has been formalized and generalized to arbitrary graphs in Ref. [30] to derive the so-called pumping restriction theorem (PRT).)

Detailed balance implies further constraints. From Eqs. 3.2 and 2.34 we have

$$e^{B_{ij}}\, J_{ij}(t) = e^{E_j(t)} p_j(t) - e^{E_i(t)} p_i(t).$$

Summing both sides of this equation over the edges along the cycle $\{1, 2, 3\}$ we get

$$e^{B_{12}}\, J_{12}(t) + e^{B_{23}}\, J_{23}(t) + e^{B_{31}}\, J_{31}(t) = 0. \tag{3.9}$$

We have deliberately omitted the superscript ps to indicate that the above relation holds whether or not the system has reached the periodic steady state. Indeed, Eq. 3.9 remains true even if the external driving is not periodic, and even if the barriers are time-dependent. (A generalized form of Eq. 3.9 for arbitrary graphs was used in Ref. [32] to derive a pumping-quantization theorem (PQT) for integrated probability currents.)

Returning to the periodic steady state with fixed barriers $\{B_{ij}\}$, we integrate Eq. 3.9 over one period τ to get

$$e^{B_{12}}\,\Phi_{12}^{ps} + e^{B_{23}}\,\Phi_{23}^{ps} + e^{B_{31}}\,\Phi_{31}^{ps} = 0. \tag{3.10}$$

Combined with Eq. 3.8 this gives

$$\left(e^{B_{12}} + e^{B_{23}} + e^{B_{31}}\right)\Phi = 0. \tag{3.11}$$

Hence $\Phi = 0$, and all the integrated probability currents Φ_{ij}^{ps}'s in the system are zero.

3.1.2 General Proof

Consider a connected graph G with N vertices and E edges. As before, we assume that the $2E$ transition rates satisfy detailed balance at all times, hence they can be written in the form $R_{ij} = e^{-(B_{ij}-E_j)}$ with $B_{ij} = B_{ji}$. We now imagine that the state energies $E_i(t)$ are varied periodically with time, while the barriers energies B_{ij} are held fixed. After the system has reached a periodic steady state,

$\mathbf{p}^{ps}(t+T) = \mathbf{p}^{ps}(t)$, integration of Eq. 3.1 over one period yields

$$\sum_{j \neq i} \Phi_{ij}^{ps} = 0 \quad \text{for all } i. \tag{3.12}$$

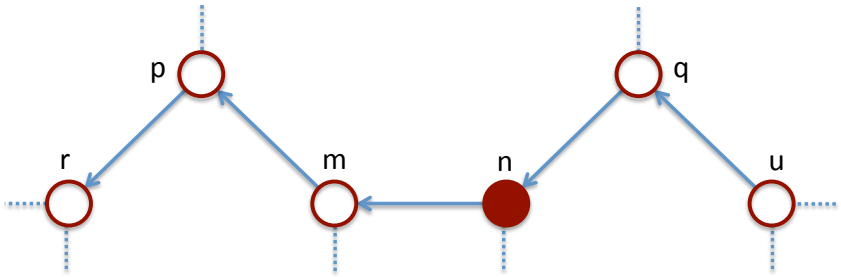
As with Eq. 3.7 only $(N-1)$ of these N equations are independent. Moreover, Eq. 3.12 implies that if $\Phi_{ij}^{ps} > 0$ for a connected pair of states (i,j), then there must exist at least one other vertex k such that $\Phi_{ik}^{ps} < 0$, as the flow of probability into state i must be balanced by the flow of probability out of that state.

As in our illustration, detailed balance implies further constraints. Summing over, and then integrating with time, the instantaneous currents along the edges of any cycle $c = \{i_1, \ldots, i_M\}$ we get (compare with equation 3.10)

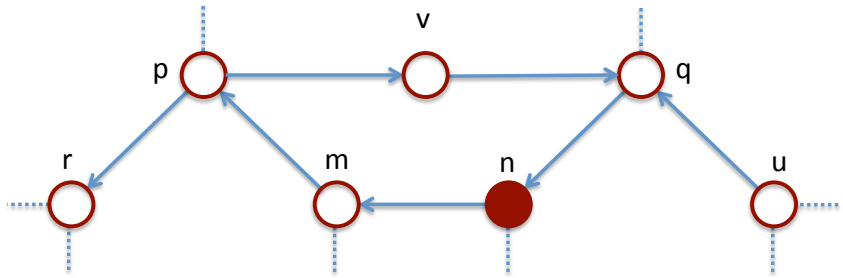
$$\sum_{j=1}^{M} e^{B_{i_j i_{j+1}}} \Phi_{i_j i_{j+1}}^{ps} = 0 \quad , \quad i_{M+1} \equiv i_1. \tag{3.13}$$

This implies that if one edge (i_j, i_{j+1}) of c has $\Phi_{i_j,i_{j+1}}^{ps} > 0$ then there must exist at least one other edge (i_k, i_{k+1}) in c with $\Phi_{i_k i_{k+1}}^{ps} < 0$. Thus, for any cycle, the non-zero $\Phi_{i_l i_{l+1}}^{ps}$'s cannot all have the same sign. We now prove that Eqs. 3.12 and 3.13 jointly imply $\Phi_{ij}^{ps} = 0$ for all edges. We establish this below by contradiction, assuming the existence of at least one edge (m,n) with $\Phi_{mn}^{ps} > 0$.

To formulate our argument, let us introduce the following convenient construction on G. Along every edge, say (r,s), with non-zero Φ_{rs}^{ps}, we draw an arrowhead indicating the positive direction of the integrated probability current, as shown in Fig. 3.3 (a). By assumption, G contains at least one arrow, pointing from n to m. Eq. 3.12 then implies the existence of another edge (p,m), such that $\Phi_{mp}^{ps} < 0$, or equivalently, $\Phi_{pm}^{ps} > 0$. Thus we must have another arrow pointing from m to some



(a)



(b)

Figure 3.3: **Part of an N state graph with arbitrary topology.** (a) Illustration of the construction of arrows. An arrow pointing along an edge, e.g. from n to m, indicates a positive integrated probability current from n to m, $\Phi^{ps}_{mn} > 0$. (b) One of the possible cycles, $\{m, p, v, q, n\}$, with all arrows pointing the same way.

$p \neq n$. Similarly there must be another arrow from some $q \neq m$ to n, to prevent the depletion of probability from state n. Refer to Fig. 3.3 (a) for illustration.

Consider now the set $\mathcal{D}_m$ of all vertices that can be reached from m by following the arrows. In Fig. 3.3 (a) $\mathcal{D}_m = \{p, r, \ldots\}$. Consider also set $\mathcal{S}_n$ of all vertices from which n can be reached by following the arrows. In Fig. 3.3 (b) $\mathcal{S}_n = \{q, u, \ldots\}$. These two sets must have at least one element in common, otherwise there will be a constant drainage of probability from $\mathcal{S}_n$ to $\mathcal{D}_m$ which is inconsistent with a periodic steady state. Let v denote this common element.

The existence of a common element has an interesting consequence. Starting from state m, we can reach state v by following the arrows (since $v \in \mathcal{D}_m$), and from there we can reach state n by continuing to follow arrows (since $v \in \mathcal{S}_n$). Since an arrow points from n to m, we conclude that there exists a cycle $\{m, \ldots, v, \ldots, n\}$ consisting of edges with arrows all pointing in the same direction $\{m \to \ldots \to v \to \ldots \to n \to m\}$. By construction, the $\Phi^{ps}_{i_j i_{j+1}}$'s along this cycle are all positive. One such cycle $\{m, p, v, q, n\}$ is shown in Fig. 3.3 (b).

However, this contradicts Eq. 3.13. We conclude that the existence of a non-zero Φ^{ps}_{mn} is inconsistent with our starting assumptions, and this completes our proof.

In summary, the proof is based on the idea that if a non-zero integrated current is generated along some edge of the graph, then this edge must be part of a closed loop along which probability is conveyed in one direction: all the $\Phi^{ps}_{i_j i_{j+1}}$'s along the cycle have the same sign. This in turn is inconsistent with the assumption of detailed balance with fixed energy barriers (which gives Eq. 3.13).

3.2 Open stochastic pumps

Closed stochastic pumps, discussed so far, have a fixed number of constituents. Now we consider the case of open stochastic pumps, where the system is connected to several chemical reservoirs, thereby having a variable number of constituents. E.g. consider a collection of quantum dots that are connected to each other and also to several electrodes with time-dependent electrical potentials; in the limit of high temperature the electronic transport through this collection is essentially Markovian and therefore the whole setup is an open stochastic pump [81]. The rest of this chapter is devoted to an extension of the NPT to such systems.

One particular way to arrive at an open stochastic pump is to connect a closed stochastic pump to several chemical reservoirs. This passage is somewhat analogous to the passage from the canonical to the grand canonical description in the equilibrium statistical mechanics: there is a large accompanied enhancement in the resulting state space. A complete stochastic description in this enlarged state space, in terms of an appropriate master equation, is both cumbersome and unnecessary for our purpose. We therefore rely on a kinetic level of description, instantaneous (ensemble) average occupation numbers of the discrete states of the corresponding closed system serving as the dynamical variables, in the same way that chemical reactions are described by deterministic equations in lieu of stochastic means. [4] In the following, we first derive the appropriate kinetic equations (Sec. 3.2.1), then

[4] A proof of the extended NPT in the complete framework has been recently obtained by Nikolai Sinitsyn (private communications). Our treatment is independent of his.

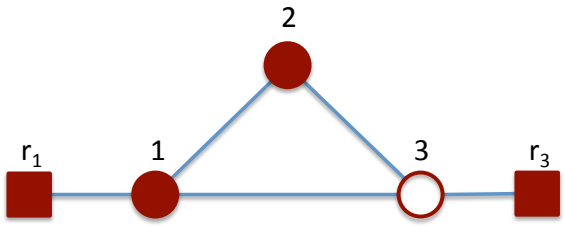


Figure 3.4: **A simple open stochastic system with $N = 3$ states and $R = 2$ reservoirs.** It is an open system extension of the system in Fig. 3.2. The boxes represent the reservoirs.

illustrate our proof with a simple example (Sec. 3.2.2), and finally, present the general analysis (Sec. 3.2.3).

3.2.1 Kinetic equations

The kinetic description of an open system is similar to the complete stochastic description of the corresponding closed system in many ways: we still have a finite number of distinct states; in place of occupation probability of a state we have its average occupation number; and in place of the probability currents between the states we have particle currents between them. The presence of the reservoirs, however, makes a crucial difference: there is no equivalent of the normalization condition, $\sum_i p_i = 1$, in the open systems.

Let $n_i(t)$ denote the average occupation number of any state i, and r_i denote the chemical reservoir connected to it with chemical potential μ_i. If $\mu_i = \mu_j$, then the associated states i and j can be considered to be connected to the same reservoir. We have depicted the setup with a simple open system in Fig 3.4. This is an open

system generalization of the closed system depicted in Fig. 3.2. The reservoirs are denoted by boxes. They have been drawn filled because the density of particles in a reservoir is unaffected by the number of particles it may exchange with the system.

Let us now introduce the transition rates of a single particle. The conditional rate of transition from a state j to a connected state i is denoted by R_{ij} [5], with the usual assumptions of reversibility, ergodicity and detailed balance, $R_{ij} = e^{-B_{ij}+E_j}$. The conditional rate of transition from the reservoir r_i to state i and from i to r_i are denoted by α_i and β_i, respectively, with the detailed balance dictating

$$\alpha_i = e^{-(B_i-\mu_i)} \quad , \quad \beta_i = e^{-(B_i-E_i)}, \tag{3.14}$$

where B_i is the effective free energy barrier between i and r_i (see Fig. 3.5 for an illustration). These rates and free energy parameters are well defined if the particles do not interact with each other. We can admit only those interactions which can be effectively incorporated by simply making the energy parameters dependent on occupation numbers.

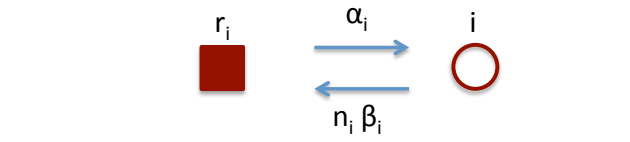
These transitions lead to particle currents and therefore time evolution of average occupation numbers $n_i(t)$. The instantaneous net current from j to i is given by

$$J_{ij}(t) = R_{ij}\, n_j(t) - R_{ji}\, n_i(t) \quad , \quad J_{ji}(t) = -J_{ij}(t), \tag{3.15}$$

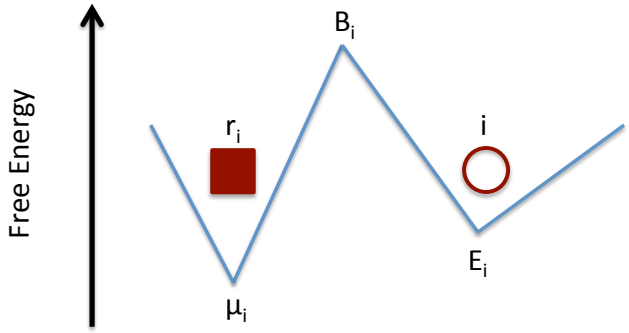
and that from a reservoir r_i to state i by

$$J_i(t) = \alpha_i - n_i(t)\, \beta_i. \tag{3.16}$$

[5] We use the same notation as for a closed systems because we do not return to the latter in the rest of this chapter.



(a)



(b)

Figure 3.5: **Particle exchange with a reservoir.** (a) α_i and $n_i(t)\beta_i$ denote the instantaneous average particle currents in directions $r_i \to i$ and $i \to r_i$, respectively. (b) Under the assumption of (local) detailed balance the free energy diagram of Fig. 2.2 can be extended to include the reservoirs. μ_i is the chemical potential of the reservoir r_i, E_i is the effective free energy of state i, and B_i is the effective free energy barrier between them.

$n_i(t)$'s evolve in time according to the kinetic equations

$$\dot{n}_i(t) = \sum_{j \neq i} J_{ij}(t) + J_i(t). \tag{3.17}$$

(Compare Eq. 3.1 for a closed system.)

To make connection to the framework of stochastic pumps we now consider periodic variation of the elements $\{E_i, \mu_i, B_{ij}, B_i\}$. We assume that the system relaxes to a periodic steady state, $n_i^{ps}(t+\tau) = n_i^{ps}(t)$, just like the closed systems considered before [6]. We then consider the net integrated particle current over a time period τ from one state j to another state i, Φ_{ij}^{ps}, and from a reservoir r_i to the associated state i, Φ_i^{ps}:

$$\Phi_{ij}^{ps} = \int_\tau \mathrm{d}t\, J_{ij}^{ps}(t) \quad , \quad \Phi_i^{ps} = \int_\tau \mathrm{d}t\, J_i^{ps}(t). \tag{3.18}$$

The extended NPT states that all integrated currents are zero if either

(i) all the energies and chemical potentials $\{E_i, \mu_i\}$ are held fixed, all chemical potentials are the same, and only the barriers, $\{B_{ij}(t), B_i(t)\}$, are varied periodically in time, or

(ii) all the barriers $\{B_{ij}, B_i\}$ are kept fixed, energies and chemical potentials $\{E_i(t), \mu_i(t)\}$ are varied periodically in time, and all the integrals $\int_\tau \mathrm{d}t\, e^{\mu_i(t)}$ are the same.

Note that the last condition in case (ii), namely the equality of all the integrals $\int_\tau \mathrm{d}t\, e^{\mu_i(t)}$, while restrictive, can still include realistic situations different from the

[6] This assertion actually needs an independent proof from that of Sec. 2.3 because of the reservoir terms in Eq. 3.17.

trivial case where all the $\mu_i(t)$'s are equal. Consider for example the case of the quantum dots discussed in the beginning of this section; if all the electrical potentials are varied sinusoidally in time with the same amplitude and period but with arbitrary phase difference, the above condition is still satisfied.

The relation of the extended NPT to the original NPT is essentially obvious. One considers the limits $B_i \to \infty$ in Eq. 3.14 so that all $J_i = 0$ in Eq. 3.17. Then the fraction of total particles in states i, $f_i = n_i / \sum_j n_j$, can be treated as probabilities satisfying the same master equation as p_i's in Eq. 2.2. In the non-interacting case, one may alternatively consider the dynamics of a single particle and the corresponding probability distribution; the latter follows the same equation as the f_i's (i.e. Eq. 2.2). It is also important that the limit $B_i \to \infty$ be taken at the level of the evolution equation, and not the periodic steady state; this is because the two limits $t \to \infty$ and $\{B_i\} \to \infty$ do not commute with each other at the level of the solution.

Case (i) above is easy to analyze. In absence of any variation in the free energies E_i and the chemical potentials μ_i, the numbers n_i relax to the equilibrium distribution $n_i^{eq} = e^{\mu_i - E_i}$ (note that all μ_i's are assumed to be equal in this case) and all the instantaneous currents, and hence also the integrated currents are zero:

$$\begin{aligned} J_{ij}^{eq} &= e^{-B_{ij}+E_j} e^{\mu_j - E_j} - e^{-B_{ij}+E_i} e^{\mu_i - E_i} = 0, \\ J_i^{eq} &= e^{-B_i+\mu_i} - e^{\mu_i - E_i} e^{-B_i+E_i} \qquad\qquad = 0. \end{aligned} \tag{3.19}$$

Effectively, the system is immersed in a single chemical bath, and it relaxes to the corresponding equilibrium distribution. In the next section we concentrate on the

less obvious case of varying free energies and chemical potentials, i.e. case (ii).

3.2.2 Illustration

Consider the system depicted in Fig. 3.4. We assume that the barriers $\{B_{ij}, B_i\}$ are kept fixed in time while one or more of the effective free energies and chemical potentials $\{E_i(t), \mu_i(t)\}$ are varied periodically in time. Furthermore, the two chemical potentials are assumed to satisfy

$$\int_\tau \mathrm{d}t\, e^{\mu_1(t)} = \int_\tau \mathrm{d}t\, e^{\mu_3(t)}. \tag{3.20}$$

From Eq. 3.17 we have the following equations for the time evolution of the average occupation numbers

$$\begin{aligned} \dot{n_1} &= J_{12}(t) - J_{31}(t) + J_1(t) \\ \dot{n_2} &= J_{23}(t) - J_{12}(t) \\ \dot{n_3} &= J_{31}(t) - J_{23}(t) + J_3(t). \end{aligned} \tag{3.21}$$

Under the assumption of periodic steady state, the net change in the average occupation number over a time period τ is zero: $\int_\tau \dot{n}_i^{ps} \mathrm{d}t = 0$. Hence, by integrating Eq. 3.21 over τ in periodic steady state we get

$$\begin{aligned} 0 &= \Phi_{12}^{ps} - \Phi_{31}^{ps} + \Phi_1^{ps} \\ 0 &= \Phi_{23}^{ps} - \Phi_{12}^{ps} \\ 0 &= \Phi_{31}^{ps} - \Phi_{23}^{ps} + \Phi_3^{ps} \end{aligned} \tag{3.22}$$

Note that, unlike the case of closed systems (Eqs. 3.6 and 3.7), all three equations in either Eq. 3.21 or Eq. 3.22 are independent.

We can derive further restrictions on the integrated currents from the conditions of detailed balance, Eqs. 2.34 and 3.14, and the additional condition in Eq. 3.20. From Eqs. 3.15, 2.34 and 3.14

$$e^{B_{ij}} J_{ij}(t) = e^{E_j(t)} n_j(t) - e^{E_i(t)} n_i(t).$$

Summing both sides over all three edges of the cycle $c = \{1, 2, 3\}$ we get

$$e^{B_{12}}\, J_{12}(t) + e^{B_{23}}\, J_{23}(t) + e^{B_{31}}\, J_{31}(t) = 0, \tag{3.23}$$

which is the analogue of Eq. 3.9. As in the analysis of the closed systems (Sec. 3.1.1), we have deliberately omitted the superscript "ps" to indicate that Eq. 3.23 holds whether or not the system has reached the periodic steady state. Indeed, Eq. 3.23 remains true even if the external driving is not periodic, and even if the barriers are time-dependent.

In the periodic steady state with fixed barriers $\{B_{ij}, B_i\}$, we can integrate Eq. 3.23 over one period τ to get

$$e^{B_{12}} \Phi_{12}^{ps} + e^{B_{23}} \Phi_{23}^{ps} + e^{B_{31}} \Phi_{31}^{ps} = 0. \tag{3.24}$$

(compare Eq. 3.10). Eqs. 3.14 and 3.20 impose one further restriction on the integrated currents. We can combine Eqs. 3.18, 3.15, 2.34, 3.14, and 3.20 to derive

$$e^{B_3} \Phi_3^{ps} - e^{B_1} \Phi_1^{ps} = e^{B_{31}} \Phi_{31}^{ps}. \tag{3.25}$$

In the present case we have 5 integrated currents – $\{\Phi_{12}^{ps}, \Phi_{23}^{ps}, \Phi_{31}^{ps}, \Phi_1^{ps}, \Phi_3^{ps}\}$ – and we have derived 5 linear homogeneous equations for them – Eqs. 3.22, 3.24 and 3.25. It is easy to verify that these equations imply that all the integrated currents

are zero:

$$\Phi_{12}^{ps} = \Phi_{23}^{ps} = \Phi_{31}^{ps} = \Phi_{1}^{ps} = \Phi_{3}^{ps} = 0.$$

3.2.3 General Proof

Consider now a general open system with N states, E edges, and R reservoirs. We assume the general detailed balance conditions

$$R_{ij} = e^{-(B_{ij}-E_j)} \quad , \quad \alpha_i = e^{-(B_i-\mu_i)} \quad , \quad \beta_i = e^{-(B_i-E_i)} \tag{3.26}$$

where we vary one or more of the state energies and chemical potentials $\{E_i(t), \mu_i(t)\}$ periodically in time with the constraints

$$\int_\tau \mathrm{d}t\, e^{\mu_i(t)} = \int_\tau \mathrm{d}t\, e^{\mu_j(t)} \text{ for all } i, j, \tag{3.27}$$

and keep the barriers $\{B_{ij}, B_i\}$ fixed. The average occupation numbers $n_i(t)$ satisfy Eq. 3.17 and are assumed to relax to a periodic steady state, $n_i^{ps}(t+\tau) = n_i^{ps}(t)$, after sufficient time.

Integrating Eq. 3.17 over a time period τ in periodic steady state we get

$$\Phi_i^{ps} + \sum_{j \neq i} \Phi_{ij}^{ps} = 0 \tag{3.28}$$

for each state i. Eq. 3.28 has an interesting implication: If there is one non-zero integrated current, there must also be another non-zero integrated current, with opposite sign, associated with the same state; this is to ensure that the positive incoming flow of particles due to one current is balanced by the positive outgoing flow due to the other. To facilitate the proof we then carry out an exercise related to this observation. If there is any non-zero integrated current, we add an arrowhead

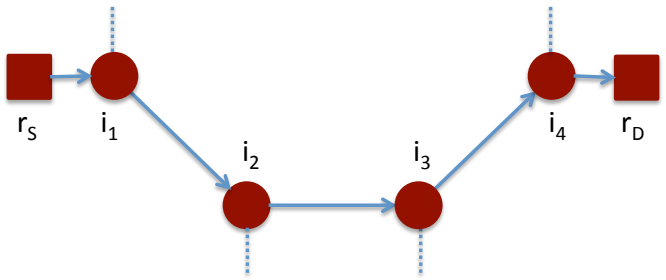


Figure 3.6: **Flow from a source to a drain reservoir.** A sample directed path is shown. Note that, the two reservoirs r_S and r_D need not be different.

to the corresponding edge to show the direction of the flow. E.g. $\Phi_i^{ps} > 0$ leads to an arrow from r_i to i; $\Phi_{ij}^{ps} > 0$ leads to an arrow from j to i; and so on. According to the discussed implication of Eq. 3.28 there cannot be just one arrow in the system.

If there is no arrow associated with any reservoir, all non-zero integrated currents in the system, if any, can flow only inside the systems, and hence along just cycles to ensure return to the same average occupation number at each state after each period. We already considered this case in the context of closed systems in Sec. 3.1.2 and derived NPT using detailed balance conditions. On the other hand, if there is an arrow with a reservoir, there must be a reservoir with an opposite arrow. More precisely, a reservoir with an arrow pointing from it ("source") to the system implies the existence of a reservoir with an arrow pointing to it ("drain") from the system; this is needed to ensure return to the same number of particles in the system after each period. Furthermore, each source reservoir must be connected to a drain reservoir through a directed path; one should be able to reach some drain from any source by following the arrows. Let us assume one such directed path to be

$\{r_S \to i_1 \to i_2 \to \ldots \to i_{n-1} \to i_n \to r_D\}$, r_S and r_D being source and drain reservoirs, respectively; see Fig. 3.6 for an illustration. Note that the two reservoirs r_S and r_D need not be different as two states may be connected to the same reservoir. Using the conditions of detailed balance, Eq. 3.26 and the definition of integrated currents, Eqs. 3.18, 3.15, 3.16, one can derive the following

$$e^{B_{i_1}} \Phi^{ps}_{i_1} + \sum_{j=1}^{n-1} \left(e^{B_{i_{j+1} i_j}} \Phi^{ps}_{i_{j+1} i_j} \right) + e^{B_{i_n}} \left(-\Phi^{ps}_{i_n} \right) = \int_\tau \mathrm{d}t \, e^{\mu_{i_1}(t)} - \int_\tau \mathrm{d}t \, e^{\mu_{i_n}(t)} \qquad (3.29)$$

The right hand side of Eq. 3.29 vanishes because of the conditions 3.27. But each term in the above summation is supposed to be positive: all integrated current – $\Phi^{ps}_{i_1}$, $\{\Phi^{ps}_{i_{j+1} i_j}\}$, and $\left(-\Phi^{ps}_{i_n}\right)$ – are positive because of the arrows on them (see Fig. 3.6), and the coefficients, being exponential of real numbers, are positive too. The only resolution to this inconsistency is to have no source or drain reservoirs at all. As NPT is already proven for this case, Sec. 3.1.2, our proof of extended NPT is complete.

Chapter 4

Hybrid pumps

---

[1]Until now, discrete state Markov processes have been considered to be adequate for the description of molecular machines. They are assumed to reside in one of their metastable states and make random instantaneous transitions driven by thermal fluctuations. In reality, a transition between two metastable states involves mechanical motion, and therefore can not be instantaneous. Recent experiments [18, 82] using time-resolved vibrational spectroscopy to study the movement of a molecular shuttle between two docking stations, provide evidence that this motion is described more accurately as a rapid, one-dimensional random walk than as an instantaneous jump. This motivates us to introduce a new class of models of molecular machines in which the system makes diffusive (rather than sudden) transitions.

The incorporation of diffusive dynamics during the transitions introduces an essential change in the state space of the system. Instead of being composed of discrete points, as in Fig. 2.1, the state space itself is now represented by both these discrete points and the continuous line segments joining them. The metastable states still correspond to the discrete points whereas the mechanical pathways of diffusion between the metastable states correspond to the continuous line segments.

[1]This chapter is based on Ref. [91].

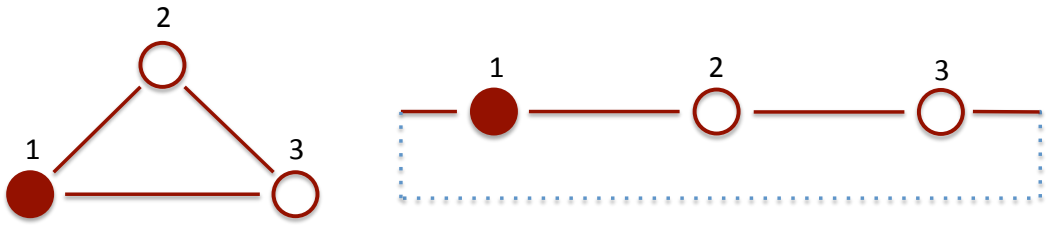
Because of this hybrid nature of state space in the resulting class of models, we will refer to them as *hybrid* models. The purpose of the present chapter is to introduce a consistent framework for this class of models.

How does one demonstrate the validity and utility of such a model? As a first step, we choose to establish the no-pumping theorem (NPT) of the previous chapter in this new framework. This will not only illustrate the feasibility and usability of the model, but also provide a more refined theoretical justification of the experimentally observed no-go conditions in Ref. [13].

A general description of the framework is notationally complicated, so we defer it till the end of the chapter. First we consider a simple hybrid model in Sec. 4.1, motivated by the [2]catenane complex of Ref. [13]. In Sec. 4.2 we discuss the conditions of detailed balance in this model; as in the previous models, we see a natural effective free energetic picture emerging out of these conditions. Sec. 4.3 gives the statement and proof of the NPT. Finally, in Sec. 4.4 we consider the general discussion of the model.

4.1 Hybrid Model of a [2]catenane

The [2]catenane we are going to deal with was introduced schematically in Fig. 3.1. It has two interlocked rings, the smaller ring having three binding sites on the bigger ring. Because of thermal fluctuations the small ring makes random transitions among these states. In the hybrid framework we assume these transitions to be diffusive, namely, the small ring can jump out of a metastable state, perform



(a) (b)

Figure 4.1: **Hybrid model of [2]-catenane.** (a) The [2]catenane is represented by a cycle with $n_s = 3$ stations and $n_T = 3$ tracks. (b) A linear representation of the same model. Periodicity in the state space is indicated by the dotted line.

diffusive motion along the arm of the large ring, and then gets captured by one of the two metastable states at the ends.

We can represent the hybrid models graphically as in the case of discrete state models. The metastable states are represented by the nodes of a graph. The edges now, however, serve as diffusive pathways for the finite time transition dynamics. Thus, in addition to providing the allowed transitions, the edges become part of the system state space. An appropriate graph for the [2]catenane is given in Fig. 4.1.

To facilitate the discussion we designate the binding sites as the *stations*, and the edges, or the diffusive pathways, as the *tracks*. Whenever in a station, the small ring, or the system henceforth, makes jumps to the nearest ends of the adjacent tracks at certain rates. Whenever in a track, the system performs diffusive motion with the stations at the ends acting as sinks. In the [2]catenane model shown in Fig. 4.1(a) there are three stations $i = 1, 2, 3$. There are three tracks, each track being denoted by the same index as the preceding station in the clockwise sense.

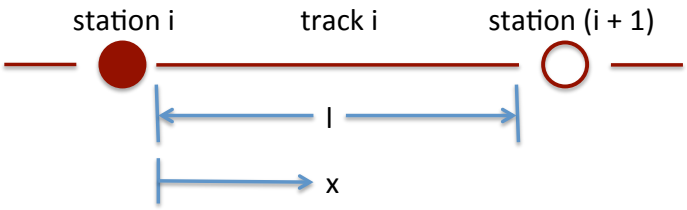
For simplicity we have represented the system in Fig. 4.1(b) as a linear system with periodic boundary condition, as indicated by the dotted line.

Let $P_i(t)$ denote the probability to find the system in station i at time t, and let $p_i(x,t)$ be the probability density to find the system at a position x along track i at time t. In our notation, a given track is designated by the same index as the station on its left; x specifies the distance along a track; and for simplicity we assume each track to be of length l. See Fig. 4.2(a) for an illustration. Because of the periodic nature of the state space, we make the identifications: $i+1 \equiv 1$ if $i = 3$ and $i-1 \equiv 3$ if $i = 1$. The total probability is normalized to unity:

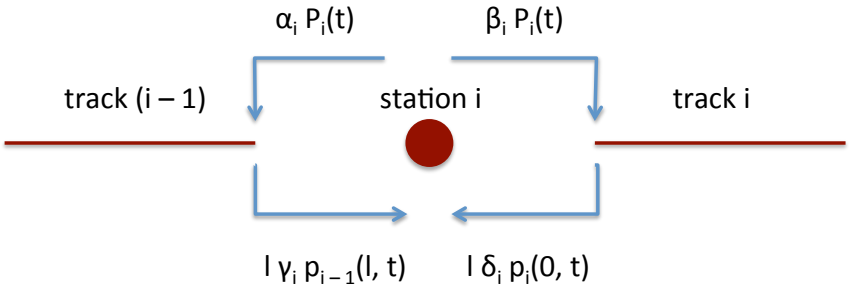
$$\sum_{i=1}^{n_S} P_i(t) \; + \; \sum_{i=1}^{n_T} \int_0^l \mathrm{d}x \, p_i(x,t) = 1. \tag{4.1}$$

Here, $n_S = 3$ is the number of stations and $n_T = 3$ is the number of tracks, but in general these need not be equal (see Sec. 4.4).

We now specify the dynamics of our model. When the system is in station i, it has a probability per unit time α_i to make a leftward transition to the location $x = l$ on track $i-1$, and similarly a probability rate β_i to make a rightward transition to the location $x = 0$ on track i; see Fig. 4.2(b). When the system is on one of the tracks, it performs diffusive motion with a fixed diffusion constant D, with reflective (hard-wall) boundary conditions at $x = 0$ and $x = l$. Upon reaching either end of the track the system might jump into the adjacent station. These jumps are characterized by probability rates γ_i (for transitions from track $i-1$ to station i) and δ_i (from track i to station i). More precisely, the probability per unit time for the system to make a transition from track $i-1$ to station i is given by the product



(a)



(b)

Figure 4.2: **Elements of hybrid model for 2-catenane.** (a) Probability at station i is denoted by $P_i(t)$, and probability density on track i by $p_i(x,t)$, where distance x is measured from station i. Each track is assumed to be of length l. (b) Rate parameters of the model: α_i and β_i denote the transition rates from station i to tracks $(i-1)$ and i, respectively; $l\gamma_i$ and $l\delta_i$ are associated with the reverse transitions, as described in the text.

$l\,\gamma_i\,p_{i-1}(l,t)$, where the factor l (length of each track) is introduced on dimensional grounds, so that the parameter γ_i has units of (time)$^{-1}$. A similar expression holds for transitions from track i to station i; see Fig. 4.2(b).

These transitions give rise to a flow of probability between stations and adjacent tracks. The net current from track $i-1$ to station i is given by

$$J_{\to i}(t) = l\gamma_i p_{i-1}(l,t) - \alpha_i P_i(t), \tag{4.2}$$

and that from station i to track i by

$$J_{i\to}(t) = \beta_i P_i(t) - l\delta_i p_i(0,t). \tag{4.3}$$

(The subscripts "$\to i$" and "$i \to$" indicate rightward probability current into and out of station i, respectively.) The diffusive current at position x along track i is

$$J_i^d(x,t) = -D\frac{\partial}{\partial x}p_i(x,t), \tag{4.4}$$

and the reflective boundary conditions imply that

$$J_i^d(0,t) = J_i^d(l,t) = 0. \tag{4.5}$$

These currents generally lead to changes in the probability distribution. The rate of change of the probability to find the system at station i is the difference between the incoming and the outgoing currents,

$$\frac{\mathrm{d}P_i(t)}{\mathrm{d}t} = J_{\to i}(t) - J_{i\to}(t), \tag{4.6}$$

and that of probability density along track i obeys a diffusion equation with a source and sink:

$$\frac{\partial p_i(x,t)}{\partial t} = -\frac{\partial}{\partial x}J_i^d(x,t) + \delta(x-0)J_{i\to}(t) - \delta(x-l)J_{\to i+1}(t). \tag{4.7}$$

The nature of the source and sink terms becomes clearer if we use Eqs. 4.2, 4.3 and 4.4 to express the currents and reorganize the terms on the right of Eq. 4.7 to get

$$\begin{aligned}\frac{\partial p_i(x,t)}{\partial t} &= D\frac{\partial^2 p_i}{\partial x^2} \\ &\quad + [\delta(x-0)\beta_i P_i(t) + \delta(x-l)\alpha_{i+1}P_{i+1}] \\ &\quad -l\,[\delta(x-0)\delta_i + \delta(x-l)\gamma_i]\,p_i(x,t).\end{aligned} \tag{4.8}$$

The second term on the right is a source term, and the third is a sink term. Eqs. 4.6 and 4.7 form a set of six coupled, linear equations (taking $i = 1, 2, 3$) which collectively constitute the master equation describing the stochastic evolution of the system.

4.2 Constraints Imposed by Detailed Balance

Since our model is meant to represent a system immersed in a thermal reservoir, the dynamics described by our master equation should have the property that when the rate parameters α_i, β_i, γ_i and δ_i are held fixed, the system relaxes to a state of equilibrium in which all currents are zero. This condition of detailed balance imposes constraints on the rate parameters, which we now explore.

Let P_i^{eq} and $p_i^{eq}(x)$ denote, respectively, the station probabilities and track probability densities in the equilibrium state. According to the condition of detailed balance, when these values are substituted into the right sides of Eqs. 4.2, 4.3 and 4.4, the currents appearing on the left sides of those equations vanish. This leads to

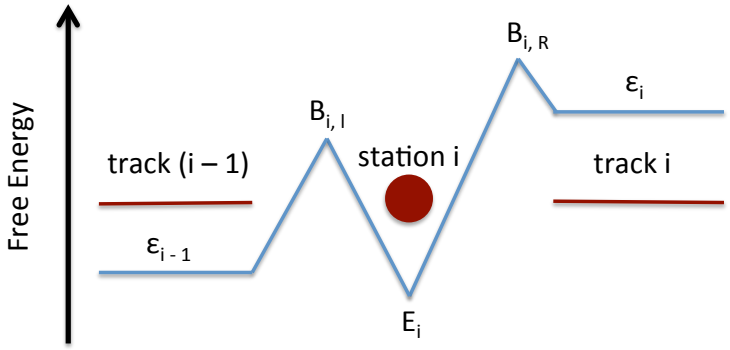


Figure 4.3: **Free energies and barriers.** Station i has energy E_i, and track i energy ϵ_i. Track $(i-1)$ and station i are separated by barrier $B_{i,L}$. Station i and track i by $B_{i,r}$.

the relations,

$$l\gamma_i p_{i-1}^{eq}(l) = \alpha_i P_i^{eq} \equiv \eta \exp[-B_{i,L}] \tag{4.9}$$

$$l\delta_i p_i^{eq}(0) = \beta_i P_i^{eq} \equiv \eta \exp[-B_{i,R}] \tag{4.10}$$

$$\frac{\partial}{\partial x} p_i^{eq}(x) = 0. \tag{4.11}$$

Eqs. 4.9 and 4.10, together with an arbitrary frequency scale η, define the dimensionless parameters $B_{i,L}$ and $B_{i,R}$, while Eq. 4.11 implies that the equilibrium probability density is uniform along each track. Introducing the dimensionless parameters

$$E_i \equiv -\ln P_i^{eq} \quad \text{and} \quad \epsilon_i \equiv -\ln\left(l p_i^{eq}\right) \tag{4.12}$$

now allows us to rewrite Eqs. 4.9 and 4.10 as follows:

$$\begin{aligned} \alpha_i &= \eta\, e^{-B_{i,L}+E_i} \quad , \quad \beta_i = \eta\, e^{-B_{i,R}+E_i}, \\ \gamma_i &= \eta\, e^{-B_{i,L}+\epsilon_{i-1}} \quad , \quad \delta_i = \eta\, e^{-B_{i,R}+\epsilon_i}. \end{aligned} \tag{4.13}$$

As before, because of the resemblance of these expressions to the Arrhenius form for chemical reaction rates, it is natural to interpret E_i (or ϵ_i) as the effective free

energy of the small ring when it is at station i (or on track i); and $B_{i,L}$ (or $B_{i,R}$) as the height of the effective free energy barrier that separates station i from the track immediately to its left (or right). These energies are given in units of kT.

4.3 Statement and Proof of NPT

Stochastic pumping corresponds to the periodic variation of the transition rates $\{\alpha_i, \beta_i, \gamma_i, \delta_i\}$, subject to the constraints imposed by detailed balance. In the energetic picture introduced above, this translates to the periodic variation of the energies and barriers $\{E_i, \epsilon_i, B_{i,L}, B_{i,R}\}$, that is, $E_i(t+\tau) = E_i(t)$, etc., where τ is the period of the pumping. Under these conditions the system relaxes to a unique periodic steady state,

$$P_i^{ps}(t+\tau) = P_i^{ps}(t) \quad , \quad p_i^{ps}(x, t+\tau) = p_i^{ps}(x,t), \tag{4.14}$$

characterized by time-periodic currents passing through the stations, $J_{\to i}^{ps}(t)$ and $J_{i\to}^{ps}(t)$, and along the tracks, $J_i^{d,ps}(x,t)$. We are interested in the *integrated* currents,

$$\Phi_{\to i}^{ps} \equiv \int_\tau \mathrm{d}t\, J_{\to i}^{ps}(t) \quad , \quad \Phi_{i\to}^{ps} \equiv \int_\tau \mathrm{d}t\, J_{i\to}^{ps}(t) \quad , \quad \Phi_i^{d,ps}(x) \equiv \int_\tau \mathrm{d}t\, J_i^{d,ps}(x,t), \tag{4.15}$$

where $\int_\tau$ denotes an integral over one period of pumping. Here, $\Phi_{\to i}^{ps}$ represents the net flow of probability from track $i-1$ into station i over one pumping cycle, and $\Phi_{i\to}^{ps}$ and $\Phi_i^{d,ps}(x)$ have similar interpretations; the integrated currents thus measure the extent to which the pumping of the energies and barriers drives a non-zero current around the cycle depicted in Fig. 4.1(a). Physically, these currents measure our ability to generate directed mechanical motion of the small macrocycle around the

large macrocycle, by the periodic variation of external parameters, with positive and negative Φ's corresponding to clockwise and counterclockwise motion, respectively; see Fig. 3.1.

These considerations apply to the time-periodic pumping of any combination of the parameters $\{E_i, \epsilon_i, B_{i,L}, B_{i,R}\}$. In the subsequent analysis, however, we will assume that the track energies ϵ_i remain constant with time, while the station and/or barrier energies (the E's and B's) are varied periodically. Thus, we treat the tracks as fixed conduits for diffusive motion from one station to another; this is in keeping with the relevant experimental studies [13, 18], which did not include any time-dependent track energies. The no-pumping theorem (NPT) that we now prove states that, in order to generate non-zero integrated currents, we must vary some combination that includes *both* station energies (the E_i's) *and* barrier energies (the $B_{i,L}$'s and/or $B_{i,R}$'s). In other words: (1) if we vary only the barrier energies, while keeping the station energies fixed, or (2) if we vary only the station energies, while keeping the barrier energies fixed, then in either case all the integrated currents will be zero. To prove the NPT we now consider these cases separately.

The first case is easy to analyze. Let the term *instantaneous equilibrium distribution* denote the equilibrium distribution corresponding to the instantaneous values of the parameters. By Eq. 4.12, this distribution depends only on the state energies, and not on the barrier energies. Thus when the E_i's and ϵ_i's are held fixed, the instantaneous equilibrium distribution $\{P_i^{eq}, p_i^{eq}(x)\}$ is invariant with time and is a stationary solution of the dynamics. Since the periodic steady state is unique for any pumping protocol, it follows that when the state energies are held fixed the

system relaxes to the fixed instantaneous equilibrium distribution, no matter how the barrier energies are varied. In this state, all the instantaneous currents are zero, and therefore the integrated currents also vanish. Note that similar observations were made in the contexts of closed and open discrete pumps in Secs. 3.1 and 3.2.1, respectively.

Now consider the situation in which the station energies are varied periodically in time, $E_i(t+\tau) = E_i(t)$, and the barriers $\{B_{i,L}, B_{i,R}\}$ (together with the track energies ϵ_i) are kept fixed. The NPT then follows from a combination of two conditions: the *detailed balance* constraints, Eq. 4.13, and the *periodicity* of the probability distribution, Eq. 4.14, as we now show.

Let us first explore the consequences of the detailed balance constraints. Combining Eq. 4.13 with the expressions for the instantaneous currents, Eqs. 4.2 and 4.3, we derive for each station i,

$$e^{B_{i,L}} J_{\to i}(t) + e^{B_{i,R}} J_{i\to}(t) = \eta\, l \Big[e^{\epsilon_{i-1}} p_{i-1}(l,t) - e^{\epsilon_i} p_i(0,t) \Big]. \tag{4.16}$$

Note that the superscript *ps* does not appear here, as this relation is valid whether or not the system has reached a periodic steady state. Summing both sides over i, we get

$$\sum_i \Big[e^{B_{i,L}} J_{\to i}(t) + e^{B_{i,R}} J_{i\to}(t) \Big] = \eta\, l \sum_i e^{\epsilon_i} \Big[p_i(l,t) - p_i(0,t) \Big]. \tag{4.17}$$

The definition of the diffusive current, Eq. 4.4, implies

$$-\frac{1}{D} \int_0^l \mathrm{d}x\, J_i^d(x,t) = \Big[p_i(l,t) - p_i(0,t) \Big], \tag{4.18}$$

which combines with Eq. 4.17 to give

$$\sum_i \left[e^{B_{i,L}} J_{\to i}(t) + e^{B_{i,R}} J_{i\to}(t) + \frac{\eta\, l}{D} e^{\epsilon_i} \int_0^l \mathrm{d}x\, J_i^d(x,t) \right] = 0. \tag{4.19}$$

(This is the analogue of Eqs. 3.10 and 3.24 for closed and open discrete models, respectively.) If we now assume the system has reached a periodic steady state, and we integrate this relation over one period, we get

$$\sum_i \left[e^{B_{i,L}} \Phi^{ps}_{\to i} + e^{B_{i,R}} \Phi^{ps}_{i\to} + \frac{\eta\, l}{D} e^{\epsilon_i} \int_0^l \mathrm{d}x\, \Phi_i^{d,ps}(x) \right] = 0. \tag{4.20}$$

Now we explore the implications of the periodicity of the probability distribution, Eq. 4.14. Since the probability to find the system in station i returns to the same value after each period, the integrated current that enters that station from the left must be balanced by the integrated current that exits from the right: $\Phi^{ps}_{\to i} = \Phi^{ps}_{i\to}$. This value is in turn equal to the integrated current entering track i from the left. Along track i the integrated current $\Phi_i^{d,ps}(x)$ must be the same for any two points x_1 and x_2, otherwise there would be a net accumulation or depletion of probability in the interval between those points, over each period. Proceeding in this manner around the entire circuit we conclude that the integrated current is uniform all along:[2]

$$\Phi^{ps}_{\to i} = \Phi^{ps}_{i\to} = \Phi_i^{d,ps}(x) = \Phi^{ps}_{\to i+1} \cdots \equiv \Phi. \tag{4.21}$$

[2]A derivation of Eq. 4.21 directly from the master equations requires a more careful treatment, in which the source and sink terms in Eq. 4.7 are displaced slightly from the track-ends. As this does not contribute conceptually to the main line of the proof we present the analysis in Appendix A.

It follows immediately from Eqs. 4.20 and 4.21 that all integrated currents are zero. Thus the NPT is established for the three-state model depicted in Fig. 4.1.

4.4 Generalizations

We now generalize our discussion along two different directions. First, following Ref. [18], we allow for a spatially non-uniform (but still time-independent) free energy landscape along each of the tracks. Secondly, we move beyond the simple three-station, three-track network shown in Fig. 4.1, and extend our model to encompass an arbitrary, finite network of stations and tracks. The analysis involved in these more general situations is similar to that presented in Sec. 4.3; therefore, to avoid repetition, we sketch only the key ideas in the following discussion.

First we allow a nonuniform energy landscape $V_i(x)$ along each track i, instead of constant ϵ_i, again in units of kT; this leads to the expressions,

$$J_i^d(x,t) = -D\left[\frac{\partial p_i(x,t)}{\partial x} + \frac{\partial V_i(x)}{\partial x} p_i(x,t)\right] \tag{4.22}$$

$$\gamma_i = \eta\, e^{-B_{i,L}+V_{i-1}(l)} \quad , \quad \delta_i = \eta\, e^{-B_{i,R}+V_i(0)}. \tag{4.23}$$

(compare Eqs. 4.4 and 4.13, respectively). When all the state energies are held fixed and only the barrier energies are varied with time, the arguments presented earlier apply here without modification, and we can conclude that all currents vanish in the steady state. When instead the barrier energies are fixed and the station energies are varied periodically, in place of Eq. (4.16) we have

$$e^{B_{i,L}} J^{ps}_{\to i}(t) + e^{B_{i,R}} J^{ps}_{i\to}(t) = \eta\, l\left[e^{V_{i-1}(l)} p^{ps}_{i-1}(l,t) - e^{V_i(0)} p^{ps}_i(0,t)\right], \tag{4.24}$$

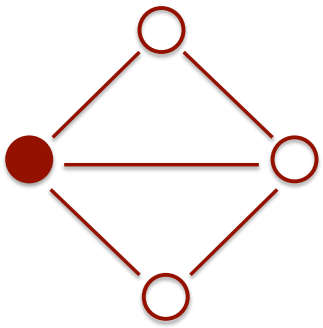


Figure 4.4: **A generic hybrid model.** There are $n_S = 4$ stations and $n_T = 5$ tracks.

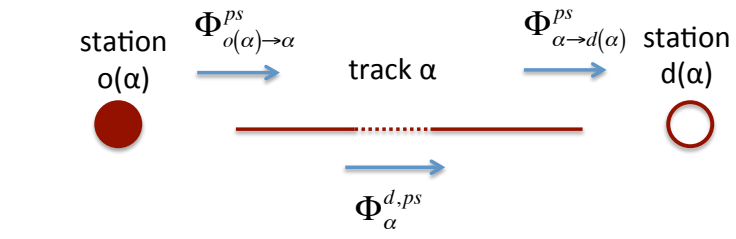
which leads to a modified version of Eq. (4.20),

$$\sum_i \left[e^{B_{i,L}} \Phi^{ps}_{\to i} + e^{B_{i,R}} \Phi^{ps}_{i\to} + \frac{\eta\, l}{D} \int_0^l \mathrm{d}x\, e^{V_i(x)} \Phi_i^{d,ps}(x) \right] = 0. \tag{4.25}$$

The periodicity of the probability distribution, Eq. 4.14, again implies a uniform integrated current, Eq. 4.21. The combination of Eqs. 4.21 and 4.25 in turn immediately implies that all integrated currents vanish, and thus the NPT is established.

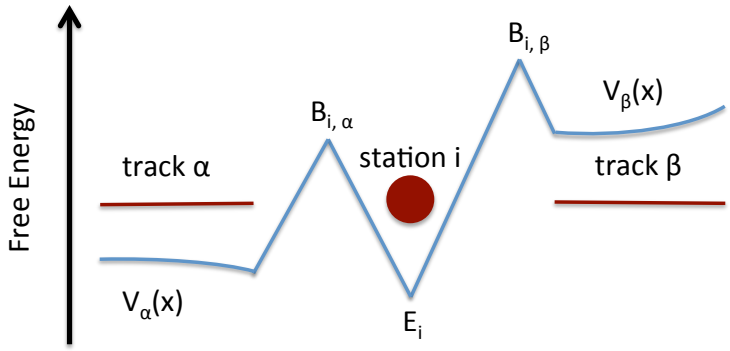
Now consider a more general network, composed of n_S stations and n_T tracks, which need not be equal. As before, we assume this network to be *connected*: for any pair of stations i, j, there exists at least one path – a sequence of alternating tracks and stations – that connects station i to station j.[3] Fig. 4.4 illustrates a connected network, with $n_S = 4$ and $n_T = 5$. The [2]catenane model analyzed in Sections 4.1 - 4.3 had only one *cycle* – a closed loop of stations and tracks (see Fig. 4.1) – but in the more general case considered here there can be more than one cycle, as shown in Fig. 4.4.

[3]This assumption is not restrictive, as any finite network can be decomposed into two or more connected networks, each of which satisfies the NPT.



(a)

(b)



(c)

Figure 4.5: **Notations for a generic hybrid network.** (a) origin $o(\alpha)$ and destination ($d(\alpha)$) of a track α, and associated currents. (b) Current between a station and the adjacent tracks. Although we show a station connected to two tracks, in general a station can be adjacent to as few as one and as many as n_T tracks. (c) Energetics for a hybrid network with non-uniform free energy along the tracks; compare Fig. 4.3.

We will now use Roman indices $i \in \{1, \ldots n_S\}$ to denote stations and Greek indices $\alpha \in \{1, \ldots n_T\}$ to denote tracks. To each track α we assign an arbitrary direction, such that the station at one end of the track is viewed as the *origin* $o(\alpha)$ and the station at the other end is the *destination* $d(\alpha)$, as illustrated in Fig. 4.5(a). The integrated current from track α (or station i) to station i (or track α) is given by $\Phi^{ps}_{\alpha \to i}$ (or $\Phi^{ps}_{i \to \alpha}$); and the integrated diffusive current on track α along the preassigned direction by $\Phi^{d,ps}_{\alpha}(x)$. These conventions are illustrated in Figs. 4.5(a) and (b). These assignments are made independently for the various tracks, and a given station can simultaneously serve as the origin of some tracks and the destination of others. Station and track energies are denoted by E_i and $V_\alpha(x)$, respectively; the barrier between station i and track α is given by $B_{i,\alpha}$. See Fig. 4.5(c).

As earlier, under the time-periodic variation of station and/or barrier energies, the system settles into a unique periodic steady state. Moreover, when all the state energies are held fixed and only the barrier energies are varied with time, then the entire system simply relaxes to a state of thermal equilibrium, in which all currents vanish. To establish the NPT, it remains to show that the integrated current vanishes when the barrier energies are kept fixed and the station energies are varied periodically.

The periodicity of the probability distribution implies, for each station i,

$$\sum_{\{\alpha | d(\alpha) = i\}} \Phi^{d,ps}_{\alpha \to i} = \sum_{\{\alpha | o(\alpha) = i\}} \Phi^{d,ps}_{i \to \alpha}; \tag{4.26}$$

the sum on the left represents the net integrated current into station i from all tracks

α for which it is the destination, and the sum on the right is the net integrated current out of the station into all tracks to which it is the origin. For each track α we have

$$\Phi^{ps}_{o(\alpha)\to\alpha} = \Phi^{d,ps}_{\alpha}(x) = \Phi^{ps}_{\alpha\to d(\alpha)} \tag{4.27}$$

(compare with Eq. 4.21), which ensures that probability neither accumulates nor depletes anywhere on the track, with each cycle. Eqs. 4.26 and 4.27 have an interesting consequence: if there exists a non-zero integrated current in the system, then it must be part of a cycle (of alternating stations and currents) along which all of the integrated currents point in the same direction. The intuition is straightforward: to prevent the systematic accumulation of probability within the network, current must flow in a circle. We now formalize and establish this statement, and then use it to prove that all integrated currents must be zero (when the barriers are fixed and the station energies are varied periodically), following arguments similar to those presented in Sec. 3.1.2.

Without loss of generality, suppose that a particular track α supports a positive integrated current: $\Phi^{d,ps}_{\alpha} > 0$. (Equivalent arguments would apply if the sign were negative.) Eq. 4.27 then implies positive integrated currents $\Phi^{ps}_{o(\alpha)\to\alpha}$ and $\Phi^{ps}_{\alpha\to d(\alpha)}$. As probability cannot deplete on station $o(\alpha)$ over a complete cycle, $o(\alpha)$ must have neighboring track(s) β such that $\Phi^{ps}_{\beta\to o(\alpha)} > 0$. Similarly, to avoid accumulation of probability, $d(\alpha)$ must have neighboring track(s) γ such that $\Phi_{d(\alpha)\to\gamma} > 0$. The periodic conservation of probability in turn establishes the directionality of the integrated currents along these tracks: on track β, the integrated current must flow

toward station $o(\alpha)$, and on track γ it must flow away from station $d(\alpha)$. Continuing in this manner, we now construct a set $\mathcal{D}(\alpha)$ of stations and tracks to which there is a positive flow of current from α; this will consist of $d(\alpha), \gamma$, and so on. Similarly, we construct set $\mathcal{S}(\alpha)$ of stations and tracks from which there is a positive flow to α; this will consist of $o(\alpha), \beta$, so on. In order to prevent the accumulation of probability in the former set and its depletion in the latter with each complete cycle, $\mathcal{D}(\alpha)$ and $\mathcal{S}(\alpha)$ must have a common element. This implies the existence of a cycle

$$c \equiv \alpha \to d(\alpha) \to \gamma \to \ldots \to \beta \to o(\alpha) \to \alpha, \tag{4.28}$$

along which all integrated currents flow in the same direction.

Now recall that each track in our network has been assigned a direction, pointing from its origin to its destination. By assumption, for track α this direction is parallel to the direction of probability flow around the cycle c, indicated by Eq. 4.28. However, since the assignment of track directions is arbitrary, each of the remaining tracks in the cycle $(\beta, \cdots \gamma)$ might be directed either parallel or anti-parallel to the flow along the cycle. Let us therefore introduce a factor $s_\mu = \pm 1$, defined for every track μ in the cycle c, such that $s_\mu = +1$ (or -1) if track μ is oriented parallel (or anti-parallel) to the flow in the cycle. Then $\Psi^{d,ps}_\mu \equiv s_\mu \Phi^{d,ps}_\mu > 0$ for each track μ in the cycle c, and $\Psi^{ps}_{o(\mu)\to\mu} \equiv s_\mu \Phi^{ps}_{o(\mu)\to\mu}$ and $\Psi^{ps}_{\mu\to d(\mu)} \equiv s_\mu \Phi^{ps}_{\mu\to d(\mu)}$ are positive as well, by Eq. 4.27. The analogue of Eqs. 4.20 and 4.25 for the cycle c is given by

$$\sum_\mu \left[e^{B_{o(\mu),\mu}} \Psi^{ps}_{o(\mu)\to\mu} + e^{B_{d(\mu),\mu}} \Psi^{ps}_{\mu\to d(\mu)} + \frac{\eta\, l}{D} \int_0^l \mathrm{d}x\, e^{V_\mu(x)} \Psi^{d,ps}_\mu(x) \right] = 0, \tag{4.29}$$

where the sum is taken over all tracks μ in cycle c. Since all the Ψ's are positive, Eq. 4.29 cannot be satisfied. Hence our starting assumption, the existence of a

positive integrated current along track α, must be invalid. This establishes the no-pumping theorem.

Chapter 5

Szilard's engine

---

[1]In this chapter and the following we focus on the topic of Maxwell's demon. As discussed in the Introduction, this is a paradoxical thought experiment proposed by James Clark Maxwell where an intelligent being " whose faculties are so sharpened that he can follow every molecule in its course" can generate a heat flow against a thermal gradient without expenditure of work [53], in violation of the second law of thermodynamics (specifically its Clausius statement [75]). This puzzle has lead to nearly 150 years of debates and discussions [52, 53, 55–58, 60, 61, 73, 83]. A consensus has developed based on the works of Rolf Landauer [58], Oliver Penrose [59] and Charles H. Bennett [60], who concluded that the demon gathers information about the molecular motion during its operations and discarding this information has a minimum entropic cost that makes up for the violation.

Generically, the term "Maxwell's demon" refers not only to the original setting of Maxwell, but to any situation where the operations of an intelligent being leads to a violation of the second law. An interesting example in this context was provided by Leo Szilard in 1929 [56]. He considered a single-molecule gas in an isothermal chamber and the following sequence of actions:

[1]This chapter is based on Ref. [92].

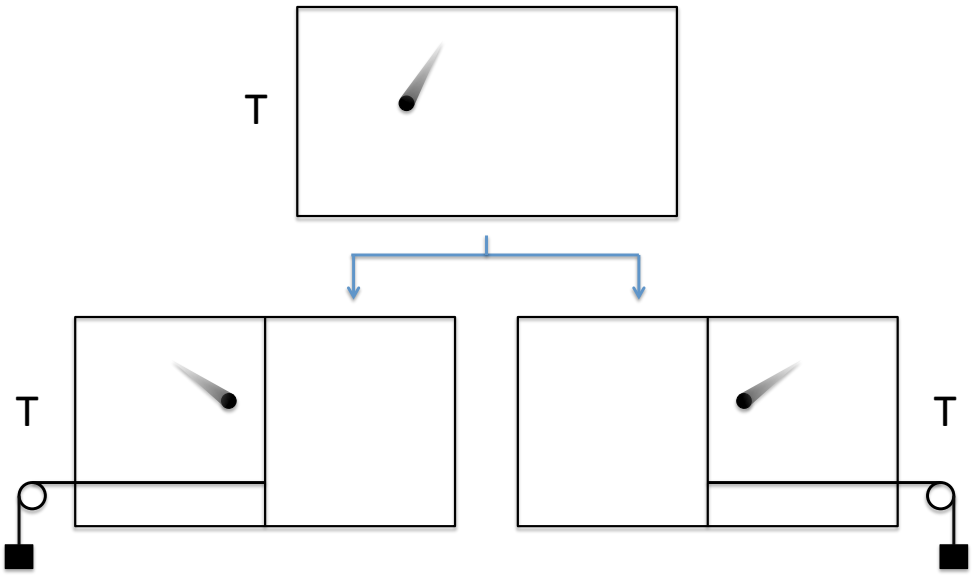


Figure 5.1: **Szilard's engine.** One starts with a one-molecule gas inside an isothermal chamber. A frictionless partition is quickly inserted in the middle of the chamber. If the molecule is trapped in the left, a mass is attached from the left so that the mass is raised because of the pressure of the gas. If the molecule is trapped in the right, the mass is connected from the right. In both case a positive amount of work is extracted. One can remove the partition, let the particle thermalize, and repeat the whole process several times to gain as much work as desired.

(i) A frictionless partition is inserted quickly in the middle of the chamber,

(ii) If the molecule is trapped in the left half of the chamber, a small mass is attached to the partition from the left (as shown in Fig. 5.1) such that the mass is raised as the partition moves to the right because of the pressure of the gas. Alternatively, if the molecule is trapped in the right half of the chamber, the mass is attached from the right. Again, the mass is raised because the partition now moves to the left because of the pressure of the gas.

(iii) The partition is removed.

A positive amount of work is done in each sequence of actions. One can repeat the sequence to perform as much work as desired. For a repetitive process the molecule undergoes a periodic evolution and therefore has the same energy on the average. The required energy for work must come from the surrounding reservoir. This is puzzling because the sole result of the process seems to be the extraction of heat from a single reservoir and its conversion into work, which is explicitly prohibited by the Kelvin-Planck statement of the second law [75]. The puzzle is resolved by noting the following set of facts: (1) taking the binary decision about connecting the mass, whether from the left or the right, requires the knowledge about the molecule's whereabouts, (2) the molecule's position must be measured and recorded prior to this decision, and (3) erasure of the recorded information comes at a thermodynamic cost, which is $kT \ln 2$ per bit of information (*Landauer's principle*) [58].

In this chapter we propose an exactly solvable autonomous model that can achieve the same feat as the Szilard's engine. Before we go into the details of the

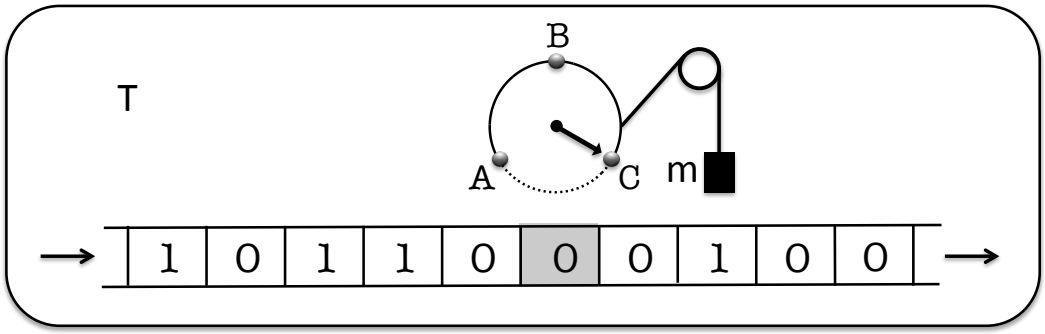
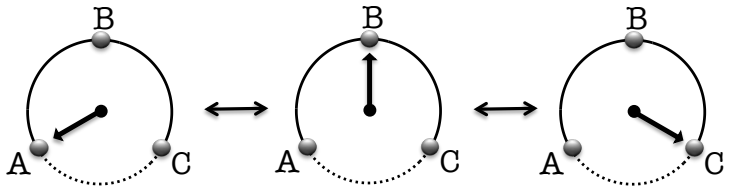


Figure 5.2: **Our model akin to Szilard's engine.** The three-state system (our demon) interacts with three other components: a mass that can be lifted or lowered, a stream of bits (see footnote 2) that pass by the demon in sequence, and a thermal reservoir, the surrounding.

model, let us have a brief overview. There are four principal components (Fig. 5.2): (a) a three-state system which we will term as the demon, (b) a mass attached to the demon, (c) a collection of two-state systems (the *bits*[2]), and (d) a heat reservoir in which all the previous components are immersed. The model has three parameters: (a) δ describing the initial statistical distribution of the bits, (b) ϵ characterizing the weight of the mass, and (c) τ giving the duration of interaction of the demon with each bit. For any set of values (δ, ϵ, τ) the model reaches a unique periodic steady state distribution, characterized by an average rate of work performed on the mass (by raising or lowering it), and an average rate of information written to the bit stream (characterized by a change in their statistical distribution).

In the next section we specify the details of our model. In Sec. 5.2 we present

[2]Note the rather unconventional use of the term "bit". Here it refers to two-state systems each of which can encode one bit (in the sense of unit) of information. The meaning of the term should be clear from the context.



(a)

(b)

Figure 5.3: **Models of our demon and each bit.** (a) The demon is shown as an arrow that can point to three directions: A, B and C. In absence of a bit, transitions are allowed only along the continuous lines ($A \leftrightarrow C$ are not allowed). (b) A bit is shown as an arrow that can point either up (state 1) or down (state 0). No transitions are possible in absence of the demon.

an outline of the analysis. In Sec. 5.3 we present the qualitative behavior of our model through its phase diagram. We see that the model is versatile: in addition to behaving like a Szilard engine, it can also act as a Landauer eraser. In Sec. 5.4 we discuss the modified expression of the second law.

5.1 Model

We now specify the details of our model. We represent the three states of the demon as an arrow pointing in one of three directions (A, B, or C) on the face of a dial; see Fig. 5.3(a). We model the transitions among the three states as Poisson

processes: if the demon is in a state $i \in \{A, B, C\}$, its rate of transition to some other state j is given by a real number $R_{ji} \geq 0$. We assume the transitions $A \leftrightarrow C$ are not allowed in absence of bits, hence

$$R_{CA} = R_{AC} = 0. \tag{5.1}$$

To keep our analysis simple, we assume all other rates to be unity

$$R_{AB} = R_{BA} = R_{BC} = R_{CB} = 1. \tag{5.2}$$

In the following we refer to these transitions as intrinsic dynamics of the demon as they do not involve the bits or the mass. Eqs. 5.1 and 5.2 imply that the effective free energies of the three states A, B and C are equal.

Each bit has two states, 0 and 1, with equal effective free energies. One can think of an arrow which can point either up (state 1) or down (state 0). We assume that no transition is possible between these states in absence of a demon. In other words, the bits do not have any intrinsic dynamics; left to themselves, they will maintain their present state for an indefinite time. See Fig. 5.3(b) for an illustration of the model.

We now specify the interaction between the demon and a bit: the demon can make the transition from C to A if the bit simultaneously flips from 0 to 1; also, the demon can make the reverse transition from A to C if the bit simultaneously flips from 1 to 0. We have shown this in Fig. 5.4.

Fig. 5.4 also shows a mass m which is now attached to the demon such that the mass is raised by a height Δh if the demon goes from C to A, and is lowered by the same amount if the demon goes from A to C. The presence of the mass favors the

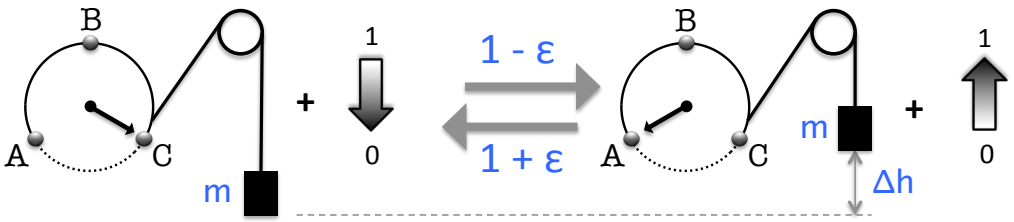


Figure 5.4: **Detailed balance.** Whenever the demon makes a transition $C \to A$, the interacting bit (closest to the demon) flips from 0 to 1, and the mass gets raised. For the opposite transition, $C \leftarrow A$, the bit flips from 1 to 0, and the mass gets lowered. The mass favors the leftward transitions over the rightward transitions. If $(1-\epsilon)$ and $(1+\epsilon)$ denote the corresponding rates, respectively, this is reflected in the ratio $(1-\epsilon)/(1+\epsilon) = \exp(-mg\Delta h/kT)$.

C to A transitions over the A to C transitions. This is reflected in the ratio of the corresponding rates: if $(1-\epsilon)$ is the rate of transition from C to A (simultaneously with $0 \leftarrow 1$) and $(1+\epsilon)$ is the rate of transition from A to C (simultaneously with $1 \to 0$), from local detailed balance (Eq. 2.36), we have

$$\frac{1-\epsilon}{1+\epsilon} = e^{-mg\Delta h/kT}, \quad \epsilon = \tanh\frac{mg\Delta h}{2kT}, \tag{5.3}$$

where g is acceleration due to gravity. Rates associated with the intrinsic transitions of the demon, $A \leftrightarrow B$ and $B \leftrightarrow C$, remain unaffected.

We can summarize the joint dynamics of the demon and an interacting bit by the following master equation

$$\dot{p}_i = \sum_{j\neq i} \left(R_{ij}p_j - R_{ji}p_i\right), \quad i,j \in \{A0, \ldots C1\}. \tag{5.4}$$

Here $A0$ indicates the joint state of the demon in state A and the bit in state 0, similarly for the other combinations. All the rates have been discussed already: the

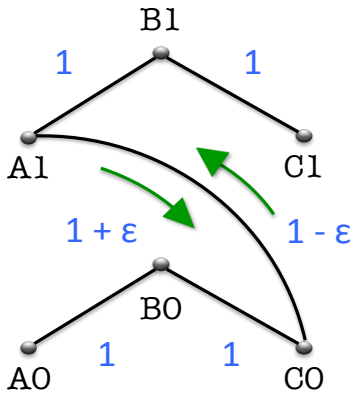


Figure 5.5: **Graph of the joint master equation.** Solid lines indicate allowed transitions. The numbers in blue indicate the associated rates. The edge that connects $A1$ and $C0$ represents the coupling between the demon and a bit. The mass goes up (down) whenever here is a joint transition $C0 \to A1$ ($A1 \to C0$).

concerted rates are $(1 \pm \epsilon)$, as shown in Fig. 5.4, and the rates for the intrinsic dynamics of the demon are unity (Eq. 5.2). We have given the graphical representation of the joint dynamics in Fig. 5.5. The six joint states $\{A0, \ldots C1\}$ are represented by six vertices; the allowed transitions are shown by the edges. The transitions $A0 \leftrightarrow B0$, $B0 \leftrightarrow C0$, $A1 \leftrightarrow B1$ and $B1 \leftrightarrow C1$ are due to the intrinsic dynamics of the demon; transitions $C0 \leftrightarrow A1$ correspond to the interaction of the demon with a bit; the mass gets raised (lowered) in transitions $C0 \to A1$ ($A1 \to C0$).

As mentioned earlier, our setup involves a *stream* of bits, arranged at equally spaced intervals along a tape that is pulled at a constant speed, for instance by a frictionless flywheel. The demon remains at a fixed location, and interacts, in a manner described above, only with the bit that is currently nearest to it. Let τ^{-1} denote the rate at which the bits pass by the demon, each interacting with the

demon for a time interval of duration τ, an *interaction interval*, before the next bit in the stream takes its place. Thus τ determines the extent to which the joint system of the demon and the interacting bit approaches equilibrium during one such interaction interval. For $\tau \ll 1$ the system hardly evolves during the interval, whereas for $\tau \gg 1$ the demon and the bit effectively reach equilibrium. The latter is just the steady state solution of Eq. 5.10:

$$p_{A0}^{eq} = p_{B0}^{eq} = p_{C0}^{eq} = \frac{1+\epsilon}{2}, \quad p_{A1}^{eq} = p_{B1}^{eq} = p_{C1}^{eq} = \frac{1-\epsilon}{2}. \tag{5.5}$$

When a new bit comes at the end of an interaction interval, the change may show up as a transition in the joint state of the demon and the interacting bit, *not* along any of the solid lines in Fig. 5.5. For the purpose of illustration, consider the situation in Fig. 5.6. In subfigure (a) we have shown a possible situation at the moment t_n^- i.e. just before the end of the n^{th} interaction interval. The demon is in state B, the outgoing (n^{th}) bit is in state 1, and hence the joint state is $B1$. At the moment t_n^+, the $(n+1)^{th}$ bit is nearest to the demon and happens to be initiated in state 0. The joint state at time t_n^+ must be $B0$. This is because the demon does not have any time to change its state during the instantaneous switching from the old to the new bit. However, the switching appears as a vertical transition from $B1$ to $B0$ in the second figure of Fig. 5.6(b). Clearly, there has been *no physical transition*, and the vertical transition above is an artifact of the reduced description of the demon and the stream of bits in terms of the demon and just the interacting bit. There is no thermodynamic significance of these artificial transitions.

The incoming bits are statistically independent of one another, each with a

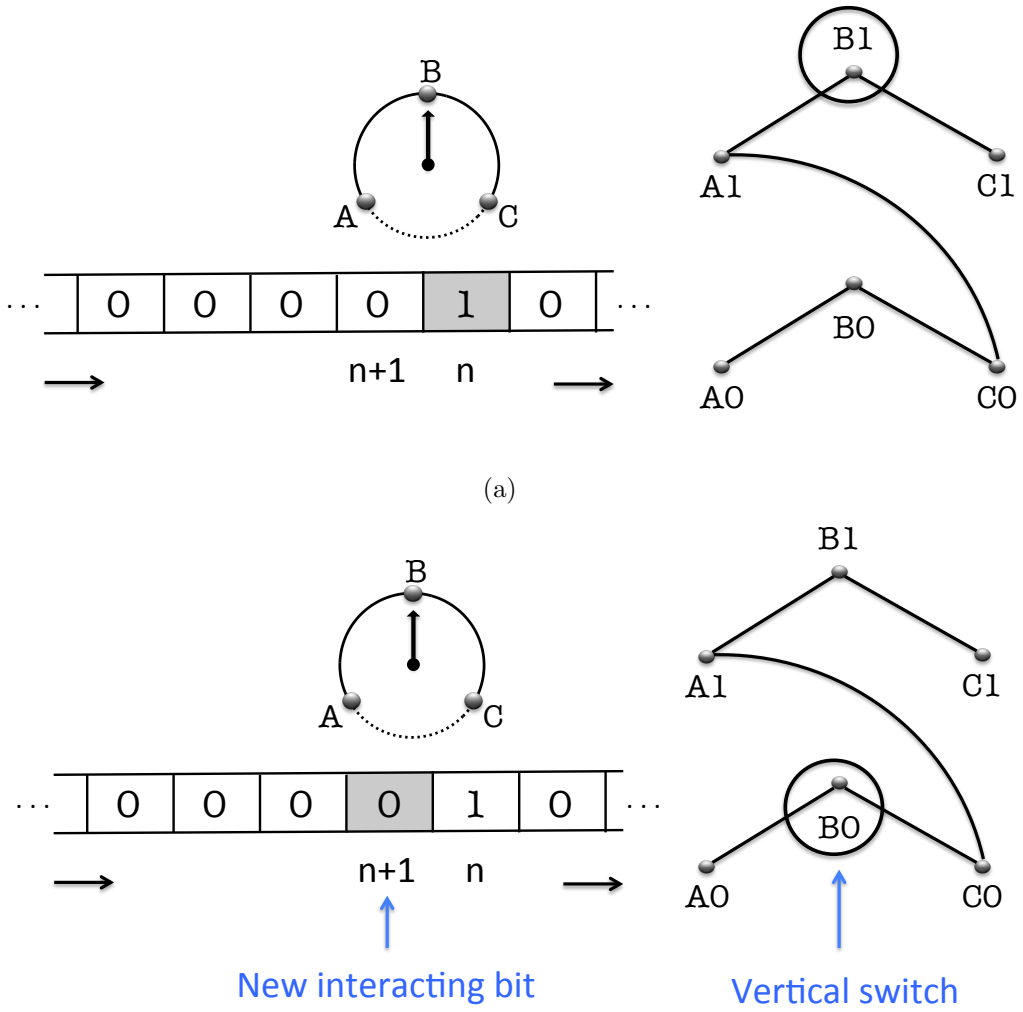


Figure 5.6: **Illustration of "unphysical" transitions.** (a) At the moment t_n^- the n^{th} bit is nearest to the demon and the joint state is $B1$. (b) At the moment t_n^+, the $(n+1)^{th}$ bit is nearest, the demons cannot not change its state during this instantaneous switch, and hence the joint state is $B0$. This is not a physical transition, just an artifact of the mode of description we have chosen.

probability p_0 to be in state 0 and p_1 to be in state 1. The *excess parameter*

$$\delta = p_0 - p_1 \tag{5.6}$$

quantifies the excess of 0's in the incoming bit stream. Let b_n and b'_n denote, respectively, the incoming and outgoing state of the n^{th} bit in the stream. The state of the bit can change only when it is interacting with the demon.

The graph shown in Fig. 5.10 forms a linear chain. Because this chain has no closed loops, the demon cannot perform continuous directed rotation, $\ldots A \to B \to C \to A \ldots$ for clockwise (CW) rotation and $\ldots A \to C \to B \to A \ldots$ for counterclockwise (CCW) rotation, with a single bit. Yet, the "unphysical" transitions discussed above provide a ratchet-like mechanism to make such a motion possible when the demon interacts with a stream of bits. To understand how this works, in the rest of this paragraph, we consider the case where every bit in the incoming stream is set to 0. To keep track of the net clockwise (CW) rotation we introduce an integer variable χ whose value increases by unity whenever the demon makes the transition $C \to A$ and decreases by unity whenever the demon makes the transition $A \to C$. The demon interacts with the n^{th} bit during the n^{th} interaction interval, $t_n = n\,\tau \leq t < t_{n+1}$. At the start of this interval, the composite system begins in state $A0$, $B0$ or $C0$, since $b_n = 0$. From $t = t_n$ to t_{n+1} the system evolves among the network of states depicted in Fig. 5.5. It might repeatedly pass forward and back along the edge connecting $C0$ to $A1$, resulting in alternating increments and decrements of the counter $\chi(t)$. At the end of the interaction interval, if the system is found in state $A0$, $B0$ or $C0$ (i.e. if $b'_n = 0$) then we can infer that every transition

$C0 \to A1$ was balanced by a transition $A1 \to C0$, hence $\Delta\chi_n \equiv \chi(t_{n+1}) - \chi(t_n) = 0$. If the system instead ends in $A1$, $B1$ or $C1$ ($b'_n = 1$), then the counter has advanced by one net unit: $\Delta\chi_n = +1$. At $t = t_{n+1}$, the n^{th} bit is replaced by the $(n+1)^{th}$ bit, and the next interval commences. Thus, if the composite system is in state $B1$ at the end of one interval, then at the start of the next interval it is in state $B0$ (This is the "unphysical transition" we referred to in the beginning of this paragraph. Compare also Fig. 5.6). Over time, the demon interacts with a sequence of bits, all initialized to 0, and the outgoing bit stream contains a record of the demon's rotary motion: each occurrence of an outgoing bit in state 1 indicates one full CW rotation, $\Delta\chi = +1$. *Since the value of the counter can only increase or remain unchanged from one interval to the next*, in the long run $\chi(t)$ grows with time and the demon undergoes directed CW rotation.

If the incoming stream were instead composed entirely of 1's, then full CW rotations would be prohibited; full CCW rotations would be documented as outgoing 0's; and $\chi(t)$ would be decreasing in time.

For a more general distribution of incoming bits, the net change in the counter during the n^{th} interaction interval is

$$\Delta\chi_n = b'_n - b_n, \tag{5.7}$$

and the outgoing stream provides partial information regarding the demon's gyrations. The demon eventually attains a periodic steady state distribution in which its statistical behavior is the same from one interval to the next (discussed in the next section). If the outgoing bit stream is then characterized by values p'_0, p'_1 and

$\delta' \equiv p_0' - p_1'$, then the average number of full CW rotations per interaction interval is

$$\Phi \equiv \langle \Delta\chi_n \rangle = p_1' - p_1 = \frac{1}{2}(\delta - \delta'). \tag{5.8}$$

We will use Φ as our measure of directed rotation, and we will call it the *circulation*.

5.2 Analysis

In this section, we first prove that the demon relaxes to a periodic steady state distribution when the incoming bits have the same initial distribution. Then we give a brief outline of the derivation of Φ (as defined in Eq. 5.8).

Let $\mathbf{p}^D(0) \equiv (p_0^A, p_0^B, p_0^C)^T$ be the initial ($t = 0$) probability distribution of the demon. The joint distribution of the demon and the first bit at $t = 0$ is given by

$$\mathbf{p}^{DB}(0) \equiv \mathcal{M}\,\mathbf{p}^D(0) \quad , \quad \mathcal{M} = \begin{pmatrix} p_0 & 0 & 0 \\ 0 & p_0 & 0 \\ 0 & 0 & p_0 \\ p_1 & 0 & 0 \\ 0 & p_1 & 0 \\ 0 & 0 & p_1 \end{pmatrix} = \begin{pmatrix} p_0\,\mathbb{I} \\ p_1\,\mathbb{I} \end{pmatrix}, \tag{5.9}$$

where $\mathbb{I}$ is the 3×3 identity matrix. $\mathcal{M}$ here encodes the initial distribution of the bit. The joint distribution can be written in this product form (Eq. 5.9) because all the incoming bits are assumed to be statistically uncorrelated to each other and hence to the demon. The joint distribution evolves according to the master equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{p}^{DB}(t) = \mathcal{R}\,\mathbf{p}^{DB}(t) \tag{5.10}$$

where the explicit form of the rate matrix is given by

$$\mathcal{R} = \begin{pmatrix} -1 & 1 & 0 & 0 & 0 & 0 \\ 1 & -2 & 1 & 0 & 0 & 0 \\ 0 & 1 & -2+\epsilon & 1+\epsilon & 0 & 0 \\ 0 & 0 & 1-\epsilon & -2-\epsilon & 1 & 0 \\ 0 & 0 & 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix}. \tag{5.11}$$

Elements of the rate matrix have already been discussed in the last section. The joint state at the end of the interaction interval i.e., at $t = \tau$ is obtained by solving the master equation 5.10 to obtain

$$\mathbf{p}^{DB}(\tau) = e^{\mathcal{R}\tau}\mathbf{p}^{DB}(0). \tag{5.12}$$

The distribution of the demon is obtained by summing over the final state of the bit

$$\mathbf{p}^{D}(\tau) = \mathcal{P}_D\mathbf{p}^{DB}(\tau) \quad , \quad \mathcal{P}_D = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \mathbb{I} & \mathbb{I} \end{pmatrix}. \tag{5.13}$$

We can combine Eqs. 5.9, 5.12 and 5.13 to obtain

$$\mathbf{p}^{D}(\tau) = \mathcal{T}\mathbf{p}^{D}(0) \quad , \quad \mathcal{T} = \mathcal{P}_D e^{\mathcal{R}\tau}\mathcal{M}, \tag{5.14}$$

so the distribution of the demon can be inferred from its initial distribution if we know the matrix $\mathcal{T}(3 \times 3)$ given in the second relation in Eq. 5.14. Clearly, $\mathcal{T}$ can be interpreted as the transition matrix whose component T_{ij} gives the probability for the demon to make a transition to state $i \in \{A, B, C\}$ from state j over the

interaction interval. Because all the incoming bits are in the same distribution and they interact with the demon according to the same rule (Eq. 5.10), the matrix $\mathcal{T}$ is the transition matrix of the demon for *any* interaction interval. Therefore, the distribution of the demon at the end on n interaction intervals (for any non-negative integer n) i.e., at $t = n\,\tau$ is simply given by

$$\mathbf{p}^D(n\,\tau) = \mathcal{T}^n\,\mathbf{p}^D(0). \tag{5.15}$$

From its very interpretation $\mathcal{T}$ is a positive matrix: The demon can make a transition to any state from any other state over an interaction interval with finite probability. From the Perron-Frobenius theorem [86] we can infer that the distribution of the demon at the end of an interaction interval in the limit of large n is given by the unique steady state eigenvector of $\mathcal{T}$:

$$\lim_{n\to\infty} \mathbf{p}^D(n\,\tau) = \mathbf{q} \quad , \quad \mathcal{T}\mathbf{q} = \mathbf{q}. \tag{5.16}$$

Because $\mathbf{q}$ is independent of n the distribution of the demon becomes periodic with respect to time after a sufficiently large number of interaction intervals ($n \to \infty$):

$$\mathbf{p}^{D,ps}(n\,\tau) = \mathbf{q} \quad , \quad \mathbf{p}^{D,ps}(n\,\tau \le t < (n+1)\tau) = e^{\mathcal{R}u}\mathbf{q} \quad , \quad u = t \bmod \tau. \tag{5.17}$$

We now give the method to obtain the circulation Φ. We solve for the periodic steady state of the demon (Eqs. 5.16 and 5.17); then we use that solution to determine the distribution of outgoing bits (p'_0, p'_1); and finally use Eq. 5.8. The distribution (p'_0, p'_1) is obtained from the steady state $\mathbf{q}$ by

$$\begin{pmatrix} p'_0 \\ \\ p'_1 \end{pmatrix} = \mathcal{P}_B e^{\mathcal{R}\tau}\mathcal{M}\mathbf{q} \quad , \quad \mathcal{P}_B \equiv \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ \\ 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix}. \tag{5.18}$$

Here, $\mathcal{M}\mathbf{q}$ gives the joint distribution of the demon and an incoming bit at the beginning of their interaction interval (after the demon has reached the periodic steady state distribution), the factor $\exp \mathcal{R}\,\tau$ evolves it over the subsequent interaction interval so that $\exp \mathcal{R}\,\tau\mathcal{M}\mathbf{q}$ is the joint distribution of the demon and the bit at the end of their interaction interval, and finally $\mathcal{P}_B$ sums over the states of the demon to give the distribution of the outgoing bit.

This calculation involves a straightforward if tedious exercise in the spectral decomposition of $\mathcal{R}$, which we detail in Appendix B. The final result is the following:

$$\Phi(\delta, \epsilon; \tau) = \frac{\delta - \epsilon}{2}\eta(\delta\epsilon, \tau), \quad \eta(\delta\epsilon, \tau) = \left[1 - \frac{1}{3}K(\tau) + \frac{\epsilon\delta}{6}J(\tau, \epsilon\delta)\right], \tag{5.19a}$$

where

$$K(\tau) = e^{-2\tau}\frac{(1 + 8\alpha + 4\sqrt{3}\beta) - (2 + 7\alpha + 4\sqrt{3}\beta)e^{-2\tau}}{3 - (2 + \alpha)e^{-2\tau}}, \tag{5.19b}$$

$$J(\tau, \epsilon\delta) = \frac{(1 - e^{-\tau})[2e^{-2\tau}(\alpha + \sqrt{3}\beta - 1)]^2}{[3(1 - \epsilon\delta e^{-\tau}) - (1 - \epsilon\delta)(2 + \alpha)e^{-2\tau}][3 - (2 + \alpha)e^{-2\tau}]}, \tag{5.19c}$$

and $\alpha = \cosh(\sqrt{3}\tau)$, $\beta = \sinh(\sqrt{3}\tau)$. These results extend to negative values of ϵ if we interpret these as indicating that gravity exerts a CW torque[3].

There is an interesting feature in Eq. 5.19. The quantity η can be shown to be non-negative from the irreversibility of dynamics. The sign of Φ is therefore determined by the difference $(\delta - \epsilon)$. One may think of two effective forces: the bias induced by the incoming bit stream, which favors $\Phi > 0$ when $\delta > 0$, and the gravitational force due to the mass, which favors $\Phi < 0$ when $\epsilon > 0$.

The thermodynamic behavior of our device is characterized by the the average

[3]For $\epsilon < 0$, the mass is raised in $A1 \to C0$ transitions and lowered in $C0 \to A1$ transitions.

work done by the demon per bit, namely,

$$W = m\,g\,\Delta h\,\Phi = k\,T \ln\left(\frac{1+\epsilon}{1-\epsilon}\right)\Phi, \tag{5.20}$$

as the mass is raised (or lowered) by Δh each time $\Delta\chi = +1(\text{or} -1)$. We have used Eq. 5.3 to write down the second equation.

To discuss the information theoretic behavior of our device we introduce the quantities

$$S_b = -\sum_{i=0,1} p_i \ln p_i \quad \text{and} \quad S_b' = -\sum_{i=0,1} p_i' \ln {p_i}'. \tag{5.21}$$

(Recall that p_0 (p_0') is the proportion of 0's in the incoming (outgoing) bit-stream, and similarly for p_1 and p_1'.) For convenience we will call these the *disorder* (per bit), although this terminology ignores correlations between successive bits in the outgoing stream. The quantity S_b is the information content of the incoming stream, and is related to its capacity to record new information, in the following sense. When $S_b = 0$ the incoming stream is a blank slate composed entirely of 0's (or entirely of 1's), and the outgoing stream contains a faithful record of CW (or CCW) rotations, as discussed earlier. When $S_b = \ln 2$ (its maximum possible value) the incoming stream is saturated with an equal mixture of 0's and 1's, and in this case the outgoing stream does not chronicle the demon's rotations. We will interpret the difference $\Delta S_B \equiv S_b' - S_b$ as a measure of the degree to which new information is written in the flow of bits, as it interacts with the demon. If we define the function

$$S(X) = -\frac{1-X}{2}\ln\frac{1-X}{2} - \frac{1+X}{2}\ln\frac{1+X}{2} \tag{5.22}$$

we can rewrite ΔS_B as

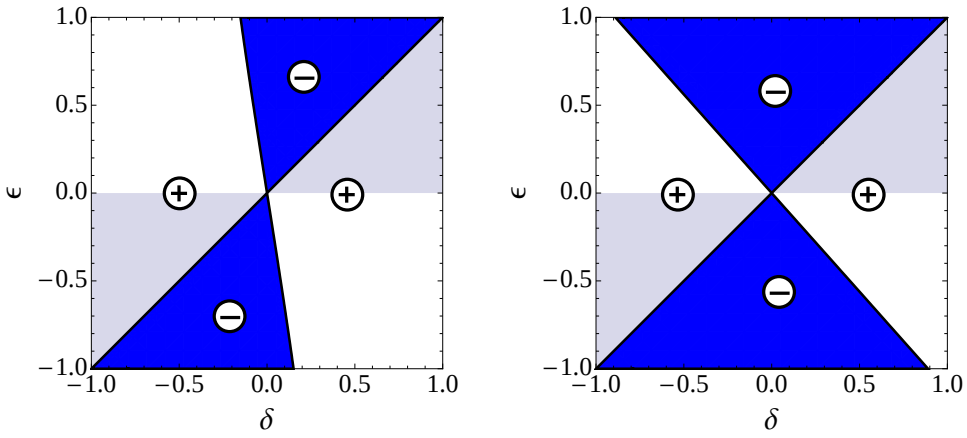
$$\Delta S_B = S(\delta - 2\Phi) - S(\delta). \tag{5.23}$$

5.3 Modes of operation

Our device displays different "useful" modes of operation. It can act both as an engine, $W > 0$, and memory eraser, $\Delta S_B < 0$. In the former case, it extracts heat from the reservoir and converts into work by raising the mass against its weight. In the eraser case, the mass gets lowered and the demon utilizes the corresponding increase in reservoir entropy to decrease the information content per bit. We discuss the details of each case in the following.

Consider the square region representing allowable values of the excess parameter δ and the weight parameter ϵ, depicted in Fig. 5.7 for $\tau = 1$ and 10. The line $\epsilon = \delta$ is the contour of zero steady state rotation; to the left of this line rotation is CCW ($\Phi < 0$), and to the right of this line the rotation is CW ($\Phi > 0$). Work per bit is positive ($W > 0$) in the two lightly shaded triangles, and thus the device acts as an engine. We can clearly see the interplay between the two parameters δ and ϵ. For example, when $\delta > \epsilon > 0$ gravity exerts a CCW torque, but the excess of incoming 0's generates a greater CW torque.

Since the rotation of the demon couples tightly to the flipping of bits (Eq. 5.7), the line $\epsilon = \delta$ (where $\Phi = 0$) is a contour along which $\Delta S_B(\delta, \epsilon; \tau) = 0$; here, there is no net rotation and no net change in the bit statistics: $p_0' = p_0$ and $p_1' = p_1$. The other solid line depicted in Fig. 5.7, running from the upper left to the lower right, is also a contour along which $\Delta S_B = 0$, representing the *inversion* of bit statistics: $p_0' = p_1$ and $p_1' = p_0$. The two lines divide the (δ, ϵ)-square into four regions, with the $+$'s and $-$'s in Fig. 5.7 denoting the sign of ΔS_B in these regions.



(a) $\tau = 1$ (b) $\tau = 10$

Figure 5.7: **Phase diagram of the engine model.** Behavior of the model as a function of δ and ϵ, for $\tau = 1$ and 10. The demon can act as an engine (lightly shaded region), an eraser (darkly shaded) or a dud (unshaded). These regions are delineated by the lines $\epsilon = 0$ and $\epsilon = \delta$, together with a third line (see text), shown passing through the second and fourth quadrants, which depends on τ and is nearly but not exactly straight. The symbols $+$ and $-$ indicate the sign of ΔS, the average change in disorder per bit. The circulation Φ is positive (CW) in the lower right half of the figure, and negative (CCW) in the upper left.

We see in Fig. 5.7 that $\Delta S_B > 0$ whenever our device acts as an engine. This is consistent with the proposition that a mechanical demon, in order to convert heat to work, must write information to a memory register. Indeed, Fig. 5.7 shows that the greater the storage capacity of the incoming bit stream, the larger the mass the demon can hoist against gravity: when presented with a blank slate ($\delta = \pm 1$) the demon can lift any mass; but when the incoming bit stream is saturated ($\delta = 0$) the demon is incapable of delivering work. Thus, a blank or partially blank memory register acts as a thermodynamic resource that gets consumed when the demon acts as an engine.

In the above description, the demon is an active rectifying agent and the bit stream merely a passive receptacle for information. From another perspective, however, the interaction with the demon presents an opportunity for the bits to evolve to a more disordered sequence of 0's and 1's. The bits' role then appears more assertive: their evolution toward greater randomness is what drives the engine, and the demon simply facilitates the process.

In the darkly shaded regions in Figs. 5.7, the demon acts as an *eraser*, removing information from the memory register: $\Delta S_B < 0$. For example, if $\delta = 0$, $mg\Delta h \gg kT$ (i.e. $\epsilon \approx 1$) and $\tau \gg 1$, then the bits arrive in an equal mixture of 0's and 1's, but each bit has sufficient time to equilibrate with the demon, hence at the end of each interaction interval the composite system is almost certainly in state $A0$, $B0$ or $C0$ (Eq. 5.5). As a result, the outgoing bits are nearly all 0's, and the memory is effectively wiped clean as the mass drops by a distance $\Delta h/2$ (on average) per interaction interval.

Our model thus reflects the interplay between two effective forces, one associated with the randomization of the bits and the other with the pull of gravity. When our model acts as an engine, it consumes one resource – a blank or partially blank memory register – to build up another: the gravitational potential of the mass. When it acts as an eraser the roles are reversed. In the unshaded regions in Figs. 5.7, both resources are squandered (the mass falls and the bits' disorder increases) and our model is a dud, accomplishing nothing useful.

5.4 Modified second law of thermodynamics

As will be proved shortly (beginning with Eq. 5.26) the model satisfies the inequality

$$W \leq kT\Delta S_B, \tag{5.24}$$

for any ϵ, δ and τ, with the equality holding only when $\epsilon = \delta$. Thus, the increase in the information content of the bit stream places an upper limit on the work that can be delivered, when the model is an engine. Analogous inequalities arise in the context of feedback control, where an external agent manipulates the system on the basis of outcomes of explicit measurements [39–50]. When our model acts as an eraser ($\Delta S_B < 0$), the relation 5.24 reveals the minimum amount of work that must be *supplied*, by the falling mass, in order to reduce the information content by a given amount. In the case of full erasure ($S'_b = 0$) this becomes Landauer's principle, $|W| > kTS_b$. Note that if we are willing to assign thermodynamic meaning to the randomness in a string of data, the relation 5.24 can be interpreted as the

second law of thermodynamics (or rather as a weak statement of it, since S'_b ignores correlations between outgoing bits): the decrease in the entropy of the reservoir, $-\Delta S_R = W/kT$, must not exceed the increase in the entropy of the bit stream:

$$\Delta S_R + \Delta S_B \geq 0. \tag{5.25}$$

While both sides of the relation 5.24 approach zero as $\epsilon \to \delta$, their ratio approaches unity in that limit (The proof of this assertion is given at the end of this section beginning with Eq. 5.36). Thus in the immediate vicinity of the line $\epsilon = \delta$, the bound represented by the relation 5.24 becomes saturated, and our model behaves with maximal efficiency, acting as a thermodynamically *reversible* engine or eraser. Note however that the *rate* at which the demon either delivers work or erases information approaches zero in this reversible limit. In the following, we first give the derivation of Eq. 5.24 and then the proof of the reversible limit.

Recall from Eq. 5.20 that

$$W = kT \ln\left(\frac{1+\epsilon}{1-\epsilon}\right)\Phi. \tag{5.26}$$

To establish the inequality 5.24 we must prove the non-negativity of the *dissipation* function:

$$\Omega \equiv \Delta S_B - \Phi \ln\frac{1+\epsilon}{1-\epsilon} \geq 0. \tag{5.27}$$

We will first prove this for the quasistatic case $\tau \to \infty$, and then extend it to finite τ.

In the quasistatic limit (specified below by the subscript "∞") we have

$$\Phi \longrightarrow \frac{\delta - \epsilon}{2} \equiv \Phi_\infty \quad , \quad \delta' \longrightarrow \epsilon \quad , \quad \Delta S_B \longrightarrow S(\epsilon) - S(\delta), \tag{5.28}$$

hence

$$\Omega \longrightarrow S(\epsilon) - S(\delta) - \frac{\delta - \epsilon}{2} \ln \frac{1+\epsilon}{1-\epsilon} \equiv \Omega_\infty. \tag{5.29}$$

Now note that

$$\Omega_\infty = 0 \quad \text{for } \epsilon = \delta \qquad \text{and} \qquad \frac{\partial}{\partial \epsilon}\Omega_\infty = \frac{\epsilon - \delta}{1-\epsilon^2} \begin{cases} > 0 & \text{if } \epsilon > \delta \\ \\ < 0 & \text{if } \epsilon < \delta \end{cases} \quad . \tag{5.30}$$

Thus for any fixed value of δ, the function $\Omega_\infty(\delta, \epsilon)$ is zero at the point $\epsilon = \delta$, and as a function of ϵ it decreases when $\epsilon < \delta$ and increases when $\epsilon > \delta$. This establishes that $\Omega_\infty \geq 0$.

We have verified by explicit numerical investigation that $\eta(\delta\epsilon, \tau)$ in Eq. 5.19 satisfies

$$0 \leq \eta \leq 1. \tag{5.31}$$

While we have not been able to establish this analytically, we believe it is related to the fact that all eigenvalues of the transition rate matrix $\mathcal{R}$ are real and non-positive (Eq. B.2), with the consequence that the composite demon-and-bit system relaxes monotonically toward equilibrium during each interaction interval.

For finite τ, the excess parameter δ' for the outgoing stream is a linear average of δ and ϵ:

$$\delta' = \delta - 2\Phi = (1-\eta)\,\delta + \eta\,\epsilon \quad , \tag{5.32}$$

using Eqs. 5.8, 5.31 and 5.28. Since $S(X)$ (Eq. 5.22) is concave ($\mathrm{d}^2S/\mathrm{d}X^2 < 0$),

$$\begin{aligned} S'_b = S(\delta') &\geq (1-\eta)\,S(\delta) + \eta\,S(\epsilon) \\ &= S(\delta) + \eta\,\big[S(\epsilon) - S(\delta)\big]. \end{aligned} \tag{5.33}$$

From Eqs. 5.28, 5.29 and the non-negativity of Ω_∞, we have

$$S(\epsilon) - S(\delta) \geq \Phi_\infty \ln \frac{1+\epsilon}{1-\epsilon}. \tag{5.34}$$

Combining Eqs. 5.33 and 5.34 we get

$$S'_b \geq S(\delta) + \eta\, \Phi_\infty \ln \frac{1+\epsilon}{1-\epsilon} = S_b + \Phi \ln \frac{1+\epsilon}{1-\epsilon} \quad , \tag{5.35}$$

which is the result we set out to establish (Eq. 5.27).

Finally, setting $kT = 1$ for convenience, we establish the result

$$\lim_{\epsilon \to \delta} \frac{W}{\Delta S} = 1 \tag{5.36}$$

to prove the reversibility of our device near $\epsilon = \delta$ line. Taking the partial derivatives of the quantities

$$W = \Phi \ln \frac{1+\epsilon}{1-\epsilon} \quad \text{and} \quad \Delta S = S(\delta') - S(\delta) \tag{5.37}$$

with respect to ϵ, at fixed δ and τ, we get

$$\begin{aligned} \frac{\partial W}{\partial \epsilon} &= \frac{\partial \Phi}{\partial \epsilon} \ln \frac{1+\epsilon}{1-\epsilon} + \frac{2\Phi}{1-\epsilon^2} \\ \frac{\partial \Delta S}{\partial \epsilon} &= \frac{\partial \Phi}{\partial \epsilon} \ln \frac{1+\delta'}{1-\delta'} \end{aligned} \tag{5.38}$$

(using $\delta' = \delta - 2\Phi$). Along the line $\epsilon = \delta$ we have $W = \Delta S = 0$ as well as

$$\frac{\partial W}{\partial \epsilon} = \frac{\partial \Delta S}{\partial \epsilon}. \tag{5.39}$$

Eq. 5.36 then follows by l'Hôpital's rule.

Chapter 6

Maxwell's refrigerator

---

[1]In the previous chapter a variant of Maxwell's demon, the so-called Szilard's engine [56], was discussed and an exactly solvable, autonomous model of this variant [85] was introduced. More recently, an analogous model for the original demon has been developed. Just as in Maxwell's proposal, this new model can effect a heat transfer against a thermal gradient without any external energy. This model too is completely autonomous, exactly solvable, and, to the best of our knowledge, the first of its kind in the field.

A schematic diagram of the model is shown in Fig. 6.1. Its consists of four components: a two-state system that plays the role of Maxwell's demon, a memory register, and two thermal reservoirs at temperatures T_c and $T_h > T_c$. The memory register is a sequence of bits (two-state systems) spaced at equal intervals along a tape that slides frictionlessly past the demon. The demon interacts with the nearest bit and with the reservoirs, as we describe in detail in the following paragraphs, and effects heat transfer between the reservoirs. If the the initial state of the memory register contains a sufficient fraction of 0's the demon is capable of transferring heat from the cold to the hot reservoir by randomizing the state of the register. As in the previous case, the present model is versatile: for certain combinations of the model

[1]This chapter is based on Ref. [93].

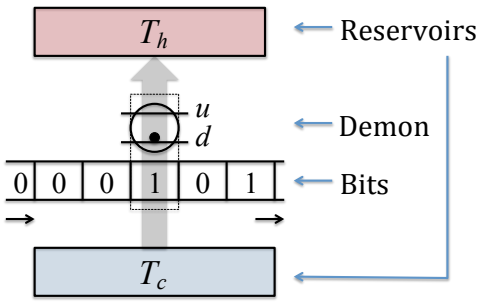


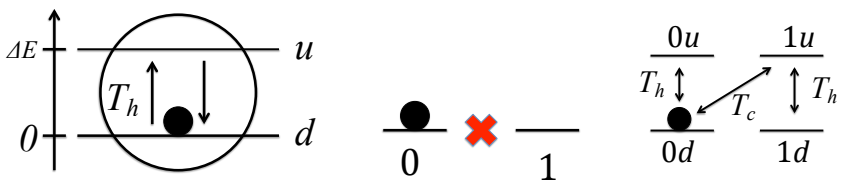
Figure 6.1: **Illustration of the setup.** The device, or demon, interacts with a sequence of bits, one at a time, while exchanging energy with two thermal reservoirs.

parameters, the demon can also act as a memory eraser.

In Sec. 6.1 a detailed description of the model is given. Sec. 6.2 gives the analyses of the model. In Sec. 6.3 the analytical results from the previous section have been utilized to represent the qualitative behavior of the model with a simple phase diagram. Sec. 6.4 addresses the irreversible behavior of our model and its relation to an effective form of the second law of thermodynamics (Eq. 6.27).

6.1 Model

The demon in this model is a two-state system with states u and d; they are characterized by an energy difference $\Delta E = E_u - E_d > 0$. The demon can make random transitions between these two states by exchanging energy with the hot reservoir, as illustrated in Fig. 6.2(a). We will refer to these transitions as the *intrinsic* dynamics of the demon, to emphasize that they involve the demon but not the flow of bits. The intrinsic transition rates satisfy the requirement of detailed



(a) (b) (c)

Figure 6.2: **Illustration of the dynamics.** (a) The demon makes intrinsic transitions between states d and u, while exchanging heat with the hot reservoir. (b) A bit cannot make intrinsic transitions between states 0 and 1. (c) The demon and bit make cooperative transitions $0d \leftrightarrow 1u$ (diagonal arrows) by exchanging heat with the cold reservoir. Vertical arrows correspond to intrinsic transitions of the demon.

balance[2] [76],

$$\frac{R_{d\to u}}{R_{d\leftarrow u}} = e^{-\beta_h \Delta E}, \tag{6.1}$$

where $\beta_h = 1/kT_h$. We parametrize these rates as

$$R_{d\to u} = \gamma(1-\sigma), \quad R_{d\leftarrow u} = \gamma(1+\sigma), \quad \sigma = \tanh\frac{\beta_h \Delta E}{2} \tag{6.2}$$

where $\gamma > 0$ sets a characteristic rate for these transitions, and $0 < \sigma < 1$.

Each bit (Please refer to footnote 2 in Chap. 5 for a clarification of the use of the term.) has two states, 0 and 1, with equal energies. We assume there are no intrinsic transitions between these two states (Fig. 6.2(b)). That is, the state of the bit can change only via interaction with the demon, as we now discuss.

At any instant in time, the demon interacts only with the nearest bit. As

[2]If these were the only dynamics of the demon Eq. 6.1 ensures that the demon approaches an equilibrium distribution given by the appropriate Boltzmann distribution: $p_u^{D,\,eq} = \exp(-\Delta E/kT_h)/\mathcal{N}, \quad p_u^{D,\,eq} = 1/\mathcal{N}, \quad \mathcal{N} = 1 + \exp(-\Delta E/kT_h)$.

a result, it interacts sequentially with the bits as they pass by. The duration of interaction with each bit is $\tau = l/v$, where l is the spacing between bits and v is the constant speed of the tape. During one such *interaction interval*, the demon and the nearest bit can make *cooperative* transitions: if the bit is in state 0 and the demon is in state d, then they can simultaneously flip to states 1 and u, and vice-versa (Fig. 6.2(c)). We will use the notation $0d \leftrightarrow 1u$ to denote these transitions, which are accompanied by an exchange of energy with the cold reservoir. The corresponding transition rates must satisfy the detailed balance condition[3]

$$\frac{R_{0d\to 1u}}{R_{0d\leftarrow 1u}} = e^{-\beta_c \Delta E}\,, \tag{6.3}$$

where $\beta_c = 1/kT_c$, and we will parametrize them as follows:

$$R_{0d\to 1u} = 1-\omega, \quad R_{0d\leftarrow 1u} = 1+\omega, \quad \omega = \tanh\frac{\beta_c \Delta E}{2}\,, \tag{6.4}$$

with $0 < \omega < 1$. For later convenience, we also define

$$\epsilon = \frac{\omega - \sigma}{1-\omega\sigma} = \tanh\frac{(\beta_c - \beta_h)\Delta E}{2}\,, \tag{6.5}$$

whose value, $0 < \epsilon < 1$, quantifies the temperature difference between the two reservoirs.

Finally, we assume that the incoming bit stream contains a mixture of 0's and 1's, with probabilities p_0 and p_1, respectively, with no correlations among the bits.

[3]If T_c were the only reservoir in the model Eq. 6.3 would have been a requirement for the model to relax to the Boltzmann distribution with respect to T_c. We assume that the coupling of the intrinsic transitions of the demon to a different reservoir T_h does not affect this requirement. If T_h were made equal to T_c, Eq. 6.3 would be required to guarantee a relaxation to the Boltzmann distribution of the joint system of the demon and the bit with respect to the common temperature.

Let

$$\delta \equiv p_0 - p_1 \tag{6.6}$$

quantify the excess of 0's among incoming bits.

We thus have the following dynamics. When a fresh bit arrives to interact with the demon, its state is 0 or 1. The demon and bit subsequently interact for a time τ, making the transitions shown in Fig. 6.2(c), thereby exchanging energy with the reservoirs. The state of the bit at the end of the interaction interval is then preserved as the bit joins the outgoing stream, and the next bit in the sequence moves in to have its turn with the demon. The parameters γ, σ and ω define the intrinsic and cooperative transition rates (Eqs. 6.2, 6.4), τ gives the duration of interaction with each bit, and δ specifies the statistics of the incoming bits. Under these dynamics, the demon evolves to a periodic steady state, in which its behavior is statistically the same from one interaction interval to the next.

Before proceeding to the solution of these dynamics, we discuss heuristically how our model can achieve the systematic transfer of heat from the cold to the hot reservoir. For this purpose let us assume that each incoming bit is in state 0, hence $\delta = 1$. At the start of a particular interaction interval, the joint state of the demon and newly arrived bit is either $0u$ or $0d$. The demon and bit then evolve together for a time τ, according to the transitions shown in Fig. 6.2(c). If the joint state at the end of the interaction interval is $0u$ or $0d$, then it must be the case that every transition $0d \to 1u$ was balanced by a transition $0d \leftarrow 1u$, hence no net energy was absorbed from the cold reservoir. If the final state is $1u$ or $1d$, then we can infer

that there was one *net* transition from $0d$ to $1u$, and a quantity of energy ΔE was absorbed from the cold reservoir. This amounts to *thermal rectification*: over the course of one interaction interval, energy can be withdrawn from the cold reservoir but not delivered to it. Moreover, a record of this process is imprinted in the bit stream, as every outgoing bit in state 1 indicates the absorption of energy ΔE from the cold reservoir. *Since the demon also exchanges energy with the hot reservoir, and since energy cannot accumulate indefinitely within the demon*, in the long run we get a net flux of energy from the cold to the hot reservoir, proportional to the rate at which 1's appear in the outgoing bit stream.

More generally, when the incoming bit stream contains a mixture of 0's and 1's, a simple rule emerges: over each interaction interval, the net change in the value of the interacting bit (-1, 0 or 1) determines the net amount of energy absorbed from the cold reservoir ($-\Delta E$, 0 or ΔE). As a result, an excess of 0's in the incoming bit stream (that is, $\delta > 0$) produces a statistical bias that favors the flow of heat from the cold to the hot reservoir, while an excess of 1's ($\delta < 0$) produces the opposite bias. This bias either competes with or enhances the thermodynamic bias due to the temperature difference between the two reservoirs. The demon thus affects the flow of energy between the reservoirs, and modifies the states of the bits in the memory register. We now investigate quantitatively the interplay between these two effects.

Once the demon has reached its periodic steady state, let p'_0 and p'_1 denote the fractions of 0's and 1's in the outgoing bit stream, and let $\delta' = p'_0 - p'_1$ denote the

excess of outgoing 0's. Then

$$\Phi \equiv p_1' - p_1 = \frac{\delta - \delta'}{2} \tag{6.7}$$

represents the average production of 1's in the outgoing bit stream, relative to the incoming bit stream. Since each transition $0 \to 1$ is accompanied by the absorption of energy from the cold reservoir (Fig. 6.2(c)), the average transfer of energy from the cold to the hot reservoir, per interaction interval, is given by

$$Q_{c \to h} = \Phi \Delta E \,. \tag{6.8}$$

A positive value of $Q_{c \to h}$ indicates that our device pumps energy against a thermal gradient, like the creature imagined by Maxwell.

To quantify the information-processing capability of the demon, let

$$S(\delta) = -\sum_{i=0}^{1} p_i \ln p_i = -\frac{1-\delta}{2} \ln \frac{1-\delta}{2} - \frac{1+\delta}{2} \ln \frac{1+\delta}{2} \tag{6.9}$$

denote the information content, per bit, of the incoming bit stream[4], and define $S(\delta')$ by the same equation, for the outgoing bit stream. Then

$$\Delta S_B \equiv S(\delta') - S(\delta) = S(\delta - 2\Phi) - S(\delta) \tag{6.10}$$

provides a measure of the extent to which the demon increases the information content of the memory register. We will interpret a positive value of ΔS_B to indicate that the demon *writes* information to the bit stream, while a negative value indicates *erasure*. (More precisely, since $S(\delta')$ neglects the correlations that might arise between the outgoing bits, ΔS_B reflects the change in the Shannon information of the *marginal* probability distribution of each outgoing bit.)

[4]This is the same function as in Eq. 5.22 in the last chapter.

6.2 Analysis

As in the model of the last chapter, the demon in the present model also relaxes to a periodic steady state distribution. As the proof of this assertion follows the same line of logic as in the last chapter (given in Sec. 5.2) we will have only a sketch of it in the present section. Then we will sketch the derivation of Φ, as defined in Eq. 6.7.

Let $\mathbf{p}^D(0) = (p_u(0), p_d(0))^T$ be the distribution of the demon at $t = 0$. The joint distribution of the demon and the first bit at the same moment is given by the product distribution

$$\mathbf{p}^{DB}(0) \equiv \begin{pmatrix} p_{u0}(0) \\ p_{d0}(0) \\ p_{u1}(0) \\ p_{d1}(0) \end{pmatrix} = \mathcal{M}\,\mathbf{p}^D(0) \quad , \quad \mathcal{M} = \begin{pmatrix} p_0 & 0 \\ 0 & p_0 \\ p_1 & 0 \\ 0 & p_1 \end{pmatrix}. \tag{6.11}$$

The joint distribution evolves according to the master equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{p}^{DB}(t) = \mathcal{R}\,\mathbf{p}^{DB}(t) \quad , \quad \mathcal{R} = \begin{pmatrix} \bullet & \gamma(1-\sigma) & 0 & 0 \\ \gamma(1+\sigma) & \bullet & 1+\omega & 0 \\ 0 & 1-\omega & \bullet & \gamma(1-\sigma) \\ 0 & 0 & \gamma(1+\sigma) & \bullet \end{pmatrix}. \tag{6.12}$$

(Elements of $\mathcal{R}$ have already been discussed in the last section, and the dots are determined by the normalization restriction that elements of each column should add up to zero [76].) At $t = \tau$ we have

$$\mathbf{p}^{DB}(\tau) = e^{\mathcal{R}\tau}\mathbf{p}^{DB}(0). \tag{6.13}$$

The marginal distribution of the demon at $t = \tau$ is given by

$$\mathbf{p}^D(\tau) = \mathcal{P}_D \mathbf{p}^{DB}(\tau) \quad , \quad \mathcal{P}_D = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}. \tag{6.14}$$

Combining Eqs. 6.11, 6.13 and 6.14 we get

$$\mathbf{p}^D(\tau) = \mathcal{T} \mathbf{p}^D(0) \quad , \quad \mathcal{T} = \mathcal{P}_D e^{\mathcal{R}\tau} \mathcal{M}. \tag{6.15}$$

Clearly, $\mathcal{T}$ is the transition matrix of the demon whose component T_{ij} gives the probability for the demon to make a transition to state $i \in \{u, d\}$ from state $j \in \{u, d\}$ over the interaction interval. Because all bits are initiated in the same distribution, are independent of each other, and interact with the demon according to the same rules the matrix $\mathcal{T}$ is also the transition matrix of the demon for any interaction interval. Then the marginal distribution of the demon at $t = n\,\tau$ is

$$\mathbf{p}^D(n\,\tau) = \mathcal{T}^n\,\mathbf{p}^D(0). \tag{6.16}$$

The matrix $\mathcal{T}$ must be positive because it is a transition matrix. From the Perron-Frobenius theorem [86] we have

$$\lim_{n\to\infty} \mathbf{p}^D(n\,\tau) = \mathbf{q} \quad , \quad \mathcal{T}\mathbf{q} = \mathbf{q}. \tag{6.17}$$

As $\mathbf{q}$ is independent of n, the demon attains a periodic steady state distribution in the limit of large n:

$$\mathbf{p}^{D,ps}(n\,\tau) = \mathbf{q} \quad , \quad \mathbf{p}^{D,ps}(n\,\tau \le t < (n+1)\tau) = e^{\mathcal{R}u}\mathbf{q} \quad , \quad u = t \bmod \tau. \tag{6.18}$$

Solving for Φ (as defined in Eq. 6.7) involves first solving for the periodic steady state distribution of the demon, then using it to determine the distribution

of the outgoing bits $(p'_0, p'_1)^T$, from which Φ follows by Eq. 6.7. After the demon reaches its periodic steady state distribution, the joint distribution of the demon and the outgoing bit is given by $\exp(\mathcal{R}\tau)\mathcal{M}\mathbf{q}$: the joint distribution of the demon and incoming bit is $\mathcal{M}\mathbf{q}$ and it evolves in the subsequent interaction interval to $\exp(\mathcal{R}\tau)\mathcal{M}\mathbf{q}$. (See Eq. 6.13.) The marginal distribution of the outgoing bit is then given by

$$(p'_0, p'_1)^T = \mathcal{P}^B e^{\mathcal{R}\tau}\mathcal{M}\mathbf{q} \quad , \quad \mathcal{P}^B \equiv \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}. \tag{6.19}$$

We performed these calculations using Mathematica [87], and then simplified the results substantially by hand, finally obtaining

$$\Phi = \frac{\delta - \epsilon}{2}\,\eta \quad , \quad \eta = \frac{\nu_2 P + \nu_3 Q}{P + Q}, \tag{6.20a}$$

$$\begin{aligned}
P &= \mu_2\,(\mu_4\nu_3 + \mu_1\nu_1) && , \quad Q = \mu_3\,(\mu_4\nu_2 + \mu_1\nu_1), \\
\nu_1 &= 1 - e^{-2\gamma\,\tau} && , \quad \mu_1 = (\delta + \sigma)\,\omega, \\
\nu_2 &= 1 - e^{-(1+\gamma-\alpha)\,\tau} && , \quad \mu_2 = \alpha + \gamma + \sigma\,\omega, \\
\nu_3 &= 1 - e^{-(1+\gamma+\alpha)\,\tau} && , \quad \mu_3 = \alpha - \gamma - \sigma\,\omega, \\
\alpha &= \sqrt{1 + \gamma^2 + 2\gamma\sigma\omega} && , \quad \mu_4 = 1 - \delta\,\omega.
\end{aligned} \tag{6.20b}$$

We now give a quickly accessible analysis of our model in the limit $\gamma \to \infty$. In this limit, the intrinsic transitions of the demon are fast compared to its cooperative transitions with the current bit. In Fig. 6.2(c), this implies vertical transitions are fast compared to the diagonal transitions. As a result the two pairs of states $(u0, d0)$ and $(u1, d1)$ quickly "equilibrate" themselves with respect to the intrinsic rates,

$$\frac{p_{u0}}{p_{d0}} = e^{-\Delta E/kT_h} = \frac{p_{u1}}{p_{d1}}. \tag{6.21}$$

(These relations are valid only up to $O(1/\gamma^0)$; smaller corrections of order $O(1/\gamma^1)$ exist.) We can rewrite the associated probabilities as

$$p_{u0}(t) = \frac{e^{-\Delta E/kT_h}}{1+e^{-\Delta E/kT_h}}[\,p_{u0}(t)+p_{d0}(t)] = \frac{1-\sigma}{2}p_0(t), \tag{6.22a}$$

and similarly for the others

$$p_{d0}(t) = \frac{1+\sigma}{2}p_0(t) \quad , \quad p_{u1}(t) = \frac{1-\sigma}{2}p_1(t) \quad , \quad p_{d1}(t) = \frac{1+\sigma}{2}p_1(t) \tag{6.22b}$$

We have used the definition of σ from Eq. 6.2. Inserting these expressions in the master equation 6.12 we get the following effective equation for the bit distributions:

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} p_0(t) \\ \\ p_1(t) \end{pmatrix} = \begin{pmatrix} -a & b \\ \\ a & -b \end{pmatrix}\begin{pmatrix} p_0(t) \\ \\ p_1(t) \end{pmatrix}, \tag{6.23}$$

with $a = (1-\omega)(1+\sigma)/2$ and $b = (1+\omega)(1-\sigma)/2$. Integrating Eq. 6.23 over one interaction interval, $0 \le t \le \tau$, with the initial conditions $p_0(0) = p_0$ and $p_1(0) = p_1$, and then setting $p_1(\tau) = p_1'$ in Eq. 6.7 we get, after some algebra,

$$\Phi = \frac{\delta-\epsilon}{2}\eta \quad , \quad \eta = 1-e^{-(1-\sigma\omega)\tau}. \tag{6.24}$$

This result also follows from our general solution for η (Eq. 6.20), evaluated in the limit $\gamma \to \infty$:

$$\nu_1, \nu_3 \to 1 \quad , \quad \nu_2 \to 1-e^{-(1-\sigma\omega)\tau} \quad , \quad Q/P \to 0\,, \tag{6.25}$$

hence $\eta \to 1-e^{-(1-\sigma\omega)\tau}$.

The quantity η in Eq. 6.20a is positive:

$$\eta > 0 \tag{6.26}$$

Our general expression for η (in Eq. 6.20), while exact, is sufficiently complex that we are unable to derive this inequality directly. Instead we will show in Appendix C that the inequality 6.26 follows from *modified Clausius inequality*:

$$Q_{c\to h}(\beta_h - \beta_c) + \Delta S_B \geq 0\,. \tag{6.27}$$

(It will be derived in Sec. 6.4.) An important consequence of Ineq. 6.26 is that the sign of Φ is the same as that of $(\delta - \epsilon)$. We can think of two effective forces: the bias induced by the incoming bit stream, which favors $\Phi > 0$ when $\delta > 0$ (as discussed above), and the temperature gradient, quantified by ϵ, which favors $\Phi < 0$ (Eq. 6.8). When these compete, the winner is determined by the difference $(\delta - \epsilon)$.

6.3 Phase diagram

Here we use the results obtained in the last section to investigate the behavior of our model in the periodic steady state. To that end, we fix γ and ω and construct a phase diagram that illustrates the dependence on δ and ϵ, for three different values of τ, shown in Fig. 6.3. Let us consider the different regions of this diagram, working our way from right to left.

From Eqs. 6.8, 6.20a and 6.26 it follows that $Q_{c\to h} > 0$ when $\delta > \epsilon$, shown as the most darkly shaded region in Fig. 6.3. Here, a surplus of incoming 0's prevails over the temperature difference and our demon generates a flow of energy from the cold to the hot reservoir. Moreover, Eq. 6.27 reveals that $\Delta S_B > 0$ in this region (since $\beta_h < \beta_c$). This agrees with the consensus described earlier: in order for a physical device to act in the manner of Maxwell's demon, it must write information

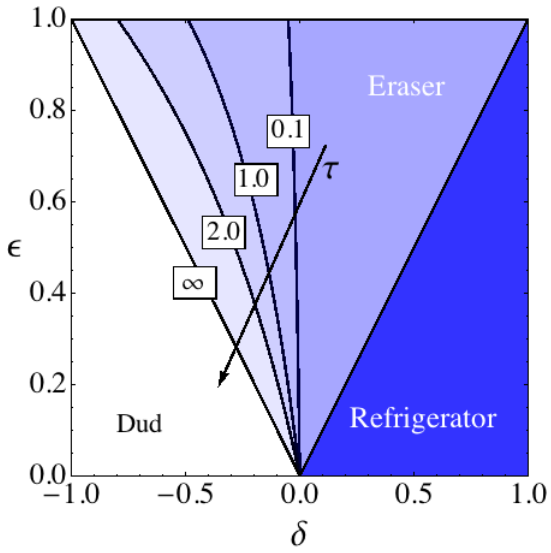


Figure 6.3: **Phase diagram of our model at fixed $\gamma = 1$ and $\omega = 1/2$.** The parameter δ specifies the incoming bit statistics, and ϵ is a rescaled temperature difference (Eq. 6.5). In the most darkly shaded region the demon acts as a refrigerator ($Q_{c\to h} > 0$), while in the lightly shaded regions it acts as an eraser ($\Delta S_B < 0$). The left boundary of the eraser region is shown for $\tau = 0.1$, 1.0, 2.0 and ∞. In the blank region at the lower left, our model exhibits neither behavior (see text).

to a physical memory register. In this sense, a bit stream with a low information content can be viewed as a thermodynamic resource, which can be expended (by writing to the available memory) in order to achieve refrigeration.

Now consider the region $\epsilon > \delta > 0$, in which the surplus of 0's in the incoming bit stream is not sufficient to overcome the temperature gradient, and energy flows from the hot to the cold reservoir. Since $\Phi < 0$ we get $\delta' > \delta > 0$ (Eq. 6.7). This in turn implies $\Delta S_B < 0$, as $S(\delta)$ is a concave function with a maximum at $\delta = 0$. In this region the demon acts as an eraser, lowering the information content of the bit stream, but the price paid for this erasure is the passage of heat from the hot to

the cold reservoir.

In the region $\delta < 0$, energy flows from the hot to the cold reservoir (Eqs. 6.8, 6.20a and 6.26), but the value of ΔS_B depends on all the model parameters. In Fig. 6.3, for three different values of τ, we show the line corresponding to $\Delta S_B = 0$. To the right of this line we have $\Delta S_B < 0$ and the device acts as an eraser. To the left we have $\Delta S_B > 0$, indicating that the information content of the bit stream increases.

Examining the phase diagram as a whole, we see that in the shaded regions our model reaches a steady state in which one thermodynamic resource is replenished at the expense of another. Either energy is pumped against a thermal gradient at the cost of writing information to memory (the refrigerator regime), or else memory is made available, by erasure, at the expense of allowing energy to flow from the hot to the cold reservoir (the eraser regime). The boundary between these two behaviors is the line $\delta = \epsilon$. In the unshaded region at the far left, both resources are consumed, as energy flows down the thermal gradient and information is written to the bit stream.

In summary, we have constructed a simple, solvable model of an autonomous physical system that can mimic the behavior of the "neat-fingered being" in Maxwell's thought experiment, generating a systematic flow of energy against a thermal gradient without the input of external work. While Maxwell's creature accomplishes this with intelligence, our inanimate device requires only a memory register to which information can be written. Alternatively, our demon can harness the flow of energy from hot to cold in order to erase information from the register.

6.4 Modified Clausius inequality

We now place our model within the context of second law of thermodynamics. In the periodic steady state we can derive the inequality 6.27

$$Q_{c\to h}(\beta_h - \beta_c) + \Delta S_B \geq 0\,. \tag{6.28}$$

The first term on the left side is the steady-state change in thermodynamic entropy due to the flow of heat, and the second term is the change in information entropy, per interaction interval. Inequality 6.28 can be viewed as a modified Clausius inequality, in which the information entropy of a random sequence of data is explicitly assigned the same thermodynamic status as the physical entropy associated with the transfer of heat. Inequality 6.28 is actually a weak version of this inequality, as we neglect correlations that might arise among the outgoing bits.[5] Thus, our model provides support for the consensus mentioned earlier [58, 59, 60], and in particular the relation 6.28 is consistent with Landauer's principle [58], which states that a thermodynamic cost must be paid for the erasure of memory. In Landauer's work

[5]It is not obvious a priori whether the outgoing bits would be correlated to each other or not. However, a heuristic argument can be proposed in favor of correlation. The demon gets correlated with the current bit because of their interactions; when the new bit comes, partial information about the outgoing bit can be transmitted via the demon, thus correlating the new bit with the previous bit. Intuitively, this correlation must be small for small interaction interval τ, because the demon hardly has any time to correlate with the bit, and also for large τ, because the demon and the bit reach a steady state distribution which is a product of their marginal distributions (Eq. 6.31) thus devoid any correlation. We have checked these intuitions numerically. An indirect analytical justification will be presented following the relation 6.40.

this cost appears as the dissipation of energy into a single thermal reservoir, whereas in our model it is the transfer of energy from a hot to a cold reservoir. In the following we present a derivation of the relation 6.28 based on the properties of the dynamics.

During any interaction interval, the joint distribution of the demon and the interacting bit evolves according to the master equation (Eq. 6.12),

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{p}^{DB}(t) = \mathcal{R}\,\mathbf{p}^{DB}(t), \tag{6.29}$$

where $\mathcal{R}$ is given in Eq. 6.12. For very long interaction intervals ($\tau \to \infty$), the combined system relaxes to the stationary state

$$\overline{\mathbf{p}}^{DB} = \frac{1}{\mathcal{N}}\left(1, \mu, \mu\nu, \mu^2\nu\right)^T, \quad \mu = \frac{1+\sigma}{1-\sigma}, \quad \nu = \frac{1-\omega}{1+\omega}, \quad \mathcal{N} = (1+\mu)(1+\mu\nu)\,, \tag{6.30}$$

which satisfies $\mathcal{R}\overline{\mathbf{p}}^{DB} = \mathbf{0}$. Note that $\overline{\mathbf{p}}^{DB}$ is actually a product of marginal distributions $\overline{\mathbf{p}}^{D}$ and $\overline{\mathbf{p}}^{B}$ for the demon and bit:

$$\overline{p}^{DB}_{ij} = \overline{p}^{D}_{i}\,\overline{p}^{B}_{j}, \quad i \in \{u, d\}, \quad j \in \{0, 1\}, \tag{6.31a}$$

$$\overline{\mathbf{p}}^{D} = (1, \mu)^T/(1+\mu), \quad \overline{\mathbf{p}}^{B} = (1, \mu\nu)^T/(1+\mu\nu). \tag{6.31b}$$

To keep our formula cleaner, we shall replace the joint distribution $\mathbf{p}^{DB}$ by simply $\mathbf{p}$ in the rest of this chapter. The irreversible approach of $\mathbf{p}(t)$ toward $\overline{\mathbf{p}}$ is described by the *relative entropy* [88],

$$D(\mathbf{p}||\overline{\mathbf{p}}) = \sum_m p_m \ln \frac{p_m}{\overline{p}_m} \geq 0\,. \tag{6.32}$$

Here and in what follows, we use the index m to indicate a joint state of the demon and the bit, $m \in \{0u, 0d, 1u, 1d\}$, reserving i and j for the demon and the bit,

respectively, as in Eq. 6.31a. A standard calculation [76] shows that D is a Lyapunov function, that is it satisfies

$$\frac{\mathrm{d}}{\mathrm{d}t} D(\mathbf{p}||\overline{\mathbf{p}}) \leq 0\,, \tag{6.33}$$

where the equality holds only when $\mathbf{p} = \overline{\mathbf{p}}$. Thus, as measured by relative entropy, any initial $\mathbf{p} \neq \overline{\mathbf{p}}$ evolves monotonically toward $\overline{\mathbf{p}}$, although for finite interaction intervals this relaxation is interrupted by the arrival of the next bit. We now use these properties to derive the inequality 6.27.

Let $\mathbf{p}_0$ and $\mathbf{p}_\tau$ denote the joint distributions of the demon and a bit at the beginning and end of a given interaction interval, respectively, and similarly define $\mathbf{p}_0^D$, $\mathbf{p}_\tau^D$, $\mathbf{p}_0^B$ and $\mathbf{p}_\tau^B$ for the marginal distributions of the demon and the bit. Eq. 6.33 implies

$$D(\mathbf{p}_0||\overline{\mathbf{p}}) - D(\mathbf{p}_\tau||\overline{\mathbf{p}}) \geq 0. \tag{6.34}$$

Using Eqs. 6.32 and 6.31a we rewrite the left side of this equation as

$$S_\tau - S_0 - \sum_{i\in\{u,d\}} \left(p_{\tau,i}^D - p_{0,i}^D\right) \ln \overline{p}_i^D - \sum_{j\in\{0,1\}} \left(p_{\tau,j}^B - p_{0,j}^B\right) \ln \overline{p}_j^B\,, \tag{6.35}$$

where $S_0 = -\sum_k p_{0,k} \ln p_{0,k}$ and $S_\tau = -\sum_k p_{\tau,k} \ln p_{\tau,k}$ are the information entropies of the joint distributions of the demon and the bit at the beginning and end of the interaction interval. Let us now evaluate Eq. 6.35, assuming the demon has reached its periodic steady state.

The joint entropy S can be written as [88]

$$S = S^D + S^B - I(D;B) \quad , \quad I(D;B) \geq 0, \tag{6.36}$$

where S^D is the marginal entropy of the demon, S^B is the marginal entropy of the bit, and the *mutual information* $I(D;B)$ quantifies the degree of correlation

between them. By construction, the demon and bit are uncorrelated at the start of the interaction interval, hence $I_0(D;B) = 0$. In the periodic steady state we have $S_\tau^D = S_0^D$, because the demon starts and ends in the same distribution. Hence the difference $S_\tau - S_0$ in Eq. 6.35 can be replaced by $\Delta S_B - I_\tau(D;B)$. We also have $\mathbf{p}_0^D = \mathbf{p}_\tau^D$ in the periodic steady state, so the first sum appearing in Eq. 6.35 vanishes.

Once the period steady state has been reached, the bit distributions $\mathbf{p}_0^B$ and $\mathbf{p}_\tau^B$ correspond to the statistics of the incoming and outgoing bit streams:

$$p_{0,j}^B = p_j \quad , \quad p_{\tau,j}^B = p_j' \quad , \quad j \in \{0,1\}\,, \tag{6.37}$$

hence $p_{\tau,0}^B - p_{0,0}^B = -(p_{\tau,1}^B - p_{0,1}^B) = \Phi$ (Eq. 6.7). The last term in Eq. 6.35 can now be rewritten, using Eqs. 6.30, 6.2 and 6.4, as

$$-\sum_{j\in\{0,1\}} \left(p_{\tau,j}^B - p_{0,j}^B\right) \ln \overline{p}_j^B = \Phi \ln(\mu\nu) = Q_{c\to h}(\beta_h - \beta_c). \tag{6.38}$$

Collecting these results, we get

$$D(\mathbf{p}_0||\overline{\mathbf{p}}) - D(\mathbf{p}_\tau||\overline{\mathbf{p}}) = \Delta S_B - I_\tau(D;B) + Q_{c\to h}(\beta_h - \beta_c), \tag{6.39}$$

which then combines with Eq. 6.34 to give us

$$Q_{c\to h}(\beta_h - \beta_c) + \Delta S_B \geq I_\tau(D;B) \geq 0. \tag{6.40}$$

An alternative derivation of this result can be constructed using the integral fluctuation theorem for total entropy production [89].

The first inequality in 6.40 is stronger than the modified Clausius statement, inequality 6.28. This underscores the fact that the inequality 6.28 is a weak statement of the second law of thermodynamics (as it applies to our model), since it

neglects correlations in the outgoing bits: the quantity ΔS_B is defined in terms of the marginal distribution of each bit. In reality the bits do develop correlations via their interactions with the demon as discussed in footnote 5 in page 103. (Let it be emphasized that explicit numerical simulations indicate that these correlations are small, but not zero.) If these correlations were to be taken into account, then the net change in the Shannon entropy per bit would have a value slightly lower than ΔS_B, and inequality 6.28 would be replaced by a somewhat stronger bound. These considerations are reflected, somewhat indirectly, by the term $I_\tau(D;B)$ in the relation 6.40.

Finally, note that

$$\frac{\overline{p}_1^B}{\overline{p}_0^B} = \mu\nu = \frac{1-\epsilon}{1+\epsilon} \quad , \quad \frac{p_1}{p_0} = \frac{1-\delta}{1+\delta}\,, \tag{6.41}$$

using Eqs. 6.30 and 6.31b, and the definitions of σ and ϵ, Eqs. 6.2 and 6.5, respectively. Thus, when $\delta = \epsilon$, the incoming bits arrive in the stationary distribution $\overline{\mathbf{p}}$. In this case, no relaxation occurs during the interaction interval; the equality holds in Eqs. 6.33 and 6.34; the outgoing bits depart with the same distribution; and $\Phi = 0$. When $\delta \neq \epsilon$, Eqs. 6.33 and 6.34 are both strict inequalities, and therefore so is the modified Clausius inequality (Eq. 6.28).

Chapter 7

Summary and future outlook

We need a strong theoretical framework for the design and control of useful efficient molecular machines. The theory of Markov processes provides the natural setting because of the inherent stochastic motion of any molecular system. Following an often employed experimental strategy of periodic modulation of external stimuli, a theoretical framework of stochastic pumping has been developed where the dynamical parameters of a Markov model are periodic functions of time. Several recent theoretical results were mentioned in the introduction. In the first part of the thesis, we have been concerned with the result of the no-pumping theorem (NPT) which states the minimal conditions necessary to generate any systematic probability current. We have given an alternative and simpler-than-the-original proof of the NPT using an elementary graph theoretic construction (Sec. 3.1). Motivated by recent experimental results, we have also proposed and analyzed a new class of "hybrid" models combining elements of both the purely discrete and the purely continuous descriptions prevalent in the field (Ch. 4). By proving the NPT for these hybrid models we have also given a detailed theoretical justification of the original experimental observation [13]. An extension of the NPT to open stochastic systems has also been developed (Sec. 3.2).

The second part of the thesis (Chs. 5 - 6) is concerned with the age-old puzzle of "Maxwell's demon". This is a notorious thought experiment proposed by James Clerk Maxwell where an intelligent being can systematically violate the second law of thermodynamics by continuous rectification of thermal fluctuations. We have proposed two exactly solvable, autonomous models that reproduce the actions of the demon without any help from an external agent or an explicit thermodynamic force. The first model can rectify the thermal fluctuations of a single heat reservoir and convert them into work (Ch. 5), in (apparent) violation of the Kelvin-Planck statement of the second law. The second model can create a heat flow against a thermal gradient without expenditure of work (Ch. 6), in (apparent) violation of the Clausius inequality. To the best of our knowledge, these are the first such models in the field.

There are indications that the equations used in our proof of the NPT can encompass some of the other theoretical results on stochastic pumps. It was noted in the main text that the detailed balance restrictions, such as Eqs. 3.9 and 3.13, are the starting relations for the derivation of the pumping quantization theorem (PQT) of Refs. [32, 37, 38]. The periodicity conditions, such as Eqs. 3.7 and 3.12, and the above detailed balance restrictions are also capable of determining the number of independent currents in any given pumping protocol (This point has not been discussed in the thesis). This result is strongly reminiscent of the pumping restriction theorem (PRT) of Ref. [30]. It will be interesting to see if the two results are equivalent to each other or one is stronger than the other.

Our discussions of the Maxwell's demon were based on schematic models rather

than realistic systems. As a continuation of the project we are presently investigating an explicitly mechanical system for the engine model (in Ch. 5) involving a contraption with paddles, axles and pulleys, immersed in a gas of particles. It will be interesting to come up with an equivalent system for the refrigerator model (in Ch. 6). There are some interesting problems in the biological context in connection with this research. Many biomolecular processes such as DNA replication and protein recognition involve information processing by small systems. Is it possible to arrive at a thermodynamic description of these processes along the line of Maxwell's demon? Can the modified forms of the second law (Eqs. 5.25 and 6.27) give realistic bounds for the efficiency of these processes? These are some interesting questions to address in future.

Appendix A

Derivation of Eq. 4.21

In Sec. 4.4 we motivated the validity of Eq. 4.21

$$\Phi^{ps}_{\to i} = \Phi^{ps}_{i\to} = \Phi^{d,ps}_i(x) = \Phi^{ps}_{\to i+1} \cdots \equiv \Phi.$$

on conceptual grounds. Here we establish it directly from the master Eqs. 4.6 and 4.7, evaluated in the periodic steady state, Eq. 4.14. We begin by assuming that the source and sink terms in Eq. 4.7 are displaced away from the track-ends, as illustrated in Fig. A.1; Eq. 4.7 then becomes

$$\frac{\partial p_i(x,t)}{\partial t} = -\frac{\partial}{\partial x} J^d_i(x,t) + \delta(x-\epsilon) J_{i\to}(t) - \delta(x-l+\epsilon) J_{\to i+1}(t), \tag{A.1}$$

where $\epsilon > 0$. We recover the situation described in the main body of the text in the limit $\epsilon \to 0$.

In the periodic steady state we have $P^{ps}_i(t+\tau) = P^{ps}_i(t)$ for each station i. Equivalently, $\int_\tau \mathrm{d}t\, \mathrm{d}P^{ps}_i(t)/\mathrm{d}t = 0$, which from Eqs. 4.6 and 6.7 implies

$$\Phi^{ps}_{\to i} = \Phi^{ps}_{i\to}. \tag{A.2}$$

This is the first part of Eq. 4.21. Similarly, for each track i we have $\int_\tau \mathrm{d}t\, \partial p^{ps}_i(x,t)/\partial t = 0$, which combines with Eq. A.1 and 6.7 to give

$$\frac{\partial}{\partial x} \Phi^{d,ps}_i(x) = \delta(x-\epsilon)\Phi^{ps}_{i\to} - \delta(x-l+\epsilon)\Phi^{ps}_{\to i+1}. \tag{A.3}$$

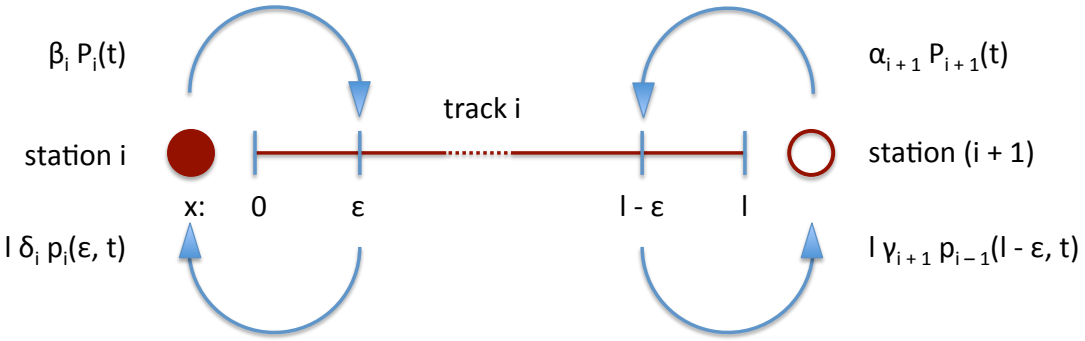


Figure A.1: **Details of source and sink.** Source and sink terms are moved away from the track-ends by a parameter $0 < \epsilon < l/2$.

From the reflective boundary conditions, $J_i^d(0,t) = J_i^d(l,t) = 0$, we have

$$\Phi_i^{d,ps}(0) = \Phi_i^{d,ps}(l) = 0. \tag{A.4}$$

Finally, solving Eq. A.3 with boundary conditions Eq. A.4 we get

$$\Phi_i^{d,ps}(x) = \begin{cases} 0 & \text{for} \quad 0 \le x < \epsilon, \\ \Phi_{i\to}^{ps} = \Phi_{\to i+1}^{ps} & \text{for} \quad \epsilon < x < l - \epsilon, \\ 0 & \text{for} \quad l - \epsilon < x \le l. \end{cases} \tag{A.5}$$

Eqs. A.2 and A.5, together with the limit $\epsilon \to 0$, lead to Eq. 4.21.

Appendix B

Derivation of Φ (as in Eq. 5.19)

As explained in Sec. 5.2 (pages 77-81) derivation of Φ involves a tedious exercise in the spectral decomposition of the rate matrix

$$\mathcal{R} = \begin{pmatrix} -1 & 1 & 0 & 0 & 0 & 0 \\ 1 & -2 & 1 & 0 & 0 & 0 \\ 0 & 1 & -2+\epsilon & 1+\epsilon & 0 & 0 \\ 0 & 0 & 1-\epsilon & -2-\epsilon & 1 & 0 \\ 0 & 0 & 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix} . \tag{B.1}$$

This matrix has six real, non-degenerate eigenvalues that are (surprisingly) independent of ϵ:

$$\{\lambda_i\} = \{0, -c, -1, -2, -3, -d\} \quad , \tag{B.2}$$

with

$$\begin{aligned} a = 1 - \sqrt{3} \quad , \quad c = 2 - \sqrt{3} \quad , \quad x = 1 + \epsilon \\ b = 1 + \sqrt{3} \quad , \quad d = 2 + \sqrt{3} \quad , \quad y = 1 - \epsilon. \end{aligned} \tag{B.3}$$

The quantities a, b, x and y will be used momentarily.

We have found the following spectral decomposition of $\mathcal{R}$ to be convenient:

$$\begin{aligned}\mathcal{R} &= \sum_{i=1}^{6} \frac{|i\rangle\lambda_i\langle i|}{\langle i|i\rangle} = UN^{-1}\Lambda V \\ &= \begin{pmatrix} \uparrow & & \uparrow \\ \mathbf{u}_1 & \cdots & \mathbf{u}_6 \\ \downarrow & & \downarrow \end{pmatrix} \begin{pmatrix} n_1^{-1} & & \\ & \ddots & \\ & & n_6^{-1} \end{pmatrix} \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_6 \end{pmatrix} \begin{pmatrix} \leftarrow & \mathbf{v}_1 & \rightarrow \\ & \vdots & \\ \leftarrow & \mathbf{v}_6 & \rightarrow \end{pmatrix}.\end{aligned} \tag{B.4}$$

Here, the columns of U are right eigenvectors of $\mathcal{R}$, and the rows of V are its left eigenvectors. We denote the right eigenvectors by $\mathbf{u}_i$ or $|i\rangle$, and the left eigenvectors by $\mathbf{v}_i^T$ or $\langle i|$. These form a biorthogonal pair of basis sets: $\mathbf{v}_i^T \cdot \mathbf{u}_j = \langle i|j\rangle = n_i\delta_{ij}$, i.e. $VU = N$. Explicitly,

$$U = \begin{pmatrix} x & 1 & x & 1 & x & 1 \\ x & -a & 0 & -1 & -2x & -b \\ x & c & -x & -1 & x & d \\ y & -c & -y & 1 & y & -d \\ y & a & 0 & 1 & -2y & b \\ y & -1 & y & -1 & y & -1 \end{pmatrix}, \quad V = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ y & -ay & cy & -cx & ax & -x \\ 1 & 0 & -1 & -1 & 0 & 1 \\ y & -y & -y & x & x & -x \\ 1 & -2 & 1 & 1 & -2 & 1 \\ y & -by & dy & -dx & bx & -x \end{pmatrix}, \tag{B.5}$$

and $\{n_i\} = \{6, 12c, 4, 6, 12, 12d\}$. Note that since $\mathcal{R}$ is not symmetric, its left and right eigenvectors differ. The matrices N and Λ are diagonal. While it is usual to normalize the left and right eigenvectors so that they are biorthonormal ($n_i = 1$), we have found that the choice of normalization given above leads to less cumbersome expressions in the subsequent analysis.

In terms of this decomposition, we have

$$\mathcal{T} = \mathcal{P}_D e^{\mathcal{R}\tau} \mathcal{M} = \Big(\mathbb{I} \quad \mathbb{I} \Big) U N^{-1} e^{\Lambda \tau} V \begin{pmatrix} p_0 \, \mathbb{I} \\ \\ p_1 \, \mathbb{I} \end{pmatrix} \tag{B.6}$$

where $\mathbb{I}$ is the 3×3 identity matrix (see Sec. 5.2 for the definition and meaning of all these matrices). An explicit evaluation yields

$$\begin{aligned} \mathcal{T} = & \frac{1}{12} \begin{pmatrix} F+G+\delta H & M-2\delta L & F-G+\delta H \\ M & M+12\sigma^3 & M \\ F-G-\delta H & M+2\delta L & F+G-\delta H \end{pmatrix} \\ & - \frac{\epsilon}{12} \begin{pmatrix} H+\delta(G-6\sigma) & -2L & H-\delta(G-6\sigma) \\ 0 & 0 & 0 \\ -H-\delta(G-6\sigma) & 2L & -H+\delta(G-6\sigma) \end{pmatrix} \end{aligned} \tag{B.7}$$

where $\sigma = e^{-\tau}$ and

$$\begin{aligned} F = 4 + 2\sigma^3 \quad , \quad G = 4\sigma^2 + \sigma^c + \sigma^d \quad , \quad H = \sqrt{3}(\sigma^c - \sigma^d) \\ L = 2\sigma^2 - \sigma^c - \sigma^d \quad , \quad M = 4 - 4\sigma^3 \end{aligned} \quad . \tag{B.8}$$

Solving the equation $\mathcal{T}\mathbf{q} = \mathbf{q}$ (see Sec. 5.2) we obtain

$$\mathbf{q} = \frac{1}{3} \begin{pmatrix} 1+N \\ \\ 1 \\ \\ 1-N \end{pmatrix} \quad , \quad N(\delta, \epsilon) = \frac{(\delta - \epsilon)(H - L)}{6 - G + \epsilon\delta(G - 6\sigma)} \quad . \tag{B.9}$$

Combining this result with Eq. 5.18 of the main text yields the statistics of the outgoing bits, (p_0', p_1'), from which we then obtain the circulation using the relation $\Phi = p_1' - p_1$.

Appendix C

Derivation of inequality 6.26

We suggested the inequality

$$\eta > 0 \tag{C.1}$$

in the relation 6.26. It is difficult to derive this inequality form the exact expression of η (in Eq. 6.20). An alternative derivation is possible from the modified Clausius inequality, Eq. 6.28 in the main text:

$$Q_{c\to h}(\beta_h - \beta_c) + \Delta S_B \geq 0\,. \tag{C.2}$$

We present this alternate derivation in this appendix.

To investigate the sign of η, let us take $\delta \neq \epsilon$ [1] and rewrite Eq. C.2 (Eq. 6.28 in the main text) in the form

$$f(\delta') > f(\delta)\,, \tag{C.3}$$

where

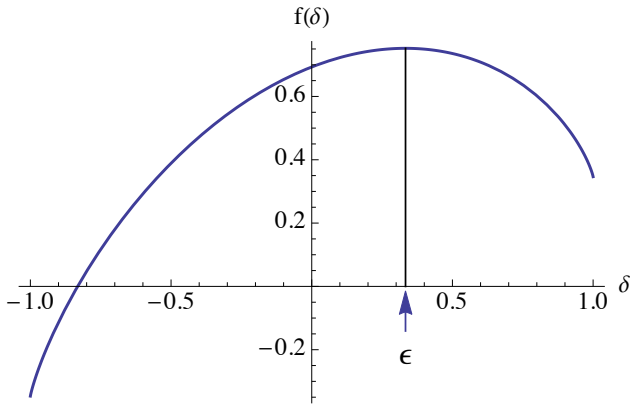
$$f(\delta) = K\delta + S(\delta) \quad , \quad K = \frac{1}{2}(\beta_c - \beta_h)\Delta E > 0. \tag{C.4}$$

Eq. C.3 follows by the direct substitution of the relations

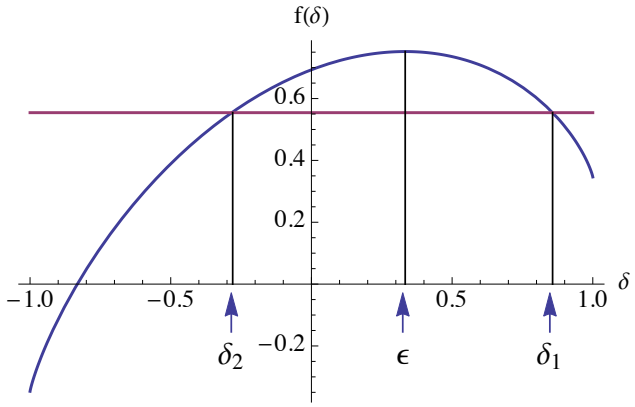
$$Q_{c\to h} = \Phi\Delta E \quad , \quad \Phi = \frac{\delta - \delta'}{2} \quad , \quad \Delta S_B = S(\delta') - S(\delta) \tag{C.5}$$

into Eq. C.2, using a strict inequality since $\delta \neq \epsilon$.

[1] When $\delta = \epsilon$, the value of η is inconsequential, by Eq. 6.20a.



(a)



(b)

Figure C.1: **Illustration of the nature of $f(\delta)$.** (a) $f(\delta)$ is concave and has a maximum at $\delta = \epsilon$, as illustrated for $\epsilon = 1/3$. (b) For a given δ_1, we must have $\delta_2 < \delta_1' < \delta_1$ to ensure $f(\delta_1') > f(\delta_1)$. Hence, both δ_1' and ϵ lie to the left of δ_1.

By construction, $\mathrm{d}^2 f/\mathrm{d}\delta^2 < 0$. Setting $\mathrm{d}f/\mathrm{d}\delta = 0$, the unique maximum of $f(\delta)$ is easily shown to occur at $\delta = \epsilon$, as illustrated in Fig. C.1(a) for $\epsilon = 1/3$. Now let δ_1 and δ_2 denote two values of δ that correspond to the same value of f, with $\delta_2 < \epsilon < \delta_1$, as shown in Fig. C.1(b). Let δ_1' describe the surplus of 0's in the outgoing bit stream, when the incoming stream is characterized by δ_1. Because the maximum of $f(\delta)$ occurs at $\delta = \epsilon$, Eq. C.3 implies that $\delta_2 < \delta_1' < \delta_1$; see Fig. C.1(b). If we instead consider incoming and outgoing bit streams described by δ_2 and δ_2', then the same argument gives us $\delta_2 < \delta_2' < \delta_1$. We therefore conclude that the incoming and outgoing bit streams necessarily satisfy

$$\mathrm{sign}(\delta - \delta') = \mathrm{sign}(\delta - \epsilon)\,, \tag{C.6}$$

in other words δ' lies on the same side as ϵ with respect to δ. Since

$$\frac{\delta - \delta'}{2} = \Phi = \frac{\delta - \epsilon}{2}\eta\,, \tag{C.7}$$

we must have $\eta > 0$.

Bibliography

[1] R. P. Feynman, "There's plenty of room at the bottom," in *Miniaturization*, H. D. Gilbert, editor (Van Nostrand Reinhold, New York, 1961).

[2] E. K. Drexler, "Molecular engineering: an approach to the development of general capabilities for molecular manipulation," *Proc. Natl. Acad. Sci. U.S.A.* **78**, 5275 (1981).

[3] E. K. Drexler, *Nanosystems* (Wiley-Interscience, New York, 1992).

[4] V. Balzani, A. Credi, F. M. Raymo, and J. F. Stoddart, "Artificial molecular machines," *Angew. Chem. Int. Ed.* **39**, 3348 (2000).

[5] R. Ballardini, V. Balzani, A. Credi, M. T. Gandolfi, and M. Venturi, "Artificial molecular-level machines: which energy to make them work?," *Acc. Chem. Res.* **34**, 445 (2001).

[6] *Molecular Machines*, T. R. Kelly, editor (Springer, Berlin, 2005).

[7] E. R. Kay, D. A. Leigh, and F. Zerbetto, "Synthetic molecular motors and mechanical machines," *Angew. Chem. Int. Ed. Engl.* **46**, 72 (2007) (and references therein).

[8] J. Bath and A. J. Tuberfield, "DNA nanomachines," *Nat. Nanotechnol.* **2**, 275 (2007).

[9] M. Mickler, E. Schleiff, and T. Hugel, "From biological towards artificial molecular motors," *ChemPhysChem* **9**, 1503 (2008).

[10] F. C. Simmel, "Processive motion of bipedal DNA walkers," *ChemPhysChem* **10**, 2593 (2009).

[11] J. Michl and C. H. Sykes, "Molecular rotors and motors: Recent advances and future challenges," *ACS Nano* **3**, 1042 (2009).

[12] P. L. Anelli, N. Spencer, and J. F. Stoddart, "A molecular shuttle," *J. Am. Chem. Soc.* **113**, 5131 (1991).

[13] D. A. Leigh, J. K. Y. Wong, F. Dehez, and F. Zerbetto, "Unidirectional rotation in a mechanically interlocked molecular rotor," *Nature* **424**, 174 (2003).

[14] M. von Delius, E. M. Geertsema, and D. A. Leigh, "A synthetic small molecule that can walk down a track," *Nat. Chem.* **2**, 96 (2009).

[15] K. Lund, A. J. Manzo, N. Dabby, N. Michelotti, A. Johnson-Buck, J. Nangreave, S. Taylor, R. Pei, M. N. Stojanovic, N. G. Walter, E. Winfree, and H. Yan, "Molecular robots guided by prescriptive landscapes," *Nature* **465**, 206 (2010).

[16] H. Gu, J. Chao, S. J. Xiao, and N. C. Seeman, "A proximity-based programmable DNA nanoscale assembly line," *Nature* **465**, 202 (2010).

[17] P. Lussis, T. Svaldo-Lareno, A. Bertocco, C.-A. Fustin, D. A. Leigh, and A.-S. Duwez, "A single synthetic small molecule that generates force against a load," *Nat. Nanotechnol.* **6**, 553 (2011).

[18] M. R. Panman, P. Bodis, D. J. Shaw, B. H. Bakker, A. C. Newton, E. R. Kay, A. M. Brouwer, W. J. Buma, D. A. Leigh, and S. Woutersen, "Operation mechanism of a molecular machine revealed using time-resolved vibrational spectroscopy," *Science* **328**, 1255 (2010).

[19] R. Pei, S. K. Taylor, D. Stefanovic, S. Rudchenko, T. E. Mitchell, and M. N. Stojanovic, "Behavior of polycatalytic assemblies in a substrate-displaying matrix," *J. Am. Chem. Soc.* **128**, 12693 (2006).

[20] J. Howard, *Mechanics of Motor Proteins and the Cytoskeleton* (Sinauer, Sunderland, Massachusetts, 2001).

[21] M. Schliwa, *Molecular Motors* (Wiley, Weinheim, Germany, 2006).

[22] *Molecular Motors: Methods and Protocols*, A. O. Sperry, editor (Humana Press, Totowa, New Jersey, 2007).

[23] *Molecular Machines in Biology: Workshop of the Cell*, J. Frank, editor (Cambridge University Press, New York, 2011).

[24] J. Schnakenberg, "Network theory of microscopic and macroscopic behavior of master equation systems," *Rev. Mod. Phys.* **48**, 571 (1976).

[25] T. L. Hill, *Free Energy Transduction in Biology* (Academic Press, New York, 1977).

[26] A. B. Kolomeisky and M. E. Fisher, "Molecular motors: A theorist's perspective," *Annu. Rev. Phys. Chem.* **58**, 675 (2007).

[27] R. D. Astumian, "Adiabatic operation of a molecular machine," *Proc. Natl. Acad. Sci. U.S.A.* **104**, 19715 (2007).

[28] N. A. Sinitsyn and I. Nemenman, "Universal geometric theory of mesoscopic stochastic pumps and reversible ratchets," *Phys. Rev. Lett.* **99**, 220408 (2007).

[29] S. Rahav, J. Horowitz, and C. Jarzynski, "Directed flow in nonadiabatic stochastic pumps," *Phys. Rev. Lett.* **101**, 140602 (2008).

[30] V. Y. Chernyak and N. A. Sinitsyn, "Pumping restriction theorem for stochastic networks," *Phys. Rev. Lett.* **101**, 160601 (2008).

[31] J. Ohkubo, "The stochastic pump current and non-adiabatic geometric phase," *J. Stat. Mech.: Theor. Exp.* (**2008**) P02011.

[32] V. Y. Chernyak and N. A. Sinitsyn, "Robust quantization of a molecular motor motion in a stochastic environment," *J. Chem. Phys.* **131**, 181101 (2009).

[33] N. A. Sinitsyn, "The stochastic pump effect and geometric phases in dissipative and stochastic systems," *J. Phys. A: Math. Theor.* **42**, 193001 (2009).

[34] N. A. Sinitsyn, A. Akimov, and V. Y. Chernyak, "Supersymmetry and fluctuation relations for currents in closed networks," *Phys. Rev. E* **83**, 021107 (2011).

[35] J. Ren, V. Y. Chernyak, and N. A. Sinitsyn, "Duality and fluctuation relations for statistics of currents on cyclic graphs," *J. Stat. Mech: Theory Exp.* (**2011**) P05011.

[36] A. V. Akimov and N. A. Sinitsyn, "Sensitivity field for nonautonomous molecular rotors," *J. Chem. Phys.* **135**, 224104 (2011).

[37] V. Y. Chernyak, J. R. Klein, and N. A. Sinitsyn, "Quantization and fractional quantization of currents in periodically driven stochastic systems. I. Average currents," *J. Chem. Phys.* **136**, 154107 (2012).

[38] V. Y. Chernyak, J. R. Klein, and N. A. Sinitsyn, "Quantization and fractional quantization of currents in periodically driven stochastic systems. II. Full counting statistics," *J. Chem. Phys.* **136**, 154108 (2012).

[39] K. H. Kim and H. Qian, "Fluctuation theorems for a molecular refrigerator," *Phys. Rev. E* **75**, 022102 (2007).

[40] T. Sagawa and M. Ueda, "Second law of thermodynamics with discrete quantum feedback control," *Phys. Rev. Lett.* **100**, 080403 (2008).

[41] T. Sagawa and M. Ueda, "Minimal energy cost for thermodynamic information processing: Measurement and information erasure," *Phys. Rev. Lett.* **102**, 250602 (2009).

[42] F. J. Cao, M. Feito, and H. Touchette, "Information and flux in a feedback controlled Brownian ratchet," *Physica A* **388**, 113 (2009).

[43] T. Sagawa and M. Ueda, "Generalized Jarzynski equality under nonequilibrium feedback control," *Phys. Rev. Lett.* **104**, 090602 (2010).

[44] S. Toyabe, T. Sagawa, M. Ueda, E. Muneyuki, and M. Sano, "Experimental demonstration of information-to-energy conversion and validation of the generalized Jarzynski equality," *Nat. Phys.* **6**, 988 (2010).

[45] M. Ponmurugan, "Generalized detailed fluctuation theorem under nonequilibrium feedback control," *Phys. Rev. E* **82**, 031129 (2010).

[46] J. M. Horowitz and S. Vaikuntanathan, "Nonequilibrium detailed fluctuation theorem for repeated discrete feedback," *Phys. Rev. E* **82**, 061120 (2010).

[47] J. M. Horowitz and J. M. R. Parrondo, "Thermodynamic reversibility in feedback processes," *Europhys. Lett.* **95**, 10005 (2011).

[48] D. Abreu and U. Seifert, "Extracting work from a single heat bath through feedback," *Europhys. Lett.* **94**, 10001 (2011).

[49] S. Vaikuntanathan and C. Jarzynski, "Modelling Maxwell's demon with a microcanonical Szilard engine," *Phys. Rev. E* **83**, 061120 (2011).

[50] T. Sagawa, "Thermodynamics of information processing in small systems," *Prog. Theor. Phys.* **127**, 1 (2012).

[51] M. V. Berry, "Quantal phase factors accompanying adiabatic changes," *Proc. R. Soc. Lond. A* **392**, 45 (1984).

[52] *Maxwell's Demon 2: Entropy, Classical and Quantum Information, Computing*, H. S. Leff and A. F. Rex, editors (IOP, Bristol, U.K., 2003).

[53] J. C. Maxwell, *Theory of Heat* (Longmans, London, 1871).

[54] W. Thomson, "Kinetic theory of the dissipation of energy," *Nature* **9**, 441 (1874).

[55] M. Smoluchowski, "Experimentell nachweisbare, der üblichen thermodynamik widersprechende molekularphänomene [Experimentally verifiable phenomena of molecules contradicting usual thermodynamics]," *Phys. Z.* **13**, 1069 (1912) German.

[56] L. Szilard, "Über die Entropieverminderung in einem thermodynamischen System bei Eingriffen intelligenter Wesen [On the decrease of entropy in a thermodynamic system by the intervention of intelligent beings]," *Z. Phys.* **53**, 840 (1929) German (English translation by A. Rapoport and M. Knoller available in Ref. [52]).

[57] M. Brillouin, "Maxwell's demon cannot operate: Information and entropy. I.," *J. Appl. Phys.* **22**, 334 (1951).

[58] R. Landauer, "Dissipation and heat generation in the computing process," *IBM J. Res. Dev.* **5**, 183 (1961).

[59] O. Penrose, *Foundations of Statistical Mechanics: A Deductive Treatment* (Pergamon Press, Oxford, 1970).

[60] C. H. Bennett, "The thermodynamics of computation – a review," *Int. J. Theor. Phys.* **21**, 905 (1982).

[61] W. H. Zurek, "Thermodynamic cost of computation, algorithmic complexity and the information metric," *Nature* **341**, 119 (1989).

[62] J. Earman and J. D. Norton, "Exorcist XIV: The wrath of Maxwell's demon. Part I. From Maxwell to Szilard," *Stud. Hist. Philos. M. P.* **29**, 435 (1998).

[63] J. Earman and J. D. Norton, "Exorcist XIV: The wrath of Maxwell's demon. Part II. From Szilard to Landauer and beyond," *Stud. Hist. Philos. M. P.* **30**, 1 (1999).

[64] J. D. Norton, "Waiting for Landauer," *Stud. Hist. Philos. M. P.* **42**, 184 (2011).

[65] H. T. Quan, Y. D. Wang, Y.-X. Liu, C. P. Sun, and F. Nori, "Maxwell's demon assisted thermodynamic cycle in superconducting quantum circuits," *Phys. Rev. Lett.* **97**, 180402 (2006).

[66] D. Andrieux and P. Gaspard, "Nonequilibrium generation of information in copolymerization processes," *Proc. Natl. Acad. Sci. U.S.A.* **105**, 9516 (2008).

[67] K. Maruyama, F. Nori, and V. Vedral, "Colloquium: The physics of Maxwell's demon and information," *Rev. Mod. Phys.* **81**, 1 (2009).

[68] B. Lambson, D. Carlton, and J. Bokor, "Exploring the thermodynamic limits of computation in integrated systems: Magnetic memory, nanomagnetic logic, and the Landauer limit," *Phys. Rev. Lett.* **107**, 010604 (2011).

[69] A. Hosoya, K. Maruyama, and Y. Shikano, "Maxwell's demon and data compression," *Phys. Rev. E* **84**, 061117 (2011).

[70] L. Granger and H. Kantz, "Thermodynamic cost of measurements," *Phys. Rev. E* **84**, 061110 (2011).

[71] L. del Rio, J. Aberg, R. Renner, O. Dahlsten, and V. Vedral, "The thermodynamic meaning of negative entropy," *Nature* **474**, 61 (2011).

[72] M. G. Raizen, "Demons, entropy and the quest for absolute zero," *Sci. Am.* **304**, 55 (2011).

[73] O. Maroney, "Information processing and thermodynamic entropy," in *The Stanford Encyclopedia of Philosophy*, E. N. Zelta, editor (http://plato.stanford.edu/archives/fall2009/entries/information-entropy/, 2009).

[74] M. Hemmo and O. Shenker, "Maxwell's demon," *J. Philos.* **107**, 389 (2010).

[75] A. B. Pippard, *The Elements of Classical Thermodynamics* (Cambridge University Press, Cambridge, U.K., 1957).

[76] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (Elsevier, Amsterdam, 3rd edition, 2007), chap. V.

[77] D.-Q. Jiang, M. Qian, and M.-P. Qian, *Mathematical Theory of Nonequilibrium Steady States*, vol. 1833 of *Lecture Notes in Mathematics* (Springer, Berlin, 2004).

[78] J. M. Horowitz and C. Jarzynski, "Exact formula for currents in strongly pumped diffusive systems," *J. Stat. Phys.* **136**, 917 (2009).

[79] C. Maes, K. Netočný, and S. R. Thomas, "General no-go condition for stochastic pumping," *J. Chem. Phys.* **132**, 234116 (2010).

[80] A. D. McNaught and A. Wilkinson, *IUPAC. Compendium of Chemical Terminology* (Blackwell Scientific Publications, 2nd edition, 1997).

[81] S. Ganeshan and N. A. Sinitsyn, "Fluctuation relations for current components in mesoscopic electric circuits," *Phys. Rev. B* **84**, 245405 (2011).

[82] M. R. Panman, P. Bodis, D. J. Shaw, B. H. Bakker, A. C. Newton, E. R. Kay, D. A. Leigh, W. J. Buma, A. M. Brouwer, and S. Woutersen, "Time-resolved vibrational spectroscopy of a molecular shuttle," *Phys. Chem. Chem. Phys.* **14**, 1865 (2012).

[83] R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman Lectures on Physics* (Addison-Wesley, Reading, Massachusetts, 1966), vol. 1, chap. 46.

[84] A. Bérut, A. Arakelyan, A. Petrosyan, S. Ciliberto, R Dillenschneider, and E. Lutz, "Experimental verification of Landauer's principle linking information and thermodynamics," *Nature* **483**, 187 (2012).

[85] 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**, 11641(2012).

[86] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra* (SIAM, Philadelphia, Pennsylvania, 2000), chap. 8, pp. 661–704.

[87] *Mathematica*, version 8.0 (Wolfram Research, Inc., Champaign, Illinois, 2010).

[88] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley-Interscience, Hoboken, New Jersey, 2006).

[89] U. Seifert, "Entropy production along a stochastic trajectory and an integral fluctuation theorem," *Phys. Rev. Lett.* **95**, 040602 (2005).

[90] D. Mandal and C. Jarzynski, "A proof by graphical construction of the no-pumping theorem of stochastic pumps," *J. Stat. Mech.: Theory Exp.* (**2011**) P10006.

[91] D. Mandal and C. Jarzynski, "Hybrid models of molecular machines and no-pumping theorem," *J. Chem. Phys.* **137**, 234104 (2012).

[92] D. Mandal and C. Jarzynski, "Work and information processing in a solvable model of Maxwells demon," *Proc. Natl. Acad. Sci. U.S.A.* **109**, 11641 (2012).

[93] D. Mandal, H. T. Quan, and C. Jarzynski, "Maxwell's refrigerator: An exactly solvable model," *Phys. Rev. Lett.* (in press).